ACADEMY OF SCIENCES OF THE ESTONIAN S.S.R.

INSTITUTE OF LANGUAGE AND LITERATURE

CONTENTS

INDEX OF AUTHORS

WAVE MECHANICS OF THE SPEECH SIGNAL

VICTOR SOROKIN

Institute for Information Transmission Problems
Academy of Sciences, USSR
Moscow 101447

ABSTRACT

Processes of control, muscular contraction, articulators deformations, and acoustical oscillations take place in continuous media. Description of these processes can be done by means of the same mathematical technique. Some important properties of the speech production processes are described.

Speech production processes proceed in different physical media : articulation control processes - on a set of α and γ motor neurons and muscle fibers, mechanical oscillation processes of the vocal folds, tongue and lips - in viscous - elastic tissues, acoustical processes - in air cavities of the vocal tract. There are certain special features in each medium which determine characteristics of the processes, but there is also a very important similarity that enables to use actually the same mathematical technique of wave mechanics. This similarity comes from the fact that on each level - acoustical, mechanical and control, the system of speech production is a system with distributed parameters.

Motor units in the articulation control system are discrete elements but their number is great, and their parameters overlap rather a wide range. Thus it is possible to consider that processes of muscle contraction control take place in a certain continious medium. In that medium, for instance, there is a possibility to control dynamic characteristics of the muscle by the working point shift of the muscular receptors and also by displacement of the area of the sarcomers contraction along the muscular fibers from one muscle end to another.

Amplitudes of eigenfunction of muscle elastic deformations change due to these displacements. Although distributed control systems have been studied in engineering, special features of the articulation control system are little known and deep investigations in the area are of great necessity. Some results concerning articulation control system properties are described in [6].

Geometrical parameters of the vocal folds are rather small. Thus folds elastic vibrations accompany waves in all three dimensions. It can be well seen in high speed motion pictures that elastic waves propagate along the focal folds and also in transverse and vertical directions. Besides, surface waves are observed after folds collapse [2]. Characteristics of the waves are determined by mechanical properties of the vocal folds tissues. For example, surface waves dissapear when cancer tumour evolves.

Mathematical models of vocal folds elastic vibrations were investigated in [4,5,6,7,9,10]. Computer modeling shows, that for the description of folds elastic oscillations in a vertical direction only one eigenfunction is sufficient, in a transversal direction two eigenfunctions are required, and oscillations along the vocal folds require three eigenfunctions. Vertical movements create a new, unknown before, excitation source - a piston source, which is active during close vocal slit interval also. A speech synthsizer excited by the described vocal source produces a speech signal with high naturalness.

Geometrical sizes of the tongue and lips are comparatively large. Thus elastic waves do not propagate in the articulators but their movements are "wave-like", as it is seen in the cinemaradiographic motion pictures. Elastic deformations of the tongue and lips are described by the same mathematical technique as elastic deformations of the vocal folds.

It is sufficient to have only one eigenfunction for description of the shape and movements of the lip. This eigenfunction for the lip is just half-wave of sine. Five eigenfunctions must be used for the tongue shape description. In the case the approximation error in the uniform metrics is about 6 - 7 %. Change of the tongue shape is achieved by means of modes control of elastic oscillations. [6]. The velum is an elastic body with distributed parameters too, thus to calculate its deformations the same mathematical technique as for the vocal folds, tongue and lips must be used.

For the frequencies above 200 - 300 Hz the vocal tract is a system with distributed parameters where waves of acoustical oscillations propagate. Fast change of speech parameters and nonstationary processes domination are properties of acoustics in the vocal tract. There is a set of pecularities of speech production acoustical processes which play crucial role both for speech synthesis and speech recognition.

First of all one must reject an idea that the vocal slit is a starting point of the vocal tract. During production of voiceless fricatives and stops the area of the vocal slit is comparable to the minimal area of the vocal tract. Even during phonation the resistance of the vocal slit turns out to be not so high as it was supposed just recently. As a result, processes in the lungs and the trachea have influence upon acoustical parameters of speech signal and, therefore, the vocal slit is located almost in the middle of the vocal tract. During phonation formant frequencies are subjected to noticeable alterations, sometimes up to 20 - 30 % and those alternations are synchronous with vocal folds oscillations [6] . The formant frequencies variations are pretty fast, therefore both in speech synthesis and speech analysis the vocal tract must be considered as a parametric system with fast alternation of parameters.

In addition to those variations the formant frequencies sometimes undergo fast alternations during articulators movements. For example, the rate of the first formant freqiency variation can be 10 times as much as the velocity of articulator movements, if the minimal area of the vocal tract cross-section is sufficiently small. As a result an abrupt change of formant frequencies is observed before and after a closure.

Vocal tract walls yielding is a cause of radial oscillations, which dominate during closure. The first formant frequency in this interval is equal to the radial resonance frequency (150 - 350 Hz) instead of zero as it were in an acoustical system with absolutely rigid walls. There is radiation of a speech signal through the yielding vocal tract walls and, as it was shown in [1] , the radiation occurs mainly in two areas - around the lips and the pharynx. Yielding of the vocal tract walls leads to the "shut" effect, when in areas with a small cross-section the propagation of low-frequency oscillations stops due to walls oscillations in the antiphase.

As a result of all above mentioned facts the only right method of description of speech production acoustics is a method nonsteady-state wave processes. Therefore the formant method of speech synthesis is unadequate in consequence of the hypothesis on steady stateness of processes and slow variations of the vocal tract parameters. This is confirmed by the low quality of formant speech synthesizers. More over, there is a limit (not very high) for improvement of such synthesis naturalness and intelligibility.

As a basis for description of acoustical processes during speech production a method of travelling waves like Kelly - Lochbaum scheme [3] can be used. Originally Kelly - Lochbaum scheme has a set of serios shortcomings. Particularly, it generates specific noise during alternation of the vocal tract area function. However it is possible to solve the noise problem by means of dynamic matching of boundary conditions between cylindrical sections which approximate the vocal tract shape [7,8] . Further improvement of that scheme should concern section and vocal tract length alternation.

Characteristics of the turbulent source of excitation are little dependent on the place of articulation. Difference in acoustical characteristics of fricatives and bursts of stops are results of various positions of the turbulent source in the vocal tract accompanied with change in values of zeroes of the vocal tract transfer function. This effect is also a consequence of the fact that the vocal tract is a system with distributed parameters.

Thus, physics of speech production processes is much more complicated then it was supposed some time ago, but cognition of those processes inspires a hope of development of high quality synthesizers and reliable speech recognition systems.

1. Fant G., Nord L., Branderud P. A note onthe vocal tract wall impedance. STL QPSR, 1976, N 4, p. 13-20.

2. Hirano M. Data from high speed motion pictures. Vocal Fold Physiology Conf. Kurume, 1980, p. 4a1-4a6.

3. Kelly J.L., Lochbaum C.C. Speech synthesis. Proc. 2 Int. Congr. Ac., 1962, paper G42, p. 1-4.

4. Sorokin V.N. Some questions of the general model of speech production generation. Proc. ARSO-8, Lvov, 1974, p. 97-100.

5. Sorokin V.N. The voice source as a system with distributed parameters. Acoust. J., 1981, v.27, N 3, p. 434-440.

6. Sorokin V.N. The speech production theory, 1985.

7. Sorokin V.N. Travelling waves in the vocal tract. Acoust. J.,1986, N 4, p. 506-510.

8. Strube H.W. Time-varying wave digital filters for modeling analog system. IEEE Trans. v.ASSP-30,1982, N6, p.864-868.

9. Titze J.R., Strong W.J. Normal modes in vocal cords tissues. JASA, 1975, v. 57, N 3, p. 736-744.

10. Titze J.R. On the mechanisms of the vocal-folds vibration. JASA, 1976, N 6, p. 1366-1380.

FUNDAMENTALS AND APPLICATIONS IN SPEECH PRODUCTION RESEARCH

OSAMU FUJIMURA
Murray Hill, New Jersey 07974 U.S.A.

AT&T Bell Laboratories

ABSTRACT

Current issues in speech production research are reviewed with some historical perspective. It is emphasized that recent progress in computational and experimental techniques has brought about a substantial change in the research methodology, and that the interaction between linguistic theory and the understanding of the nature of speech signals has substantially contributed to the progress in both abstract description and speech analysis and synthesis.

0. Introduction

In this paper, I would like to express my personal opinion about the direction of research in conjunction with a fairly wide variety of topics in speech production research. The point I would like to make is that we need a deep inquiry into the nature of speech, in its linguistic, psychological, physiological and physical aspects, taking full advantage of the emerging computational techniques, in order to pave the way for future industrial applications as well as to understand what speech is. I will argue that some basic concepts in the theory of phonology and phonetics must be revisited (*cf.* Fujimura [1980]).

1. Physical Process

1.1. Acoustical Theory

According to the acoustical theory of speech production [Fant 1960], the physical process of speech production comprises two basic components: (1) source signal generation: the process of producing the source airflow through the glottis typically for sonorant parts of the signal, and pressure variation anywhere along the vocal tract near constrictions for some of consonantal parts; and (2) vocal tract filtering: the linear process of converting the airflow/pressure source signals into outcoming acoustic waves that represent speech signals.

There have been challenges to the source-filter theory, claiming that the plane-wave assumption is not valid in reality when we consider the three-dimensional turbulence formation above the glottis [Teager 1983]. There is at least one experimental attempt at measuring the three-dimensional distribution of acoustic pressure within the vocal tract in vowel production [Firth 1986]. The Fantian acoustical theory is the only workable (approximation) theory available at present, however.

In particular, it is well known that the vocal tract transfer functions for different vowel articulation gestures can be effectively represented by the F-patterns [Fant 1956]. We have verified this using an acoustical measurement on normal subjects (see Fig. 1). This acoustic measurement of the natural vocal tract does not involve any dc airflow. To the extent the observed transfer characteristics compare with predicted characteristics of naturally produced vowel sounds, our theory captures the essence of the acoustic process.

Perceptually, also, it has been our experience for a long time that a series-type formant synthesizer captures all vowel characteristics in terms of the phonetic values that are familiar to us.

The role of formant transitions, associated characteristically with consonantal gestures, was convincingly demonstrated by

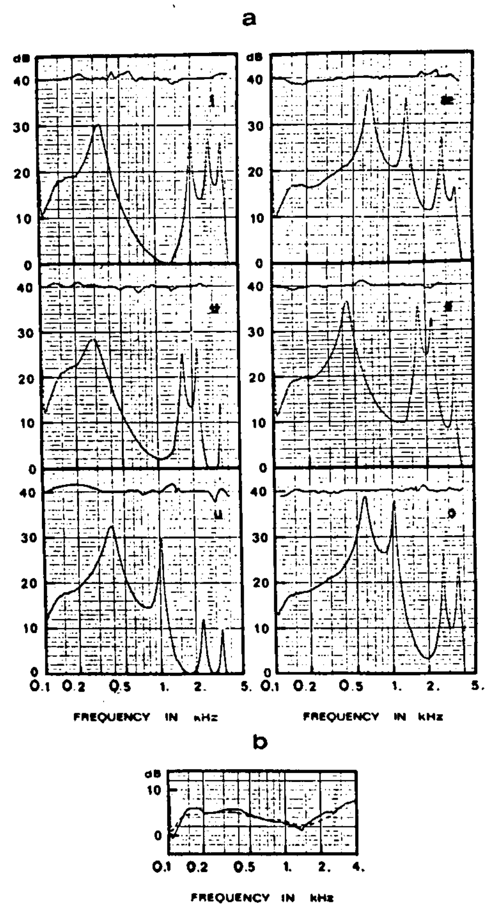


Fig. 1: Vocal tract transfer functions for Swedish vowels, estimated by a sweeptone method using a female native speaker. The curves include a constant frequency function (b), which is assumed to represent the transfer characteristic between the acoustic source (vibrator) output and the virtual excitation source above the glottis. The curve in the upper end of each frame represents the difference between the measurement and the theoretical prediction based on the series formant theory as in Fant [1961]. From Fujimura & Lindqvist [1971].

researchers at Haskins laboratories [Cooper *et al.* 1952; Liberman *et al.* 1954; Delattre *et al.* 1955]. These and related subsequent experiments led us to believe that the quasi-static formant theory was effective enough to capture basic characteristics of speech signals. For further progress, however, I believe this point has to be revisited. There is enough evidence to suspect that our current signal processing technology that is commonly used in automatic recognition schemes, for example, does not capture some crucial information including very rapid transitions for consonantal identification. Also, our knowledge of inherent signal properties of occlusive consonants (see Blumstein & Stevens [1979]) has not been utilized sufficiently in such applications.

The synthesis experiments, incidentally, not only contributed substantially to our understanding of the nature of speech signals and their phonetic perception, but also, in combination with the subsequent emergence of the idea of analysis-by-synthesis [Stevens 1960], set a rather widely applicable methodology for studying complex human information processing.

1.2. Articulatory System

The mandible is literally a basic component of the articulatory system, and our understanding of its function in speech is still far from satisfactory. Edwards [1985] made a fundamental contribution to our knowledge in this field, clarifying its movement patterns in speech involving both rotation and translation relative to the skull.

The velum is presumably the simplest case for studying the phonetic functions relative to its physiological control and the resultant physical configuration. We seem to find a one-to-one correspondence between nasality as a phonological feature and the articulatory gesture of the velum, which probably can be effectively represented by a one-dimensional measure of velum lowering. The acoustical consequences of its movement, however, is by no means simple, nor is it limited to the coupling of the nasal tract to the (proper) vocal tract, as assumed in early works [Maeda 1983] (cf. House & Stevens [1956], Hattori *et al.* [1958]). Also, velum height is affected observably by the raised tongue dorsum for palatal consonants.

Because of the relative simplicity, the lip movement patterns have been the subject of quantitative studies by many investigators (see, for example, Bell-Berti & Harris [1982]). One particularly interesting topic is the relation of lip gestures to the mandible gestures. Macchi [1985] studied this problem in relation to segmental vs. suprasegmental functions using a statistical analysis of microbeam data. She found evidence that while both articulators contribute to the lip closure, some suprasegmental functions are related more closely to the mandible gesture than to the lip proper gesture.

The tongue is the most important articulator in the sense that it determines the largest portion of the vocal tract shape, with a large number of degrees of freedom, resulting in direct acoustical consequences. It is the most complex articulatory organ anatomically, physiologically, and physically. Its phonological implications are also complex.

Since the introduction of x-ray techniques, the laterally viewed midsagittal surface shape has predominated in the discussions of vocal tract modeling. The cylinder model of the tongue for speech synthesis [Coker 1968; Mermelstein 1973] appears to be the most extensively used computational algorithm for deriving area functions out of specifications of articulatory variables. I think it is now clear, however, that we cannot capture some of the most basic principles of articulatory control unless we consider the three-dimensional nature of the articulatory structures more directly.

Let me illustrate this argument with one example, just to demonstrate the nature of the problem [Fujimura & Kakita 1979]. In articulating a high front vowel, say [i], the tongue as a whole is pushed forward by the contraction of the posterior part of the genioglossus muscle. At the same time, some other muscles, including the contraction of the anterior part of the genioglossus, which run vertically near the midsagittal plane only, are used to form a fairly stiff surface shape with a significant groove along the midsagittal line. The resultant tongue surface is bulged upward on the sides. When the tongue is pushed upward and forward keeping this local condition, the sides touch the hard palate and support the tongue against a further upward forward push, leaving the central groove that forms a long and narrow open channel. The back of the tongue is forwarded considerably, creating a wide cavity behind in the lower and middle pharyngeal region. These conditions seem to be crucial for this vowel. The contact on the sides and the relatively rigid surface shape makes the articulation stable, without requiring excessive accuracy of the muscle contractions for forming such a critical narrow passage. This may be considered a viewpoint generalizing Stevens' [1972] concept of the quantal nature of speech production.

This study of the three-dimensional effects of muscular contractions has been performed by the use of the finite-element method of computational simulation [Kiritani *et al.* 1976; Kakita *et al.* 1985], and the interpretation above was inferred from this quantitative simulation work.

This tongue shape formation is inherently a three-dimensional process. It cannot be understood by considering the midsagittal configuration only, even though, after considering all these factors, we may well be able to compute the three dimensional shape, and thereby the area function of the vocal tract, accurately enough. This computation could be carried out, once we understand the mechanism, even from, say, positions of three appropriately chosen sample flesh points of the tongue surface in the midsagittal plane.

Interaction among different anatomical/physiological components of the articulatory system is a particularly difficult issue to study with limited available data. Intricate and often annoying effects of coupling between physical correlates of different linguistic variables have been observed. The segmental effects (of vowel identity, consonantal voicing, etc.) on voice fundamental frequency have been known for a long time (see Kohler [1986] for some relevant discussion). The tongue-larynx interaction has been discussed by Honda [1983], using careful electromyographic evaluation of activities of many relevant muscles.

The hyoid bone is located in between the tongue body and the larynx, conneted to both structures as well as the mandible via muscles and a unique sliding tendon mechanism. An interesting and powerful assumption, if it were true, were that this bone behaved as an effective positional stabilizer via various sensory mechanisms. A recent study by Westbury [personal communication], however, demonstrates, using a cineradiographic observation, that this assumption is not true. Rossi and Autessere [1981] studied related issues concerning the intrinsic pitch of vowels, and provided a realistic picture of the interaction between laryngeal control and tongue gestures based on careful observations.

1.3. Source Signal Generation

Much research has been devoted to and progress has been achieved in understanding the mechanism of voice production. Publications are available, in part in proceedings of the Voice

Foundation Series of the Vocal Fold Physiology Conferences [Stevens and Hirano 1981; Bless and Abbs 1983; Titze and Scherer 1983; Baer, Sasaki and Harris 1987; Fujimura to appear]. The topics range from subcortical neural patterns (in animal vocalization) to computational modeling of the vibration mechanism. Notably, the anatomy, physiology and biophysics of the vocal folds themselves are substantially better understood in comparison to our knowledge, say, ten years ago, demonstrating the benefit of international and interdisciplinary cooperation.

The mucous membrane, the "cover" in Hirano's [1977] terminology, moves relatively independently from the muscular "body" of the folds, in the tangential direction in the speech mode, showing wave propagation along its vertical surface. Fleshpoints on its surface draw roughly elliptic trajectories. This two-dimensional picture (within the coronal plane) of vibration is not new in essense: Kirikae [1943] in his early study using a stroboscopic technique and carbon particles placed on fleshpoints of the muscosa observed clear wave propagation patterns from above in a living subject's larynx. Saito and his group at Keio University recently studied the movement patterns of fleshpoints in the cover and the body by special x-ray techniques applied to excised larynges [Saito *et al.* 1981].

Van den Berg [1957] originally discussed his experimental results about the interaction of the vocal folds with the airflow through them, revealing the basic physical principle of their vibration. Flanagan originated computational simulation of such a vibratory process using a mass and a spring to represent the vocal fold in interaction with the airflow [Flanagan & Landgraf 1968]. Ishizaka and Matsudaira [1968] theoretically demonstrated that realistic vibratory conditions under vocal tract loading can be explained only by adding another degree of freedom, and proposed a now classical two-mass model of the vocal fold vibration mechanism. This minimally approximates the three-dimensional structure by two pairs of mass and spring coupled with each other and with airflow. Flanagan and Ishizaka [1976] then produced a computational simulation of this model coupled with the vocal tract, demonstrating significant segmental effects of the vocal tract loading on the voice fundamental frequency.

Titze and Talkin [1979] approached this issue by using a more detailed part-by-part approximation. Fujimura [1981a] discussed the tension control mechanism based on the body-cover theory. Kakita *et al.*[1981] contributed biomechanical measurements of elastic properties of tissues. Titze and coworkers discussed various aspects of the vocal fold vibration mechanism such as energy exchange between the air and tissues [Titze 1985] and contributions of extralaryngeal factors to the voice fundamental frequency [Titze & Durham 1987]. Conrad [1987] proposed a functional interpretation of the fold vibration based on negative resistance as the key concept. Stevens [1974, 1975, to appear] also used circuit analogy and discussed large-amplitude as well as small-amplitude oscillatory characteristics. Rothenberg [1981] has contributed further insights into the interaction between the vibration and acoustic loading.

Fant [1983a] has been studying functional models of the source (volume flow) waveform in voicing using a unique system of parametric specifications, capturing important characteristics of the voice quality.

Experimental and computational studies of the vortices just above the glottis also are interesting from many points of view. It is a difficult area of experimental studies because of the small dimensions involved. Unfortunately, scaled-up physical experiments do not provide us with straightforwardly interpretable results.

With respect to the turbulent noise generation for fricatives, reader are referred to recent PhD dissertations by Shadle[1985] and Thomas[1985].

2. Physiological and Psychological Studies

2.1. Principles of Coordination and Control

One basic question in speech production research is what principle prescribes the time course of utterance, or the temporal pattern of motor commands for it, given the informational content of the message to be carried by it. The tacit assumption always has been that speech is a common daily activity for human life, and an uttererance must be economical in some sense [Lindblom 1983; Kent 1983]. Nelson[1983] faced this issue straightforwardly, and proposed a control-theoretical account. Based on this principle, combined with the concept of the quantal nature of speech production [Stevens 1972], Perkell and Nelson [1982] discussed some related stability issues of vowel articulation using microbeam data.

Among many topics concerning coordination of different organs in articulatory gestures, the concept of motor equivalence [Hughes & Abbs 1976; Abbs 1979] poses an interesting question with respect to high level planning and control in speech production [Abbs & Gracco 1982, 1983]. For example, a bilabial stop consonant inherently requires that the lips be closed, as its positional target gesture. For this condition to be (nearly) achieved, in terms of motor control, various patterns of activities of multiple muscles could be used. If for example the lip constriction gesture and the mandible raising gesture can be mixed in different proportions for the same goal, say lip closure, the proportional contributions of different articulators may vary from occasion to occasion. The question is what factors determine the variation and how we can describe the regularity involved.

Abbs and Gracco [to appear] report that in a repetition of a word 'sapple', the excursions of the upper and lower lips (measured at their vermilion borders) and of the mandible, for the vowel to consonant movement in the first syllable, are individually more variable than the resultant distance between the fleshpoints of the lips, over repeated utterances for each subject. They argue that such relative invariance of physical quantities that are directly related to the acoustic and perceptual consequences suggests that there is a strong role played by high level motor planning and adaptive (real time) control that combine the uses of different organs to achieve the given target gesture. According to them, the temporal coordination of a series of such target events also is fixed and is controlled at a level higher than for the movements of individual organs.

This is an appealing hypothesis. There are different feedback paths available for speech production, and they are often crucial for understanding aspects of normal speech. Unless we understand the way abstract planning and control are related to signal level phenomena, we may not be able to interpret the meaning of observed signals at an intermediate level, such as mandible movement as opposed to gestures inherently related to the control of lips proper. We may then ask which level of observation in the hierarchy of speech production control is most directly relevant to the description of the phonetic process. Random variation must exist, but identifying its existence is hardly sufficient, particularly when an independent mechanism such as saturation due to the direct contact of lips with each other contributes to statistical reduction of variability of the measured dimension. Macchi's work [ibid.], on the other hand, does suggest that variabilities of different component organs do reflect specific shares of different linguistic functions.

Several investigators recently proposed hypothetical principles for speech production coordination. Particularly at issue is how the temporal organization is designed, and what quantities remain invariant, given a phonetic identity of the speech material, resisting various causes of variation of the signals (see Perkell & Klatt [1986] for a collection of relevant discussions). Since the pioneering work by Lindblom [1964] and Öhman [1967], the basic concept of temporal organization for most investigators remained a concatenated series of positional target gestures representing phonemic segments, supplemented by a smoothiing process called coarticulation. In addition to this basic point of view, Kozhevnikov and Chistovic [1965] introduced a sequential planning model based on a statistical analysis of motor execution variation, and proposed a CV-type syllabic organizational unit. Henke[1966]'s look ahead model generalized the notion of coarticulation to include anticipation. In this connection, recently, Sternberg and his coworkers [1978, 1980] contributed a rather intriguing discovery about how the motor program for what appears to represent a stress group or a foot in English is formed prior to the utterance.

Kelso and his coworkers [1986] hypothesized a general speech production principle in accordance with a popular theory of neuromechanical control of biological systems [Haken 1977]. The basic idea is to assume simple oscillation as an underlying mechanism of speech production, and seek invariance in the phase relations among the underlying oscillatory movements of different articulators which form a task-oriented coordinative structure. They have conducted sets of experiments measuring relative contributions of lip-mandible gestures to bilabial closures. They go farther and argue that their result suggests some support of the consonant-vowel configuration as the basic phonological unit. (for my criticism and authors' reply, see [Fujimura 1986b; Kelso *et al.* 1986a].

When we consider apparent variance and invariance of specially designed and somewhat artificial (repetitive or purturbed) tasks, we need to be careful in interpreting data in different experimental situations. Different feedback paths may be used in different mixtures depending on the particular task and situation. Eliminating crucial dependence on one mechanism in one experimental situation does not lead us to conclude the lack of the use of that mechanism in other situations even for the same phonetic gesture. For example, repetitive utterance materials may introduce apparent characterization of movement control which may not properly belong to the nature of speech in general.

On the other hand, it is highly desirable, from a data-interpretation point of view, to design systematically controlled speech material, even at some cost of undetermined influence of the artificial contexts. A word paradigm, for example, comparing different vowel contexts for the same consonant in the same phonological environment, is never perfectly uniform with respect to, say, word familiarity, even if all words are natural existing words. I think we need to use both situations in such a case: natural linguistic materials in which items are not completely comparable, and systematically distributed artificial paradigms which must resort to some "phonetic performance" even by nonphoneticians, for the purpose of mutual calibration.

2.2. Neural Control of the Larynx and Sensory Mechanisms

Direct electrical access to higher level neural activities is not achievable in normal circumstances, at present, in spite of some promising new technologies such as highly sensitive magnetic field measurements. As for control of vocalization, however, animal experiments have made solid progress in our understanding about neural paths and control functions, for example relating activities at the brain stem level to laryngeal and other control in the monkey [Zealear 1987]. Sensory characteristics also have been studied by direct access to the afferent nerves. Davis and Nail [to appear] report on activities of both tonic and silent myelinated fibers of the internal laryngeal nerve of the cat, in response to carefully servo-controlled mechanical stimuli as well as chemical stimuli.

2.3. Observations in Pathologies and Speech Errors

One informative approach toward inaccessible human processes is to observe different types of pathological cases and compare them with normal cases. This is a rich and rewarding field, and in connection with the new research center with the microbeam facility at the University of Wisconsin, we expect substantial progress. For example, using the microbeam system at the University of Tokyo, Hirose and Kiritani [1985] obtained revealing data in cases of ataxia.

Another large area of study is speech errors. Recent studies take note of the fact that phoneme or feature value confusions between segments do not occur indiscriminately with respect to their role in syllable or word composition, and provide new framework for the description of the cognitive phenomena responsible for phonological performance [Kupin 1979; MacNeilage 1985]. Along the same line, developmental observations of child language and speech contain unique and valuable data.

3. Instrumental Methods

Algorithms of speech signal processing have become commonly available and several commercial systems exist for routine interactive studies of speech (acoustic) signals, using personal computers and workstations extensively. Major research groups often have more specialized advanced systems for efficient measurements of massive data. At the same time, large amounts of systematically collected speech materials are becoming available, with a large-scale effort invested into phonetic as well as some partial syntactic transcriptions of large databases such as the Brown Corpus [Frances & Kucera 1982] and the TI speech data base [Fisher *et al.* 1986].

3.1. Mechanical Measurements

The use of servomechanical adjustment of output impedance under flexible computer control for positional measurements of either flesh points or peripheral structures of articulatory organs provides us with a very powerful means for studying motor control mechanisms in speech [Muller *et al.* 1977]. For understanding the overall feedback functions under well-controlled mechanical conditions, such advanced techniques provide us with new possibilities of, for example, perturbation experiments, extending earlier explorations using the bite block conditions (see for example, Lindblom *et al.* [1979]).
Light-weight mechanical devices have been used by some investigators for obtaining the articulatory degree of nasalization [Horiguchi & Bell-Berti 1984].

3.2. Ultrasonic Measurements

Surface contour information about organs that are not externally accessible can be obtained using ultrasonic techniques, which have seen good progress for clinical purposes. Ultrasonic pulse echo as well as penetration/nonpentration information (using the reflection of beams at the tissue-air boundary) gives us relatively good quality two-dimensional observations for some organs such as the tongue and the larynx, without causing any hazard such as ionization in the subject's body. Some investigators advocate the usefulness of ultrasonic measurements for detecting muscle contraction patterns as well as the surface shape of the tongue [Sonies *et al.* 1981]. As a novel application, Kaneko *et al.* [1981] observed minute vibration of the vocal fold surface in response to external excitation.

The main limitation, in my opinion, of the ultrasonic technique applied to tongue observations lies in the mechanical loading effects on the outside skin. Ultrasonic signals are easily reflected at a boundary between a solid object or liquid and the air, and this necessitates a direct contact of the solid transducer surface onto either the skin itself or some liquid-like material as a transmission medium. This is particularly problematic for measuring movements, because of the inertia of such a medium, while it is circumventable for a carefully designed static measurement. The under surface below the floor of the tongue is quite soft, but it easily transmits force through the tongue, causing unknown dynamic deformation. With careful application, particularly in combination with other methods like x-ray microbeam for calibration, there is a good possibility of extensive use, however, since it can give different information related to the continuous surface contour as opposed to flesh-point sample positions of the tongue.

3.3. Optical Measurements

The use of a special fiberscope for laryngeal observations during speech utterances brought us new opportunities to understand the laryngeal gestures under phonetic control [Sawashima & Hirose 1968; Sawashima 1976]. Recently, in addition to the film and video recording methods in use in the past, a new technique of using a two-dimensional array structure of light-sensitive semiconductor elements (image sensors) has become feasible for high speed recording at a few thousands frames per second [Honda *et al.* 1985; Kiritani *et al.* to appear]. This makes it possible to digitally record glottal images without resorting to stroboscopic methods, which by definition is not very useful for studying any aperiodic characteristics of the vocal fold vibration.

3.4. General X-Ray Techniques

X-rays used to be the only source of information about dynamic tongue gestures, apart from the qualitative information obtained through visual inspection from the outside and subjective tactile and proprioceptive sensations. Because of the hazardous ionization effects in the body, however, the film method using fluoroscopic cineradiography (or other variants) is not generally recommended for extensive data collection of articulatory gestures. It also requires excessive analysis effort frame by frame. The video recording technique is probably significantly better but the situation is not qualitatively different. The basic problem stems from the flood exposure covering the entire image field. It should be mentioned, however, that careful and thorough examinations of limited amounts of film records in earlier years provided us with invaluable understanding of the physics and physiology of articulation [Chiba & Kajiyama 1941; Houde 1967; Perkell 1969; Wood 1979]. Some information with respect to the configuration in the lower pharyngeal regions, for example in relation to pharyngealization in Arabic languages, is also indispensable, at present, even though the available data are extremely limited. El Halees [personal communication] for this purpose used the recently developed xeroradiographic method, producing x-ray pictures of detailed structures with otherwise unimaginable clarity, but the extremely high dosage makes this method hardly applicable to more systematic studies. Rossi & Autesserre [1981] also applied this technique effectively for studying the functions of the hyoid bone.

The computed tomography [Kiritani *et al.* 1977] also provides invaluable information at the cost of a very high dose. It is possible, however, to reduce the required dose substantially, by readjusting the source intensity to barely sufficient amounts for distinguishing air from tissues, rather than using the normal conditions set optimally to differentiate tissue compositions. Another serious limitation of this method for speech research purposes is that the measurement time is inevitably very long, making even stationary vowel gestures somewhat difficult. In this respect, the nuclear magnetic resonance method has the same limitation.

3.5. X-Ray Microbeam System

Unlike the conventional film method, where flood x-rays emerge in a wide solid angle uniformly from a small x-ray generating spot on the target, the x-ray microbeam system uses a deflectable pencil beam of x-rays which is adaptively controlled by a digital computer. I invented the x-ray microbeam method out of the need to study dynamic articulatory gestures with the very minimum use of radiation and for practical feasibility of analysing extensive data. The first generation, a pilot system for testing the method, was implemented in 1968 at the University of Tokyo, with a 50-kV acceleration and a PDP-9 computer for control [Fujimura, Kiritani & Ishida 1973] (supported in part by NIH, USA). A second-generation device was implemented in 1973, with a 150-kV acceleration and a 2-mA electron beam current [Kiritani *et al.* 1975](Japnese governmental grant). This system was used for many data collection experiments, mainly by the University of Tokyo group, myself, and the speech physiology group at Haskins Laboratories in cooperation with the University of Tokyo group.

The third generation has been implemented at the University of Wisconsin, Madison, as the central research tool for a nationally shared speech research facility with research grants given by NIH (PI's: Abbs, Thompson and Fujimura, see Nadler *et al.* [1987]). This new system is designed for a 600-kV/5-mA operation, and is now being operated at 450-kV/5mA.

The reason for the high voltage is primarily twofold: (1) the geometrical design for distortionless image field requires a newly introduced transmission-type x-ray generator, and (2) to be able to cope with extraneous metal objects in the mouth, such as dental fillings, so that the experimenters are not excessively constrained about the choice of subjects in a wide range of experiments including studies of pathologies. In addition, (3) the energy absorbed by the body (*i. e.* ionization effects) is considerably less for the same detected energy, due to the better penetration of high energy photons.

The system is equipped with provisions for simultaneous acoustic and electromyographic data acquisition, and extensive uses by external groups are being scheduled under the coordination of a Users' Committee (K. S. Harris, chair).

A number of metal pellets (gold sphere or cylinder, one to three mm in cross dimension) are placed on the tongue and other articulators, and a few reference pellets are similarly placed on fixed points on the head (for head movement calibration and compensation). Pellets are searched by the microbeam automatically one by one time-sequentially, based on the past positions and according to prescribed prediction and search algorithms. In the new system, the exposure time for each position (pixel) is 2.5 to 10 microseconds, being adaptively chosen, so there will be no excessive radiation after securing a sufficient amount of photon detection. The effective frame-rate varies from pellet to pellet according to the experimenter's specification, and the microbeam is stopped by overdeflection for any moment it is not necessary for pellet identification. The radiation doses in realistic situations using the microbeam scheme are extremely small in comparison with any other x-ray methods.

In addition to obvious reasons for dose reduction due to selective exposures in space and in time, there are more subtle and still important reasons. Because of the use of the thin beam, the scatter photons are created only along the narrow beam, as opposed to the flood x-ray situation where they are created all over the volume of the exposed object, contributing to the summed-over noise registration. This results in a significantly better signal to noise ratio, and for this reason, the equivalent image quality is substantially superior. This, combined with the inherently high detector sensitivity, means that for a given task, even the local x-ray intensity at the point of exposure can be made considerably smaller than in a comparable situation (pellet position identification) using film methods.

The actual accumulative dosage in a few data acquisition sessions has been empirically evaluated using the Tokyo system. Dosimetry film and TLD mosaic have been placed on both the entry side and the exit side of the head to reveal accurate spatial distributions of accumulative dose within the image field, for two sessions each containing approximately 10-minute worth net total exposure. The total dose for such a typical session would be less than the accumulative cosmic ray exposure for the person under normal circumstances. The peak dose rate (averaged over a very small volume along the direction of photons) is really what we should pay attention to in planning experiments, taking a conservative attitude. It was found to be about 10 mR at maximum within the image field for each 1-minute worth net exposure. This means that if we take the local peak dosage as an index for conservative precaution, an hour long net or continuous data acquisition would amount to a peak dose roughly comparable to one dental bitewing shot.

3.6. Magnetic Methods

While the radiation hazards are minimized by the use of the microbeam, it would be nice if we could perform comparable tasks without using ionizing photons at all. Sonoda's early attempt used a small permanent magnet attached on the tongue, its position being determined by externally located field detection coils [Sonoda & Kiritani 1976]. This system has the basic limitation of not being capable of tracking more than one sample point simultaneously. The use of an externally created ac field picked up by a small detector coil in the mouth circumvents this constraint [Oka 1980; Schoenle *et al.* 1983]. Each detector is a 4 x 4 x 2.5-mm coil wrapped around a ferrite core with a pair of thin wires for external connections, and it is glued on the tongue surface as in the case of the microbeam pellets. Perkell and Cohen [personal communication] recently succeeded in tracking one "pellet" on the tongue yielding an extensive set of data. A practical system using a large number of "pellets" simultaneously remains to be developed in order to replace the x-ray microbeam for general purposes of articulatory studies. The crucial dependence on the attached wires leading to the outside measurement system does constitute a limitation. Also, the metal pellets for the microbeam system can be substantially smaller. The magnetic method does have a distinct advantage, however, in not being constrained by metal objects in the mouth such as dental fillings and caps, in addition to the nonuse of ionizing rays.

3.7. Electrical Methods

3.7.1. Palatography

Computerized palatography, in my opinion, is a very useful device for both research and tutorial/clinical purposes. It makes the traditional palatography applicable to moving gestures, and at the same time, the data are now recorded in computer files directly. The idea of using multiple electrodes embedded on an artificial palate, to my knowledge was first tried in Stevens' group at MIT by Rome [1964], who represented the time course pattern using the spectrographic display scheme. This dynamic palatography was then computerized using oscillographic displays [Fujii *et al.* 1971]. We studied characteristics of Japanese apical consonants [Fujimura *et al.* 1973a] and Miyawaki [1972] studied their palatalization using this method. Eek [1973] also applied computerized dynamic palatography to studies of Estonian palatalized consonants, revealing an intriguing difference in the temporal characteristics of phonetic implementations among languages.

Schemes using the same basic principle, called electropalatography or dynamic palatometry, are in use by several groups for phonetic research [Hardcastle 1972, 1974, 1984; Fletcher *et al.* 1975; Sawashima & Kiritani 1985], and for clinical applications [Shibata *et al.* 1979]. The device is now commercially available with new features, particularly with the provision for using ready-made palates as opposed to the palate specially made for the individual.

3.7.2. Glottography

Electroglottography [Fant *et al.* 1966; Smith 1981; Childers *et al.* 1984] and laryngography [Fourcin and Abberton 1977] have been used extensively for phonetic studies of vocal fold vibration patterns. While it is only an indirect indication of the condition of the vocal fold contact, its fast response and the lack of invasive elements makes it practically useful for many situations where other more direct methods of observation are not applicable.

3.7.3. Electromyography

Measurements of the muscle activities are at present the best we can do for directly observing physiological patterns above the physical levels in speech behavior. The use of hook-wire electrodes prevails in electromyographic studies [Hirano & Ohala 1969]. The interpretation of the signals representing contributions from the complex of the muscle fibers under unidentified physical conditions is difficult for rigorous quantitative discussions (see for a careful and elaborate method of single motor unit decomposition, Deluca [1975]). With appropriate care, EMG is the most powerful means for assessing speech control principles via direct measurements (see Fujimura [1979] for a review of its applications in studies of laryngeal control gestures). Combinations with other methods of physical observations are often desirable, and the new research facility at the University of Wisconsin aims at simultaneous digital data recording with the microbeam pellet position measurement.

4. Temporal Organization and Linguistic Structure

The general aim in this area of study is to separate physical constraints from linguistically motivated control. My own approach around 1960, working in Halle and Stevens' group at MIT, was to observe the articulatory dynamics as much as possible, through high-speed motion picture recording and analysis of the lip movement [Fujimura 1961]. Öhman [1967] in the same line of effort, working with Lindblom [1968] at MIT and then KTH (RIT, Sweden), analysed x-ray data of tongue movement as well as acoustic data, and proposed a quantitative model formalizing a now standard concept of coarticulation. At the same time, Öhman proposed the perturbation theory of consonantal articulation, introducing an important conceptual deviation from the classical notion of speech as a single chain of segmental units. He tried to quantify the inherently multidimensional nature of speech, by a method which later would have been called a projection principle.

4.1. Segment Concatenation and Coarticulation

Coarticulation in the Lindblom and Öhman's sense is basically the process of parameter smoothing in the physical realization of a string of phonetic segments [Stevens 1983]. If we take the phoneme to be the segmental unit, however, and expect an observable speech signal or its conventional parametric representation (as in speech synthesis experiments) to be constructed by concatenating segmental target values into a string, it does not capture some important characteristics of natural speech. The concept of coarticulation as a smoothing filter for any parameter, such as formant frequency, quite possibly with some notable asynchrony allowed, can be generalized to include the more traditional and qualitative linguistic notion of assimilation, or what we might call soft coarticulation [Fujimura & Lovins 1978]. This makes the string concatenation model more tenable, but at the same time it makes it more difficult to assess its validity; and still, it is difficult to explain observed *ad hoc* variation of phonemes in different environments *[ibid]*.

An appropriate model of concatenation and smoothing, in my opinion, can be obtained only if we describe the production system using a temporal pattern comprising multiple dimensions, each of which is related to a physiologically controllable variable. The mapping relation between such a set of control variables into the conventional speech signal parameters such as formants and pitch is likely quite complex, involving nonmonotonicity and hysteresis. Also, the control program itself is under the influence of feedback and anticipation. We need to know what these mapping characteristics are, or at least what qualitative constraints they have, before we can determine what the effective variables are for successfully relating abstract linguistic units to physical phenomena. It is a horrendous task to pursue, but recent progress in technology, particularly in computational methods, has made us feel that some progress is in sight (see for examples of research efforts along this direction, Coker [1968]; Browman & Goldstein [1985]). It would not be possible at all, however, if we had to rely entirely on the inductive approach. Since the superpositional principle is not expected to work over the independent variables unless we find an effective transformation, resorting to statistical approaches blindly does not look very promising. Fortunately, recent progress in phonological theory as referred to later, gives us good insight into this issue, and of course, in turn, any discovery in the facts of speech will contribute substantiallly to the formulation of a successful theory of phonology.

In this connection, from an engineering point of view, I believe the optimal choice of a phonetic unit as long as we remain in the segment concatenation method, is the demisyllable or something equivalent [Fujimura 1976; Fujimura *et al.* 1977a; Browman 1980; Macchi 1980]. The demisyllable was also successfully adopted in automatic speech recognition [Rosenberg *et al.* 1983]. The basic reason for the efficacy of demisyllables is that the predominant types of context sensitivity of phonemes, *i. e.* many sorts and degrees of allophonic variation, some hopelessly *ad hoc*, are effectively contained within the domain of the demisyllable apart from the prosodic effects (see *infra*).

Another apparently similar technique is to use phoneme diphones [Sivertsen 1961; Dixon 1968; Olive 1980] as the "segmental" unit for speech synthesis. Being based on the phonemic theory, this approach is originally independent from the demisyllabic one, but in practice both techniques in speech synthesis have been converging using about the same number of units stored in the inventory. Olive's diphone approach has many additional features as well as elaboration in details.

4.2. Phonology and Phonetics -- Intonation and Other Topics

One important recent development in the theory of phonology that bears strong implications on understanding the temporal organization of speech is the trend toward integrating phonetic observations with the very core of the theoretical discussion. This new trend is most strongly seen in the description of intonation/accent patterns, but it can now be found in the entire domain of (nonlinear) phonology, influencing the basic structure of phonological representation from the lexical level down. Articulatory data, collected systematically with careful speech material designs guided by the basic theoretical interest of linguistic structures may soon constitute unique objects of such discussions.

The spirit of nonlinear phonology at least in the case of so-called suprasegmental description is nothing new (see *e. g.* Hattori 1961]), and there has been a relatively long tradition of descriptive work on intonation in Europe involving different experimental methods (see for more recent examples, Vaissiere [1977]; Nishinuma & Rossi [1981]; Gårding [1983]; Thorsen [1984]; for discussion of interacting factors see Nooteboom & Terken [1982]; see also Eek (ed.) [1978] for reports on various studies on different languages). For its theoretical impact in general and formal phonology, we had to wait for the most recent progress using advanced computational environments. Some of the full-scale experimental effortss on sentential intonation by those familiar with linguistic theoretical issues was triggered by astute observations by speech researchers with engineering backgrounds (see *e. g.* [Maeda 1976]). After Liberman [1975]'s theoretical lead (see also Liberman & Prince [1977]), Pierrehumbert's dissertation [1980] established a new experimental/computational methodology of phonological/phonetic studies.

Traditionally, according to the explicit formulation due to generative phonology [Chomsky & Halle 1968], phonological rules constituting a precise body of formal specifications dealing with discrete symbols (specifically binary-valued distinctive features) produced an input to the process of phonetic implementation, which handled numerical or continuously valued variables representing physical correlates of those features. The objects of the entire phonological manipulation were the feature matrices, which separated phonemic segments as "simultaneous bundles" [Jakobson *et al.* 1951] represented by its columns. At the output level of phonology, the so-called systematic phonetic representation used numerical specifications of feature values, as a buffer representation between the symbolic and numerical computations. I do not believe that such an independent level of description is tenable [Fujimura 1970; Keating 1985]). It is now an empirical question if the separation of numerical processing from symbolic manipulation as subcomponents of a body of ordered rules or processes can be maintained (see Ladd [1986] for relevant observations). It is conceivable that the two distinct subsets of rules are not separate in terms of rule ordering but different in their formal properties.

The concept of simultaneous bundles is abstract, just as that of distinctive opposition is. The current argument is that the representation at the most abstract (lexical) level has to be inherently mutidimensional, different features covering different (abstract) temporal domains, and dimensional structures must reflect some articulatory functions [Clements, personal communication; Halle 1983, 1985]. The emerging theory of melody-skeleton association seems to provide a good bridge between our findings about articulatory movement patterns and their temporal organization on the one hand, and the abstract phonological representation necessitated out of distributional and derivational observations on the other [McCarthy personal communication; Fujimura 1986c]. At the same time, phonological representations may specify linguistically significant oppositional values abstractly and sparsely, as opposed to completely segment by segment. Such a descriptive system like those based on a marking convention [Chomsky & Halle 1968; Kean 1975], for example, may make good sense particularly if it is assisted by the syllabic framework and conventions involving resyllabification, for example a scheme proposed by Borowsky [1986].

A point of dispute, given this relaxation of the one-dimensionality (*i. e.* concatenative linearity) or the "simultaneous bundle" constraint, and the introduction of any rather specific but complex structural framework, is whether the abstract feature specifications should be given unit by unit completely, or rather specifications are inherently nonsegmental in the sense that they (whether quasi-static values or dynamic patterns as the "target" configurations) are sparsely specified in the multidimensional space down to the level where numerical implementation rules operate. In the latter case, the realization rules would compute the entire time course of each dimensional variable to specify the temporal course of physical signals for a large phrasal unit. Pierrehumbert (see *infra)* clearly takes the latter view, whereas Inkelas and her coworkers [1987] maintain the former view discussing African tone/intonation phenomena.

A point of future study related to this topic is the nature of phrasing in speech utterance. A three-level framework of phrasing hierarchy has been proposed by Pierrehumbert and Beckman [in press] (also Beckman and Pierrehumbert [1986]) in their studies of Japanese and English. In Japanese, within the minor (accent) phrase, any but the first lexically specified accent marks lose their realization, according to traditional accounts (see for rule formulation, McCawley [1968]; Haraguchi [1975]). This is usually interpreted as an erasure of such marks. Recently, a concept of catathesis was proposed by Poser [1984] (in conjunction with Pierrehumbert's descriptive framework) relating pitch contour realizations in contiguous phrases. This process, unlike the so-called pitch declination, is conditioned crucially by the existence of accent in the preceding phrase. When we handle a larger phrasal unit, according to the catathesis theory, a qualitatively similar phenomenon takes place, but the effect is not to eliminate the mark nor ignore it, but to reduce its manifestation for the subsequent phrases, if and only if there is a preceding accent (in the preceding minor phrase). This raises the following question: Is the accent deletion really a symbolic phonological operation, or is it only a relatively strong degree of reduction? Further, it could be questioned if the distinction between the smaller and larger phrasal units are something of a categorical nature, as expected from the syntactic motivation of the phrasal structures, or is it to be captured (roughly speaking at the phonetic level) as continuously varying boundary effects? That a complex set of discourse factors influence the boundary effects in numerical manners may favor the latter point of view, and as far as I know, there is no evidence contrary to this.

Another set of observations being discussed in terms of the relation between phonology and phonetics concerns the neutralization of phonemic distinctions in certain phonological environments. Dinnsen and Charles-Luce [1984] studying final obstruents in Catalan challenged the separation of phonology and phonetics, claiming that the phonological rule devoicing obstruents must apply after the phonetic implementation rule that accounts for speaker-dependent final devoicing. Similarly, the final consonantal tense/lax or voiced/voiceless opposition in German has been studied by several investigators, both in production and perception [Fourakis & Iverson 1984; Port & O'Dell 1985]. The concept of neutralization was revisited by Fourakis [1984].

Keating [1985] discussed the same difficulty in her study of vowel duration and voice onset timing patterns and has proposed a modification of the theoretical framework, allowing the grammar of a language to control "all aspects of phonetic form". This view may seem necessary to explain what is observed using the traditional framework of phonetic description. The question crucial to the theory of phonology is, however, not just what is sufficient for the description of the observed patterns, but how we can transform observable signal characteristics to units and structures that are effective for phonological representation. If we do not pursue an answer to this linguistic question, we will simply have to yield to the more complex data as we become capable of the more sophisticated measurements.

What is important here, however, is the fact that the fundamental concepts of phonological representation are being challenged, as the result of accurate enough quantitative observations of actual physical signals, together with the technical capability of comparing exactly realized complex mathematical schemes by computation. The conceptual process of speech synthesis by rule is a concrete technical experience in our present-day research environment, and it has emerged, in part, as the result of an engineering interest in a machine that relates lexical representations (often given in orthographic text) to speech signals.

4.3. Articulatory Aspects of Prosodic Control

Traditionally, prosodic effects on speech characteristics have been discussed in connection with their manifestation in voice pitch modulation and temporal patterning of segmental units. Thus, the segmental units (phonemes in most discussions) displayed their inherent physical correlates when they were concatenated into a temporal string, with the coarticulation or smoothing with the resultant reduction or undershooting as the only modification, while pitch and durational modulation were superimposed onto this representation of the speech signal. Some minor (presumably universal) interactions of laryngeal control with articulatory characteristics have also been considered. This picture is typically represented in the tradition of speech synthesis by rule [Liberman *et al.* 1959; Holmes *et al.* 1964].

Recent studies clearly show that this classical view only reflects a lack of precise enough data, or at best, careful avoidance by the phoneticians of the intruding complexity of "nonessential" factors in the phonetic description. Every phonetician has known that samples of vowels could not be collected from different contexts, segmental, suprasegmental, or extralinguistic, for physical measurements to yield valid comparison of contrasting phonemes. Even a narrow phonetic transcription cannot be performed mechanically, because supposedly identical phonetic segments of a language would vary from one condition to another. The engineering interest in designing a machine to identify words for practical purposes has compelled us to confront this outstanding problem (for some relevant discussions, see Fujimura [1984]; M. Ohala [1983]).

4.3.1. Focus and Phrasing

Contrastive emphasis placed on a particular word in a sentence utterance introduces remarkable effects not only in the pitch contour and segmental durations, but also in the articulatory movement patterns including what appears to be the target position. Fig. 2 illustrates an example of the vertical movement of a metal pellet placed on the blade of the tongue, tracked by the x-ray microbeam system at the University of Tokyo. The subject was a phoneticaly trained female speaker of a dialect (Georgia) of American English.

The two utterances demonstrate the effects of different placements of focus (contrastive emphasis). It can be readily seen that the affected words are uttered with radically different gestures. The syllable nucleus of the word "six" (see the single arrow) shows more than three times as deep a valley in utterance (a), accompanied by a considerably extended time interval between the downward and upward (transitional) movements. The

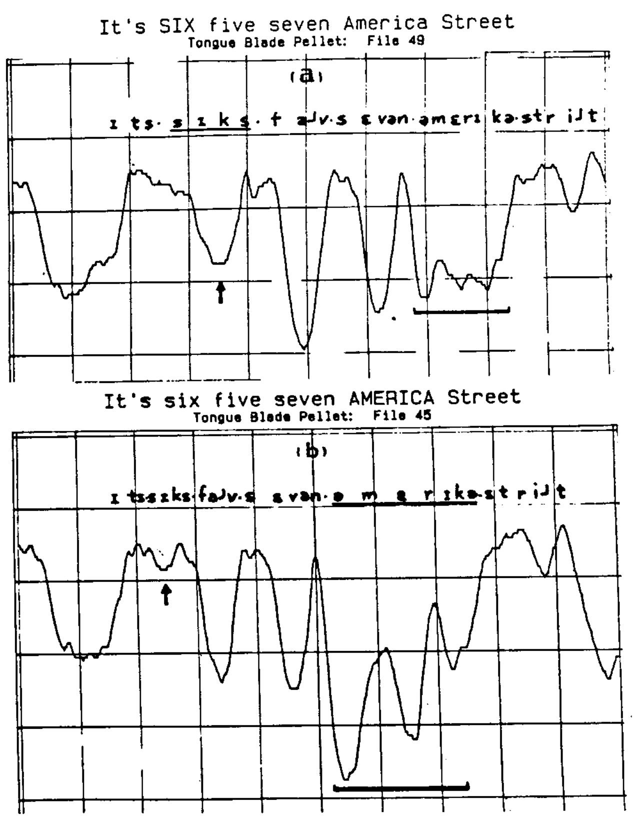


Fig. 2: Tongue blade movement (vertical position) as recorded by x-ray microbeam tracking. 'It's six five seven America Street' spoken by a female American speaker, with focus placed on (a) 'six' and (b) 'America'.

consonantal gestures on both sides of the valley also show some differences between the two versions. The portions of the utterances representing the word 'America' (see the bracket), on the other hand, demonstrates an even more dramatic difference in gesture. There, the observed tongue blade height patterns share almost nothing between the two versions. Finding correspondence between the two curves as shown by the transcription in the figure, is difficult without resorting to an examination of other pellet positions (and the acoustic signal), in spite of the fact that the two front vowels as well as the /r/ all presumably involve some inherent tongue blade gestures.

In this experiment, a few utterances were recorded for each of the four conditions placing emphasis on different focusable words. The patterns were observed to be qualitatively very consistent among different utterances for the same emphasis condition, but the gestures for emphasized words tended to vary considerably in terms of the extent of the excursion and temporal expansion.

A somewhat similar modulation with less change in the depth of nucleus valleys was observed when distinct phrasing patterns were used for the same word sequence in arithmetic formulae such as

(5 + 5) x 5 (yielding 50) vs. 5 + (5 x 5) (yielding 30).

4.3.2. Iceberg Patterns

In the course of studying temporal organizations of articulatory movement patterns, we realize it is rather difficult to define reliable land marks which we can rely on in comparing utterances of the same phonetic segmental material under different environments. The familiar notion of a segmental boundary (for representative examples see Lehiste [1970]; Umeda [1975]) displayed as acoustic events such as voice onset, consonantal implosion and explosion, do not find corresponding discontinuities in the articulatory time course. This is particularly true, presumably, because we use selected flesh points on the articulators such as the tongue blade, which, depending on the context as well as the particular phoneme, may or may not represent the point of articulation. For a precise timing definition of an event that is crucial in terms of the acoustic consequences, we will have to refer to a three-dimensional measurement covering a wide spatial domain, as seen in dynamic palatographic studies.

In my opinion, however, the apparent difficulty reflects a deeper issue. The dynamics of articulatory structures is inherrently continuous, involving a set of finite quantities like force and mass. Even when the velocity of a particular part of the organ changes abruptly, for example by collision with a heavy hard structure such as the palate, the central part of the organ keeps moving rather smoothly. Apart from the possible indirect reaction through neural feedback, what determines the time course of the entire system reflecting the neural commands is the physically central rather than peripheral part of the structure.

Also, for the purpose of quantitative analyses, a smoothly changing variable is mathematically more tractable than discontinuous functions, because smooth functions can be handled at least locally by a linear approximation. This means that within a selected range of change, the system can be treated as a superpositional system, where different factors can be easily separated out by controlling contributing factors one by one. In a formidably complex process such as speech production, this is perhaps the only practical initial approach, until we have some comprehensive view of the entire system with respect to interrelations among specific parts of the system.

One more reason favoring recording smoothly changing variables is that our measurements are always noisy. A discontinuous time derivative used as the means for evaluating the crucial event is inherently susceptible to errors due to small noise in position measurement. Especially if the purpose is to determine timing values of crucial events, continuously moving parts of the time functions provide the most accurate evaluation of timing, in comparison with, for example, an evaluation of the time when a movement starts from the standstill condition. If we define an event of invariably fast movement of a sample point fixed to the structure, say a pellet, crossing across a prespecified position, say height threshold, then the accuracy using positional measurement is very high with respect to the time evaluation. In order to make phonetically meaningful measurements, however, we have to find a crucial condition, say a specific value of height for the selected flesh point that makes sense as a definition of a phonetic event.

If we have a validated model of the time course for the given observable quantity, say if the system behavior is known to be determined by second order dynamics [Fujisaki 1977, 1983], then we can use a large segment of the time changing variable that covers an interval during which system parameters can be assumed to take constant values, for a semiglobal curve fitting procedure. This is a very noise-resistive method. Some recent temporal studies (*e.g.* Ostry *et al.* [1983]) in effect assume such a simple model (locally sinusoidal change).

My approach is to try to find relatively invariant movement patterns that can be operationally defined reliably enough for the purpose of timing evaluations of landmark events. This method was motivated by the informal observation of various data from microbeam measurements. For some parts (in terms of pellet height) of the movement of the crucial articulator for place-specified consonants (the lower lip for labials and the tongue blade for apicals), fairly reproducible results seemed to emerge with respect to timing modulation of such events as the result of prosodic control [Fujimura 1981, 1986]. Using a special statistical process to automatically and empirically decide such positional ranges for a selected domain of prosodic variablility, we evaluated, for sets of data described above relative to focus and phrasing, the temporal modulation relation of each pair of utterances (see Fig. 5). The data are only preliminary, and await further verification using more data, which hopefully will become available very shortly from the Wisconsin microbeam system.

For such patterns that seem to be characteristic of the consonant-vowel combination, or more generally for a given demisyllable, where the observed articulator is crucial for the place specification of the consonant, I gave the name "iceberg", because such a movement pattern floats around fairly freely in time relative to other articulators' movement patterns, when segmental or prosodic contexts change [Fujimura 1981].

4.3.3. The Case of Velum Movement

Vaissiere [personal communication], using the microbeam data from the University of Tokyo, studied the velum movement patterns in utterances of several sentences, as well as words in isolation, spoken by two native speakers of English (General American). She interpreted the time functions representing the vertical position of a sample point of the velum surface, obtained by tracking a pellet attached on a flexible plastic strip which was placed on the velum in the nasal cavity [Fujimura, Miller & Kiritani 1977]. In prescribing the time course of velum height, she defined the "strength" of the oral consonant with respect to its tautosyllabic effects. The strength is conditioned by intrasyllabic position as well as stress. For the positional target, she concludes tentatively that there is no target values for vowels, varied positions being specified for both nasal and nonnasal consonants depending on nonsegmental conditions.

One particularly interesting observation she has made is that the strategy related to syllable reduction seems to vary basically from one speaker to another. In one speaker, the movement reduction for prosodically weak position seems to be explained by undershooting due to time constraints, while for the other speaker, velocity seems to be under control independently. It is hoped that such issues will be pursued with extensive data using many subjects in different languages.

In many languages, it has been reported that velum height for word initial position is higher than for word final position, segmentally (nearly) *ceteris paribus* [Ushijima *et al.* 1972; Fujimura 1977]. Some observation using my own articulation in Japanese shows that this initial vs. final distinction is observed for intrasyllabic position even when the nasal consonant is in word-medial position.

4.3.4. Allophonic Variation

One important issue that stems from the traditional segmental view of speech is the allophonic variation of the same phoneme depending on the context. Presumably, any universal effects of (hard) coarticulation are excluded from such descriptions of segmental variation, but that does not mean that the remaining aspects of coarticulatory processes do not involve utterance parameters. Parameters such as time constants of movement patterns, inherent strengths of influence over neighboring elements within an articultory dimension, susceptibility of a target position or a movement pattern to such influences, must vary from language to language, dialect to dialect, and part of it may well vary speaker to speaker. The patterns of use of particular articulators for the same phonological functions may also vary, as we have seen in Vaissiere's observation of velum movement patterns. Furthermore, parameters specifying a neutral (rest) position, range of movement, sensitivity to prosodic modulation, etc. of the articulators must be specified as to what we may call "phonetic disposition" to characterize each language, dialect, ideolect, etc. In order to compare different linguistic systems, in terms of phonological patterns implemented as speech, we need a complex and very sophisticated normalization method to be applied to different phonetic systems.

Precise descriptions of coarticulation and normalization processes are not known to us at present, but as a matter of principle, we may assume such well-defined processes which we could use to identify unexplained variation of phonetic values of phonological units, phonemes or syllables. A large part of such variation would be related to prosodic effects. There are known salient cases of phonetic variation of phonemic segments, however, which can be recognized as *ad hoc* in the sense that any language dependent assimilatory principle (*i. e.* even soft coarticulation) would not be expected to predict them[Fujimura & Lovins 1978]. I think most of such known allophonic variation is contained within the domain of the syllable, or in fact, the demisyllable.

My interest now is if we can find out some parts of such seemingly *ad hoc* variation to be describable in terms of a more general systematic (but of course language dependent) description of temporal characteristics of articulatory processes. I think the following observation of American English flapping may be suggestive of such a possibility.

In American English, intervocalic /t/ and /d/, typically in a stressed-reduced environment (as in 'better', see Kahn [1976]; Laferriere & Zue [1977]), are pronounced with a transient and incomplete closure accompanied by voicing for both /t/ and /d/ (tap or so-called flap, see Ladefoged [1977]). The microbeam observation with respect to the tongue blade pellet (about one cm behind the tip of the tongue) has revealed a very interesting dynamic characteristic of this articulatory gesture. Fig. 3 shows a comparison of a minimal contrast between a voiceless stop and a (voiced) tap for a pseudo-English phrase, spoken by a female speaker.

The two sets of time-functions representing coordinate values of pellets are aligned in time, in such a way that the two utterances show a fair agreement roughly, apart from the following two points [Birnbaum, personal communication]: (1) The mandible shows some raising for the stop gesture but not for tapping, and there is some tongue body movement for the stop correspondingly. (2) The tongue blade (presumably tip also) shows a distinctly different type of gesture both in the time function shape and the timing of the event as a whole relative to other articulators' temporal patterns.

Point one probably can be explained in terms of the difference in the use of the linguamandibular gesture related to both the phonological syllable margin status and the physical constraints or the physiological mechanism used for forming the apical closure. It should be emphasized that the time course of the mandible movement is practically identical except for the local difference directly reflecting the consonantal (or rather syllable margin) gesture.

Given this agreement in the timing programs, the salient difference in the blade movement is rather remarkable. In particular, the stop gesture occurs earlier and starts moving

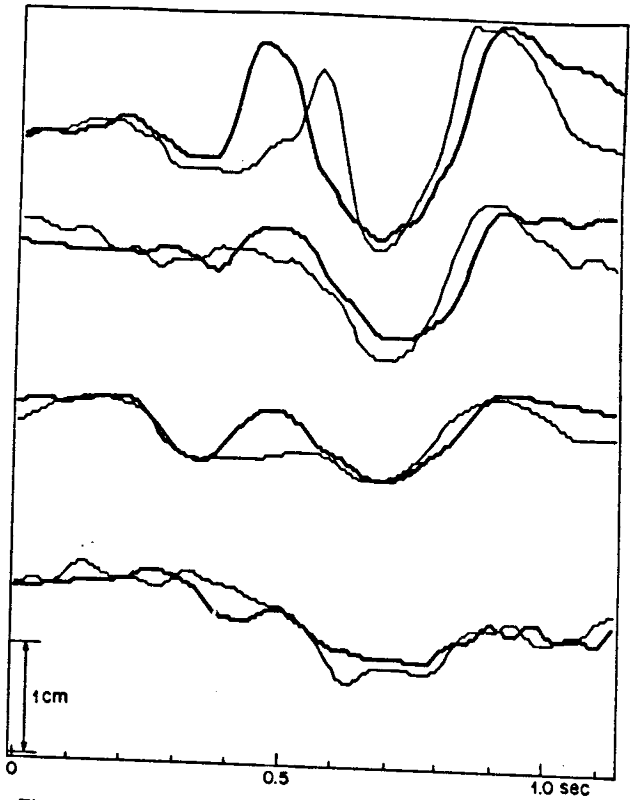


Fig. 3: Tongue blade movement (vertical position) comparing meaningless phrases 'bet aught' (thick lines) which was actually pronounced with a medial stop, and 'bed aught' (thin lines) pronounced with a medial tap by an American English speaker (female). The speaker was not given any instruction as to the manner of pronunciation, and the phrases were given in a list form. The curves represent, from top to bottom, tongue bade height, tongue dorsum height, mandible height, and tongue dorsum advancement, respectively. The two utterances are aligned in time to optimize the overall comparison for the curves except local deviations.

toward the vowel considerably earlier. The peak of the consonantal constriction (at the position of the pellet) occurs 100 msec or more earlier.

There are two categories of possible explanations. One is that the two gestures are, with respect to physiological realization, basically different. Different muscles are involved, perhaps associated with different neuromuscular and physical time constants, so that even the same time pattern of motor commands at the cortical level results in such a timing difference. Or perhaps the same muscle is used via different physiological cotrol mechanisms. Particularly, if the tapping gesture makes use of a more peripheral loop that elicits a response to some subtle change in a specific (perhaps unobserved) part of the articulator, as opposed to the stop gesture that is more or less under cortical control movement by movement, this rather qualitative difference might be plausible.

Another explanation may be that somehow the motor programming manifesting different syllable or foot structures has to be numerically different (perhaps only) in its timing configuration.

The former would suggest that the stop-flap (so-called) distinction is inherently discrete. The latter may imply that such allophonic variation is a continuous phenomenon; salient contrast evokes discretely or symbolically different perception or transcription, but depending on the context, particularly quantitatively specified phonetic parameters such as degree of emphasis or utterance speed, there may be intermediate cases from a phonetic point of view. The fact that some (even phonetically experienced) native speakers are not comfortable identifying the "stop-flap" distinction may suggest that the latter is the case. The recent study of formant characteristics of /l/-allophones by Sproat and Borowsky [1987] also seems to suggest the continuum of allophonic variation, refuting Halle and Mohanan's proposal [1985], and Borowsky's resyllabification theory basically seems promising in accounting for such phenomena.

4.4. An Elastic Model of Timing

As we have seen above, the temporal organization of speech for many purposes should be viewed as a multidimensional structure. If we obtain an effective descriptive system that uses approriate structural units for duration assignment, or events for defining timing and time intervals, then we may be able to represent in each of the dimensions the timing of each event by a linear model. That is, each time interval between contiguous events may be computed as a superposition of components due to different segmental and prosodic contributions as independent factors.

The idea of using a string of springs as a model of speech timing, or more specificallly of segmental durations, is not new. In particular, Jane Gaitenby [1965] at Haskins Laboratories discussed her data quite early using the concept of "elastic word" (see Lehiste [1980] for a review). What I would like to discuss here is a general model to describe the prosodic modulation of timing patterns of certain articulatory events. We can interpret time intervals among such events to derive durations of segmental units, whether phonemic, demisyllabic or syllabic.

After having determined the timing of each event in the time course of an utterance, we then will have to derive time functions of physical parameters, such as formant frequencies or tongue height, by looking up the inherent or segmental properties, static or dynamic as appropriate. As an example of such time function derivation processes, we may consider the case of pitch contours out of abstract tone specifications in Pierrehumbert's intonation work (see for its implementation as a synthesis rule system, Anderson *et al.* [1984]), or the prediction of velum movement patterns in Vaissiere's work discussed above.

Let us start with a simple example. Fig. 4 shows a single-dimension temporal model represented by a string of elastic springs. The length of each spring represents the time interval between two speech events to be observed, which are represented by joining points (circles) between congtiguous springs. The j-th spring is compressed or stretched deviating from its natural length x_{oj}, in response to the external force F. The extent of the response depends on the inherent stiffness k_j of the spring. Let us call the external force "prosodic force", because it is the cause of prosodic modulation in our model. The speed of utterance is directly related to the value of this prosodic force. Since the elastic system is superpositionally linear, all the increments of event intervals are proportional to the force. Also, I should emphasize here that the use of springs in the model does not imply uniform compression or stretching of the speech structure within the unit that is represented by a spring. Each spring represents only the interval between each adjacent pair of selected events.

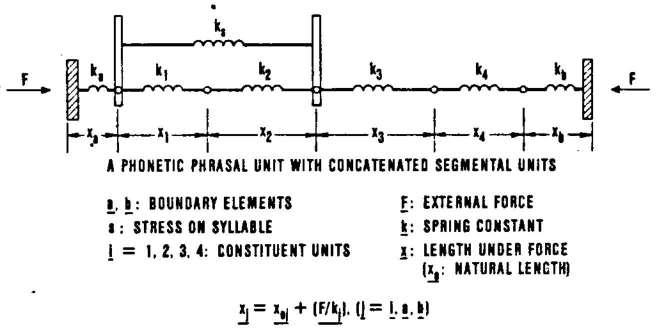


Fig. 4: A spring system composed for a simple phrasal unit (single dimension).

This model can be effective not only for representing the time interval distribution of an utterance as a whole, but also for the real time process of uttering a sentence, as far as the force is fixed throughout the utterance. This is so because in order to determine the conditions of the part of an utterance to the left (*i. e.* the past) of any joint point, we do not have to know the structure to its right (future). The boundary conditions for any substructure defined by the joint points at its ends are completely specified by the force applied to the end points. In this sense, considering the external force rather than the end positions as the boundary condition for the spring system is a crucial difference, even though mathematically the two ways of specifying boundary conditions are exactly equivalent.

We now need to devise a scheme to represent prosodic modulation. We would like to maintain that the quantities xo and k are inherent to the type of segment, or some projection of it onto a particular plane. We assume that the effect of stress is represented by a parallel spring, attached to the corresponding substring of segments representing a unit such as syllable, foot or word, to which such a stress specification is attached (see Fig. 4). This additional spring, which we may call a prosodic spring, has, as its inherent properties, natural length xo and stiffness k, representing the nature of the prosodic effect, such as the degree of stress. We may assume here that generally prosodic control is represented taking compression as the positive sense of the force. Thus a certain amount of external compression force is assumed for a neutral situation, and a relaxation or expansion of a segmental spring occurs when a parallel spring counterbalances part of the external force.

Another salient effect of prosodic modulation is the phrasal effect, in particular, phrase final lengthening. We represented in Fig. 4 such a boundary effect by adding virtual boundary springs, *a* and *b*, which are added to the segmental strings in series. These durational values are to be absorbed into the durational values of adjacent segmental units when we interpret the spring configuration as the temporal pattern of an utterance.

We can represent a hierarchical phrasal structure by embedding substructures in a larger spring system [Fujimura 1986d]. We need at least one level of phonetically motivated phrasal level to allow control of the prosodic status over the stretch of a syntactic phrase higher than words.

We studied this issue using icebergs (see *supra)* as the time marking events [Fujimura 1986a]. Fig. 5 illustrates a comparison of two utterances of the same word string, 'twenty two plus seven times four', distinguished by different phrasings corresponding to different arithmatic values. In this figure, each articulatory event (shown by a horizontal bracket) is plotted at the horizontal position representing its average timing, and at the vertical position representing the timing difference between the two utterances.

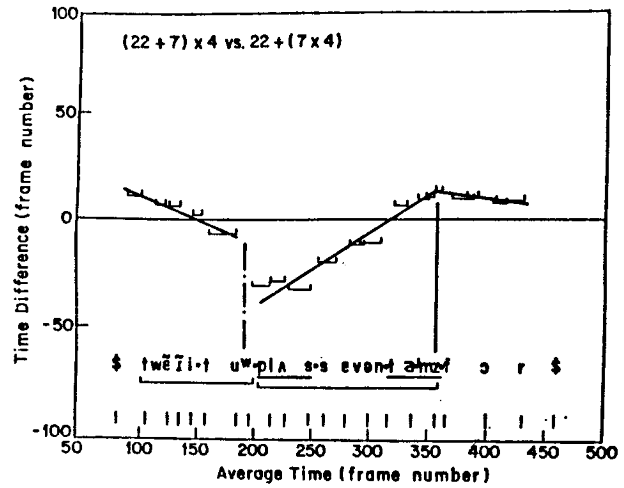


Fig. 5: A comparison of timings of corresponding events (iceberg-like movement patterns) in two utterances: along abscissa, average time for each event (shown by upward bracket), along ordinate, timing difference for each event between the two utterances. Note the time scales (in frame numbers) are different between the two axes. The frame interval is about 8 msec.

The pattern of timing difference demonstrates a piecewise linear change. This means that the ratio of the interval difference to the average time interval, *i. e.* the percentage of interval variation between the pair, was constant over each phrase but varied from one phrase to another. The local utterance speed, or equivalently the prosodic force, was varied in a manner representable by a parallel phrasal spring. Each breakpoint of straight lines indicates where a phrase boundary occurred in the sentence in question. Actually, the piecewise linear change as seen in this figure indicates that there are additional constraints imposed on the constants of the constituent springs, within the general model discussed above. At any event, this observation seems to hold for all other similar utterance pairs. Furthermore, when we compare a sentence with a contrastive emphasis placed on different words, as discussed above, similar piecewise linear patterns obtain. In such a case, the emphasized word behaves as a prosodic phrase.

In my opinion, the traditional acoustic events for timing measurements are quite useful and reliable, but do not reveal some of the important characteristics of temporal organization. I suspect strongly that events do not occur synchronously among different articulatory dimensions, as I have discussed in previous papers, because movement patterns in each dimension can reflect articulator specific temporal constraints. Some aspects of the asynchronism are probably crucial for understanding the basic nature of the phonological/phonetic structure of speech [Fujimura 1981,1986; Allwood & Scully 1982; Scully and Allwood 1985]. We need to measure articulatory events, as well as voicing control, for different articulators simultaneously. The x-ray microbeam provides us a good means to obtain useful and rather comprehenssive data.

As I mentioned before, the complex spring model is useful for describing the utterance process as long as the following two conditions are met:
(1) The prosodic force is not altered in
the middle of an integral utterance unit,

(2) The change of plan takes place
by modifying the substructure
that connects to the past part of the utterance
only through a single node.

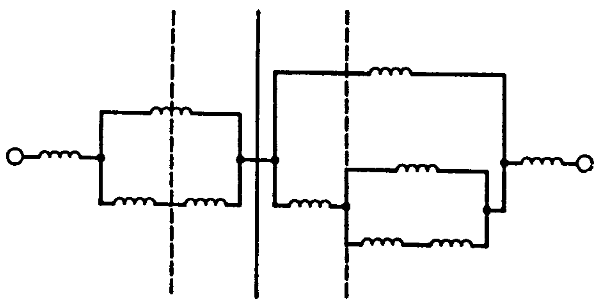


Fig. 6: A complex spring model (single dimension). Any of the solid vertical solidlines, but not broken lines, can mark a point in time as a boundary between past and future in motor programming.

Thus in Fig. 6, the solid vertical line can separate the past from the future, but broken vertical lines cannot. Time must jump from a single joint node to the next, as the motor program is sent for utterance execution. This suggests that the cognitive program controlling speech utterance is prepared as a sequence of phonetic phrasal units, formed into a simple concatenative linear string (*cf* Sternberg *et al.* [1978, 1980]). Within each phrasal unit, there can be any complex substructure involving parallel springs for prosodic modulations, but the specification of all such substructures within each phrasal unit must be complete before it gets started as an utterance.

Fig. 6 pertains only to one of the articulatory dimensions. The discussion above suggests also that the linkage among different dimensions must be solid at the phrase boundaries that serve for demarcating the motor program execution units.

5. Discourse and Intonation

In the tradition of linguistics, the sentence has played the most crucial role in the descriptive structure, defining the domain of syntax most successfully. Without delineating linguistic phenomena on the basis of the concept of sentence, the present achievement of descriptive theory could not have been imagined. This does not mean, however, that phenomena we encounter in speech research are all contained within the bounds of sentences. The more realistic we become in handling speech signals, the more severe we find the constraints. Recent research efforts have not overlooked these constraints. There are emerging findings which attempt at an ambitious challenge along this direction, even though, needless to say, it is very difficult to explore a rigorous theoretical approach, once we step out of the well-proved shelter of syntactic theory. These efforts are representatively characterized by a combination of training in artificial intelligence with the semantic, syntactic, phonological and phonetic as well as psychological disciplines. From the phonetic point of view, recent progress in intonation studies is particularly relevant, in many cases in connection with emerging thoughts about human linguistic performance and computational parsing of sentences.

Discussion of semantic references led AI researchers to the discovery of hierarchical block structures in the organization of discourse materials [Grosz 1977]. Recently, Hirschberg and Pierrehumbert [1986] (also Hirschberg [1987]) have shown that the voice pitch contours, when represented properly according to Pierrehumbert's descriptive framework, reveal the domains of such block structures. Silverman [1987] in his PhD dissertation corroborates this point, discussing related issues with extensive data from systematically controlled perceptual experiments. It is plausible that speech signals in actual conversation do carry more information than that represented in written text, in the form of pitch (and other signal aspects such as voice quality modulation as well as intensity), which significantly helps the listener in parsing the sentences correctly. Marcus and Hindle [1983] proposing their D-theory of syntactic descriiption argues that intonation breaks play a crucial role in sentence parsing, even though traditional orthographic systems ignore such information and necessitate for the readers of text a more complex parsing strategy.

6. Concluding Remarks

Speech is a physical and behavioral manifestation of linguistic structures. As such its characteristics can be evaluated only with reference to the linguistic structure that underlies it. While speech signals convey information other than linguistic codes, and the boundary between linguistic and extra- or para-lingnistic issues may not be clearcut, there is no question that the primary goal of speech research is to understand the relation between the units and organizations of linguistic forms to properties of speech signals that are uttered and perceived under different circumstances. For this goal to be achieved, it is imperative that we have an effective (probably not the only correct) theory and a feasible representation framework based thereon. In my opinion, we have not established the linguistic theory to satisfy this condition, even though we have seen remarkable progress in recent years in this field, and our understanding now is far better and more useful than it was a decade ago.

Furthermore, such theoretical endeavors must depend crucially on experimental and computational approaches, and is sensitive to the needs of industrial applications, just as, I might say, theoretical work in solid state physics is. Thus our work in speech synthesis from text, for example, can be affected immediately by any innovative development of the level theory [Kiparsky 1982] in morphology, nonlinear phonology, lexical semantics, syntactic subcategorization of verbs and so on, as well as discussions of temporal organization of articulatory events. On the other hand, there is no question that the theoretical discussion of the emerging tier/plane theory of phonological description must crucially depend on a rather accurate description of pitch contours, anatomy and physiology of articuratory systems, movement patterns of the tongue, the lips, the jaw, etc. along with a good linguistic insight and factual knowledge of a variety of natural languages, synchronic and diachronic. In addition, empirical results we obtain in engineering implementations of the theories do provide us with invaluable suggestions as to the future direction, as well as good motivations.

The more we learn about speech and language, the more strongly are we impressed by the depth of the human cognitive faculty.

7. Acknowledgment

I wish to acknowledge the benefit of daily discussions of relevant issues with all my colleagues in the Departments of Linguistics and of Artificial Intelligence Research, AT&T Bell Laboratories. In addition, Kim Silverman and David Talkin read through earlier versions of my manuscript and contributed many helpful comments. I am also grateful for many expositional corrections contributed by several readers, particularly Nina MacDonald and Joan Bachenko.

REFERENCES

Abbs, J.H.(1979), Speech Motor Equivalence: The Need for a Multi-Level Control, *Proc. 9th Int. Cong. Phonetic Sc., Vol II*, Copenhagen, Aug 6-11, 318-324.

Abbs, J.H. and Gracco, V.L.(1982), Motor Control of Multi-Movement Behaviors: Orofacial Muscle Responses to Load Perturbations of the Lips during Speech, *Soc. Neuroscience 8*, 282.

Abbs, J.H. and Gracco, V.L.(1983), Sensorimotor Actions in the Control of Multimovement Speech Gestures, in *Trends in Neuroscience 6*, 393-395.

Abbs, J.H. and Gracco, V.L.(in press), Control of Multimovement Coordination: Sensorimotor Mechanisms in Speech Motor Programming, *J. Motor Behavior*.

Allwood, E. and Scully, C.(1982), A Composite Model of Speech Production, *Proc. ICASSP '82, Vol. 2*, Piscataway N.J., IEEE Service Center, 932-935.

Anderson, M., Pierrehumbert, J.B. and Liberman, M.Y.(1984), Synthesis by Rule of English Intonation Patterns, *Proc. ICASSP '84, Vol. 1*, Piscataway N.J., IEEE Service Center,2.8.1-2.8.4.

Baer, T., Sasaki, C. and Harris, K.(eds.),(1987), *Laryngeal Function in Phonation and Respiration*, San Diego, College Hill Press.

Beckman, M.E. and Pierrehumbert, J.B.(1986), Intonational Structure in Japanese and English, in *Phonology Yearbook 3*, C. Ewen and E. Anderson(eds.), Cambridge, Cambridge U. Press,255-309.

Bell-Berti, F. and Harris, K.S.(1982), Temporal Patterns of Coarticulation: Lip Rounding, *J. Acoust. Soc. Am. 71*, 449-454.

Bless, D.M. and Abbs, J.(eds.)(1983), *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, San Diego, College Hill Press.

Blumstein, S. E. and Stevens, K. N.(1979), Acoustic Invariance in Speech Production: Evidence from Measurements of the Spectral Characterisitcs of Stop Consonants, *J. Acoust. Soc. Am. 66*, 1001-1017.

Borowsky, T.J.(1986), *Topics in the Lexical Phonology of English*, PhD Diss., Dept Linguistics, U. Massachussets, Amherst.

Browman, C.P.(1980), Rules for Demisyllable Synthesis Using LINGUA, a Language Interpreter, *Proc. ICASSP '80, Vol 2*, Piscataway N.J., IEEE Service Center,561-164.

Browman, C.P. and Goldstein, L.M.(1985), Dynamic Modeling of Phonetic Structure, in *Phonetic Linguistics -- Essays in Honor of Peter Ladefoged*, V. A. Fromkin(ed.), New York, Academic Press,35-53.

Chiba, T. and Kajiyama, M.(1941), *The Vowel, Its Nature and Structure*, Tokyo, Tokyo Kaiseikan.

Childers, D.G., Smith, A.M. and Moore, G.P.(1984), Relationships between Electroglottograph, Speech and Vocal Cord Contact, *Folia Phoniatrica 36*, 105-118.

Chomsky, N. and Halle, M.(1968), *The Sound Pattern of English*, New York, Harper & Row.

Coker, C.H.(1968), Speech Synthesis with a Parametric Articulatory Model in *Speech Symposium, Kyoto 1968*, reprinted in *Speech Synthesis*, J. L. Flanagan and L. R. Rabiner(eds.), Stroudsburg, Penn, Dowden-Hutchinson, Ross,135-139.

Conrad, W.(1987), Simplified One Mass Model with Supraglottal Resistance: A Testable Hypothesis, in *Laryngeal Function in Phonation and Respiration*, T. Baer, C. Sasaki and K. Harris(eds.), San Diego, College Hill Press, 320-338.

Cooper, F.S., Delattre, P., Liberman, A.M., Borst, J. and Gerstman, L. (1952), Some Experiments on the Perception of Speech Sounds, *J. Acoust. Soc. Am. 24*, 579-606.

Davis, P.J. and Nail, B.S. (to appear), The Sensitivity of Laryngeal Epithelial Receptors to Static and Dynamic Forms of Mechanical Stimulation, in *Voice Production*, O.Fujimura(ed.), New York, Rave Press.

Delattre, P., Liberman, A.M. and Cooper, F.S.(1955), Acoustic Loci and Transitional Cues for Consonants, *J. Acoust. Soc. Am. 27*, 769-773.

DeLuca, C.J.(1975), A Model for a Motor Unit Train Recorded during Constant Force Isometric Contractions, *Biol. Cybern. 19*, 159-167.

Dinnsen, D. A. and Charles-Luce, J.(1984), Phonological Neutralization, Phoneti. · Implementation and Individual Differences, *J. Phonetics 12*, 49-60.

Dixon, N.R. and Maxey, H.D.(1968), Terminal Analog Synthesis of Continuous Speech Using the Diphone Method of Segment Assembly, *IEEE Trans. Audio electroacoustics 16*, 40-50.

Edwards, J.(1985), *Mandibular Rotation and Translation during Speech*, PhD Diss., City U. of New York, Dept. Speech Hear. Scs.

Eek, A.(1973), Observations in Estonian Plalatalization: An Articulatory Study, *Estonian Papers in Phonetics*, 18-36.

Eek, A.(ed.)(1978), Studies on Accent, Quantity, Stress, Tone: Papers of the Symposium, Tallinn, Nov. 1978, *Estonian Papers in Phonetics*.

Eek, A. and Remmel, M.(1974), Context, Contacts and Duration: Two Results concerning Temporal Organization, *Preprints of the Speech Communication Seminar, Stockholm, Aug. 1-3, 1974*, 187-192.

Fant, G.(1956), On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies, in *For Roman Jakobson*, M. Halle, H. G. Lunt, H. McLean and C. C. van Schooneveld(eds.), The Hague, Mouton,109-120.

Fant, G.(1960), *Acoustic Theory of Speech Production*, The Hague, Mouton.

Fant, G.(1983), Preliminaries to Analysis of the Human Voice Source, *KTH-QPSR (1982), Royal Inst. Technol.*, Sweden,1-27.

Fant, G.(1983a), T... oice Source: Acoustic Modeling, *KTH-QPSR 4 (1982), R... Inst. Technol.* Sweden,28-48.

Fant, G., Ondračkova, J., Lindqvist, J. and Sonesson, B.(1966), Electrical Glottography, *KTH-QPSR 4 (1982), Royal Inst. Technol.* Sweden,15-21.

Firth, I. M.(1986), Modal Analysis of the Vocal Tract, *J. Acoust. Soc. Am. 80, Suppl 1*, S97.

Fisher, W., Doddington, G. and Goudie-Marshall, K.(1986), The DARPA Speech Recognition Database: Specifications and Status, *Proc. the DARPA Speech Recognition Workshops*, Feb. (1986), Washington, D. C., Science Applications International Corp.

Flanagan, J.L. and Landgraf, L.(1968), Self-Oscillating Source for Vocal- Tract Synthesizers, *IEEE Trans. Audio-Electronics*, 57-64.

Flanagan, J.L. and Ishizaka, K.(1976), Automatic Generation of Voiceless Excitation in a Vocal Cord/Vocal-Tract Speech Synthesizer, *IEEE Acoust. Signal Speech Processing 24*, 163-170.

Fletcher, S.G., McCutcheon, M.J. and Wolf, M.S.(1975), Dynamic Palatometry, *J Speech Hear. Res. 18*, 812-819.

Fourakis, M. (1984), Should Neutralization be Redefined?, *J. Phonetics 12*, 291-296.

Fourakis, M. and Iverson, G.K.(1984), On the 'Incomplete Neutralization' of German Final Obstruents, *Phonetica 41*, 140-149.

Fourcin, J. and Abberton, E.(1977), Laryngographic Studies of Vocal Cord Vibration, *Phonetica 34*, 313-315.

Frances, W.N. and Kučera, H.(1982), *Frequency Analysis of English Usage*, Boston, Houghton Mifflin Co.

Fromkin, V.A.(ed.)(1985), *Phonetic Linguistics, Essays in Honor of Peter Ladefoged*, Orlando, Florida, Academic Press.

Fujii, I., Fujimura, O. and Kagaya, R.(1971), Dynamic Palatography by use of a Computer and an Oscilloscope, *Proc. 7th Int. Cong. Acoust., Vol 3*, 113-116.

Fujimura, O.(1961), Bilabial Stop and Nasal Consonants: A Motion Picture Study and its Acoustical Implications, *J Speech Hear. Res. 4*, 233-2247.

Fujimura, O.(1970), Current Issues in Phonetics, *Studies in General and Oriental Linguistics*, R. Jakobson and S. Kawamoto(eds.), Tokyo, TEC Co., 109-130.

Fujimura, O.(1976), Syllable as Concatenated Demisyllables and Affixes, *Proc. Acoust. Soc. Am. 59, Suppl 1*, S55.

Fujimura, O.(1977), Recent Findings on Articulatory Processes -- Velum and Tongue Movements as Syllable Features, in *Articulatory Modeling and Phonetics*, R.Carre, R.Descout and M.Wajskop(eds.), Grenoble, GALF Groupe de la Communication Parlee,115-126.

Fujimura, O.(1979), Physiological Functions of the Larynx in Phonetic Control, *Current Issues in the Phonetic Sciences*, H. and P. Hollien(eds.), Amsterdam, John Benjamins,129-164.

Fujimura, O.(1980), Modern Methods of Investigation in Speech Production, *Phonetica 37*, 38-54.

Fujimura, O.(1981), Temporal Organization of Articulatory Movements as a Multidimensional Phrasal Structure, *Phonetica 38*, 66-83, corrected version in *Proc. Symp. Acoustic Phonetics and Speech Modeling, Part 2, Paper 5*, Inst. Defense Analysis, Princeton N.J..

Fujimura, O.(1981a), Body-Cover Theory of the Vocal Fold, in *Vocal Fold Physiology*, K. N. Stevens and M. Hirano(eds.), Tokyo, U. Tokyo Press,271-281.

Fujimura, O.(1984), The Role of Linguistics for Future Speech Technology, *Ling. Soc. Am. 104*, June 1984.

Fujimura, O.(1986), Relative Invariance of Articulatory Movements: An Iceberg Model, in *Invariance and Variability in Speech Processes*, J. S. Perkell and D. H. Klatt(eds.), Hillsdale N.J., Lawrence Erlbaum,226- 242.

Fujimura, O.(1986a), Temporal Organization of Articulatory Movements - A Multidimensional Complex Spring Model for Running Speech, *Proc. Acoust. Soc. Am. 80, Suppl 1*, S97.

Fujimura, O.(1986b), Evaluating the Task Dynamics Model, *J. Phonetics 14*, 105-108.

Fujimura, O.(1986c), A Model of Temporal Organization of Articulatory Gestures -- An X-ray Microbeam Study *Folia Phoniatrica 38*, 298, (Abst.).

Fujimura, O.(1986d), A Linear Model of Speech Timing, in *Ilse Lehiste Puhendusteos*, R.Channon and L.Shockey(eds.), Dordrecht, Holland, Foris.

Fujimura, O.(ed.)(to appear), *Voice Production - Proc. 5th Vocal Fold Physiology Conference Jan. 1987(tentative title)*, New York, Raven Press.

Fujimura, O. and Lindqvist, J.(1971), Sweep-Tone Measurements of Vocal Tract Characteristics, *J. Acoust. Soc. Am. 49*, 541-558.

Fujimura, O., Ishida, Y. and Kiritani, S.(1973), Computer-Controlled Radiography for Observation of Movements of Articulatory and other Human Organs, *Comp. Biology & Med. 3*, 371-384.

Fujimura, O., Tatsumi, I.F. and Kagaya, R.(1973a), Computational Processing of Palatographic Patterns, *J. Phonetics 1*, 47-54.

Fujimura, O., Miller, J.E. and Kiritani, S.(1977), A Computer-Controlled X-Ray Microbeam Study of Articulatory Characteristics of Nasal Consonants in English and Japanese, *Proc. 9th Int Congr. on Acoustics, Contributed Papers 1*, Madrid, 461.

Fujimura, O., Macchi, M.J. and Lovins, J.B.(1977a) Demisyllables and Affixes for Speech Synthesis, *Proc. 9th Int Congr. on Acoustics, Contributed Papers 1*, Madrid, 513.

Fujimura, O. and Lovins, J.(1978), Syllables as Concatenative Phonetic Units, in *Syllables and Segments*, A. Bell and J. B. Hopper(eds.), Amsterdam, North Holland, 107-120. An unabridged version available from Indiana U. Ling. Club.

Fujimura, O. and Kakita, Y.(1979), Remarks on Quantitative Description of the Lingual Articulation, in *Frontiers of Speech Communication Research*, S. Öhman and B. Lindblom(eds.), London, Academic Press,17-24.

Fujisaki, H.(1977), Functional Models of Articulatory and Phonatory Dynamics, in *Proc. US-Japan Sem. Dynamic Aspects of Speech Production*, M. Sawashima and F. S. Cooper(eds.), Tokyo, U. Tokyo Press,347-366.

Fujisaki, H.(1983), Dynamic Characteristics of Voice Fundamental Frequency in Speech and Singing, in *The Production of Speech*, P. F. MacNeilage(ed.), New York, Springer-Verlag, 39-56.

Gårding, E. (1983), A Generative Model of Intonation, in *Prosody: Models and Measurements*, A. Cutler and D. R. Ladd(eds.), New York, Springer-Verlag, 11-26.

Gaitenby, J.H.(1965), The Elastic Word, *Status Rep. Speech Res. 2*, New Haven, Haskins Laboratories, 3.1-3.11.

Grosz, B.J.(1977), *The Representation and Use of Focus in Dialogue Understanding*. Technical Rept 151, Menlo Park CA, SRI International.

Haken, H.(1977), *Synergetics* (Third Edition 1983), Heidelberg, Springer Verlag.

Halle, M.(1983), On Distinctive Features and their Articulatory Implementation, *Nat. Lang. Linguistic Theory 1*, 91-105.

Halle, M.(1985), Speculations about the Representation of Words in Memory, in *Phonetic Linguistics*, V. A. Fromkin(ed.), New York, Academic Press, 101-114.

Halle, M. and Mohanan, K.(1985), Segmental Phonology of Modern English, *Linguistic Inquiry 16*, 57-116.

Halle, M. and Vergnaud, J.-R.(1987), Stress and the Cycle, *Linguistic Inquiry 18*, 45-84.

Haraguchi, S.(1975), *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*, PhD Diss., Dept. Linguistics & Philosophy, MIT, Cambridge MA.

Hardcastle, W.(1972), The Use of Electropalatography in Phonetic Research, *Phonetica 25*, 192-215.

Hardcastle, W.J.(1974), Instrumental Investigation of Lingual Activity during Speech: A Survey, *Phonetica 29*, 129-157.

Hardcastle, W.(1984), New Methods of Profiling Lingual Palatal Contact Patterns with Electropalatography, *WP Phonetics Lab., U. Readings, No 4*, 1-40.

Hattori, S.(1961), Prosodeme, Syllable Structure and Laryngeal Phonemes, in *Studies in Descriptive and Applied Linguistics, Bulletin of the Summer Inst.Linguistics, 1*, Internat. Christian Univ., Tokyo, 1-27.

Hattori, S., Yamamoto, K. and Fujimura, O.(1958), Nasalization of Vowels in Relation to Nasals, *J. Acoust. Soc. Am. 30*, 267-274.

Henke, W.L.(1966), *Dynamic Articulatory Model of Speech Production Using Computer Simulation*, PhD. Diss., Dept. Linguistics & Philosophy, MIT, Cambridge, Mass.

Hirano, M.(1977), Structure and Vibratory Behavior of the Vocal Folds, in *Dynamic Aspects of Speech Production*, M. Sawashima and F. S. Cooper(eds.), Tokyo, U. Tokyo Press.

Hirano, M. and Ohala, J.(1969), Use of Hooked-Wire Electrodes for Electromyography of the Intrinsic Laryngeal Muscles, *J. Speech Hear. Res. 12*, 362-373.

Hirose, H. and Kiritani, S.(1985), A Kinesiological Study of Labial Articulatory Movements in Ataxic Patients, *Ann. Bull. U. Tokyo RILP 19*, 201-208.

Hirschberg, J. (1987), Uses of Intonational Cues in Discourse Studies, *Proc. TINLAP-87*, 86-91.

Hirschberg, J. and Pierrehumbert, J.B.(1986), The Intonational Structuring of Discourse, *Proc. 24th Meeting Ass. Comp. Linguistics*, New York, Columbia Univ.,136-144.

Holmes, J.N. and Mattingly, I.G.(1964), Speech Synthesis by Rule, *Language & Speech 7*, 127-143.

Honda, K.(1983), Variability Analysis of Laryngeal Muscle Activity, in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, I.Titze and R.Scherer(eds.), Iowa, Iowa U. Press,127-137.

Honda, K., Kiritani, S., Imagawa, J. and Hirose, H.(1985), High-Speed Digital Recording of Vocal Fold Vibration Using a Solid-State Sensor, *Ann. Bull. U. Tokyo RILP 19*, 47-53.

Horiguchi, S. and Bell-Berti, F.(1984), The Velotrace: A Mechanical Device for Tracking Velar Position, *Paper Presented at the Meeting of American Cleft Palate Assoc.*, Seatle, WA.

Houde, R.(1967), *A Study of Tongue Body Motion During Selected Speech Sounds*, PhD Diss., U. Michigan, *Speech Comm. Res. Lab. Monograph 2 (1968)*, Los Angeles.

House, A.S. and Stevens, K.N.(1956), Analog Studies of the Nasalization of Vowels, *J. Speech Hear. Dis. 21*, 218-232.

Hughes, O.M. and Abbs, J.H.(1976), Labio-Mandibular Coordination in the Production of Speech: Implications for the Operator of Motor Equivalence, *Phonetica 33*,, 199-221.

Inkelas, S., Leben, W. and Cobler, M.(1987), The Phonology of Intonation in Hausa, *NELS 17*, North-Eastern Linguistic Society.

Ishizaka, K. and Matsudaira, M.(1968), What Makes the Vocal Cords Vibrate?, *Proc. 6th Internat. Cong. Acoustics, Vol II*, Y. Kohasi(ed.), New York, Elsevier,9-12.

Jakobson, R., Fant, C.G. and Halle, M.(1951), *Preliminaries to Speech Analysis*, Cambridge, MIT Press (Third Edition 1963).

Kahn, D. (1976) *Syllable-Based Generalizations in English Phonology*, PhD Diss., Dept. Linguistics & Philosophy, MIT, Cambridge, Mass. New York, Garland Publishing, Inc. (1980).

Kakita, Y., Hirano, M. and Ohmaru, K.(1981), Physical Properties of the Vocal Fold Tissue: Measurements on Excised larynges, *Vocal Fold Physiology*, K. N. Stevens and M. Hirano(eds.), Tokyo, U. Tokyo Press, 377-397.

Kakita, Y., Fujimura, O. and Honda, K.(1985), Computation of Mapping from Muscular Contraction Patterns to Formant Patterns in Vowel Space, in *Phonetic Linguistics*, V. A. Fromkin(ed.), New York, Academic Press,133-144.

Kaneko, T., Uchida, K.,Suzuki, H., Komatsu, K., Kanesaka, T., Kobayashi, N. and Naito, J.(1981), Ultrasonic Observations of Vocal Fold Vibration, in *Vocal Fold Physiology*, K. N. Stevens and M. Hirano(eds.), Tokyo, Univ.Tokyo Press,107-117.

Kean, M. (1975), *The Theory of Markedness in Generative Grammar*, PhD Diss., Dept. Linguistics & Philosophy, MIT, Cambridge MA.

Keating, P. A.(1985), Universal Phonetics and the Organization of Grammars, in *Phonetic Linguistics*, V. A. Fromkin(ed.), New York, Academic Press, 115-132.

Kelso J.A.S., Saltzman E.L. and Tuller, B.(1986), The Dynamical Perspective on Speech Production, *J. Phonetics 14*, 29-59.

Kelso, J.A.S., Saltzman, E.L. and Tuller, B.(1986a), Intentional Contents, Communicative Context, and Task Dynamics, a Reply to Commentators, *J. Phonetics 14*, 171-196.

Kent, R.D. (1983), The Segmental Organization of Speech, in *The Production of Speech*, P. F. MacNeilage(ed.), New York, Springer-Verlag, 57-90.

Kiparsky, P.(1982), From Cyclic Phonology to Lexical Phonology, *The Structure of Phonological Representation, Part I*, H. van der Hulst and N. Smith(eds.), Dordrecht, Holland, Foris Publication,131-176.

Kirikae, I.(1943), A Study on the Vibration of the Human Vocal Cords in Phonation and the Timing Relations of the Glottal Opening-Closure by the Use of a Laryngeal Stroboscopic Motion Picture Technique, *J. Japan Otorhinolaryngology 49*, 236-268.

Kiritani, S., Itoh, K. and Fujimura, O.(1975), Tongue-Pellet Tracking by a Computer Controlled X-Ray Microbeam System, *J. Acoust. Soc. Am. 57*, 1516-1520.

Kiritani, S., Miyawaki, K. Fujimura, O. and Miller, J. E.(1976), A Computational Model of the Tongue, *Ann. Bull. U. Tokyo RILP 10*, 243-251.

Kiritani, S., Tateno, Y. and Iinuma, T.(1977), Computer Tomography of the Vocal Tract, in *Dynamic Aspects of Speech Production*, M. Sawashima and F. S. Cooper(eds.), Tokyo, U. Tokyo Press, 203-208.

Kiritani, S. Imagawa, H. and Hirose, H.(to Appear), High-Speech Digital Image Recording for the Observation of Vocal Cord Vibration, in O.Fujimura(ed.), *Voice Production*, Raven Press.

Klatt, D.H.(1978), Synthesis by Rule of Consonant-Vowel Syllables, *Proc. Acoust. Soc. Am. 64, Suppl 1*, S114.

Klatt, D.H.(1979), Synthesis by Rule of Segmental Durations in English Sentences, in *Frontiers of Speech Communication Research*, B. Lindblom and S. Ohman(eds.), New York, Academic Press, 287-300.

Kohler, K J. and van Dommelen, W. A. (1986), Prosodic Effects on Lenis/Fortis Perception: Preplosive F0 and LPC Synthesis, *Phonetica 43*, 70-75.

Kozhevnikov, V.A. and Chistovic, L.A.(1965), *Rech: Artikulyatsia I Vospriyatiye*, Nauka, Moscow, Leningrad.

Kupin, J. J.(1979), *Tongue Twisters as a Source of Information about Speech Production*, PhD Diss., Storrs CT, U. Conn.

Ladd, R.(1986), Intonational Phrasing: the Case for Recursive Prosodic Structure, *Phonology Yearbook 3*, C. Ewen and E. Anderson (eds.), Cambridge; Cambridge U. Press, 311-340.

Ladefoged, P., Cochran, A. and Disner, S.(1977), Laterals and Trills, *J. Int. Phonetic Assoc. 7*, 46-54.

Laferriere, M. and Zue, V.W,(1977), Flapping Rule in American English: An Acoustical Study, *J. Acoust. Soc. Am. 61, Suppl 1* S31.

Lassen, N.A. and Larsen, B.(1980), Cortical Activity in Left and Right Hemisphere during Language Related Brain Functions, *Proc. 9th Int. Cong. Phonetics Sci., Vol III*, Copenhagen, Aug 6-11, 137-150.

Lehiste, I. (1970), *Suprasegmentals*, Cambridge, Mass, MIT Press.

Lehiste, I.(1980), Phonetic Manifestation of Syntactic Structure in English, *Ann. Bull. U. Tokyo RILP 14* 1-28.

Liberman, A.M., Delattre, P.C., Cooper, F.S. and Gerstman, L. J.(1954), The Role of Consonant-Vowel Transitions in the Perception of Stop and Nasal Consonants, *Psychology Monographs 68*, 1-13.

Liberman, A.M., Ingemann, F., Lisker, L., Delattre, P. and Cooper, F.S.(1959), Minimal Rules for Synthesing Speech, *J. Acoust. Soc. Am. 31*, 1490-1499.

Liberman, M.Y.(1975) *The Intonational System of English*, PhD Diss., Dept. Linguistics & Philosophy, MIT, Cambridge, Mass.

Liberman, M.Y. and Prince, A.(1977), On Stress and Linguistic Rhythm, *Linguistic Inquiry 8*, 249-336.

Lindblom, B.(1964), Dynamic Aspects of Vowel Articulation, *Proc. 5th Int. Cong. Phonetic Science*, Munster, 387-388.

Lindblom, B.(1968), *On the Production and Recognition of Vowels*, PhD Diss., Lund Univ.

Lindblom, B.(1983), Economy of Speech Gestures, in *The Production of Speech*, P. F. MacNeilage(ed.), New York, Springer-Verlag,217-246.

Lindblom, B., Lubker, J. and Gay, T.(1979), Formant Frequencies of some Fixed-Mandible Vowels and a Model of Speech Motor Programming by Predictive Simulation, *J. Phonetics 7*, 147-161.

Macchi, M.J.(1980), A Phonetic Dictionary for Demisyllable Speech Synthesis, *Proc. ICASSP '80, Vol 2*, Piscataway NJ, IEEE Service Center, 565-567.

Macchi, M.J.(1985), *Segmental and Suprasegmental Features and Lip and Jaw Articulators*, PhD Diss., Dept. Linguistics, New York Univ.

MacNeilage, P.F.(ed.) (1983), *The Production of Speech*, New York, Springer-Verlag.

MacNeilage, P.F.(1985), Serial-Ordering Errors in Speech and Typing,in *Phonetic Linguistics*, V.A.Fromkin(ed.), Orlando, Florida, Academic Press.

Maeda, S.(1976), *A Characterization of American English Intonation*, PhD Diss., MIT, Cambridge MA.

Maeda, S.(1983), *Correlats Acoustiques de la Nasalization des Voyelles: Une Étude de Simulation*, Centre National D'études des Télecommunications, Centre des Recherches de Lannion, Rept. VII.

Marcus, M. and Hindle, D.(1983), D-Theory: Talking about Talking about Trees, *Proc. 21st Annual Meeting of the Ass. Comp. Linguistics*.

McCawley, J.D.(1968), *The Phonological Component of a Grammar of Japanese*, The Hague, Mouton.

Mermelstein, P.(1973), Articulatory Model for the Study of Speech Production, *J. Acoust. Soc. Am. 53*, 1070-1082.

Miyawaki, K.(1972), *A Study of Lingual Articulation by Use of Dynamic Palatography*, M.A. Thesis, Dept. Linguistics, U. Tokyo.

Müller, E., Abbs, J, Kennedy, J. and Larson, C. (1977), Significance of Perioral Biomechanics to Lip Movements during Speech, *Am. Speech Lang. Hear. Assoc.*

Nadler, R.D., Abbs, J.H., Fujimura, O.(1987), Speech Movement Research Using the New X-Ray Microbeam System, *Proc. 11th International Congress of Phonetic Sciences, Tallinn*, Paper Se 11.4.

Nelson, W.L.(1983), Physical Principles for Economies of Skilled Movements, *Biol. Cybernetics 46*, 135-147.

Nishinuma, Y. and Rossi, M.(1981), Automatisation of Prosodic Analysis in French, *Study of Sounds (The Phonetic Society of Japan) XIX*, 155-169.

Nooteboom, S. G. and Terken, J. M. B.(1982), What Makes Speakers Omit Pitch Accents? An Experiment, *Phonetica 39*, 317-336.

Ohala, M.(1983), The Machine as an Addressee: When Paralinguistics Fails, *Abstracts of the Tenth International Congress of Phonetic Sciences*, Dordrecht, Foris Publication,428.

Öhman, S.E.G. (1967), Numerical Model of Coarticulation, *J. Acoust. Soc. Am. 41*, 310-320.

Oka, D.K.(1980), *The Design and Test of a Ranging Transducer to Monitor Articulatory Movement during Speech Production*, M.S. Thesis, MIT.

Olive, J.P.(1980), A Scheme of Concatenating Units for Speech Synthesis, *Proc. ICASSP '80, Vol 2*, Piscataway N.J., IEEE Service Center,568-571.

Ostry, D. J., Keller, E. and Parush, A.(1983), Similarities in the Control of Speech Articulators and Limbs: Kinematics of Tongue Dorsum Movement in Speech, *J. Exp. Psychology: Human Percept. & Perf. 9*, 622-636.

Perkell, J.S.(1969), *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Analysis*, Res. Monograph No 53, Cambridge, Mass, MIT Press.

Perkell, J.S. and Nelson, W.L.(1982), Articulatory Targets and Speech Motor Control: A Study of Vowel Production, in *Speech Motor Control*, S. Grillner, B. Lindblom, J. Lubker, A. Persson(eds.), New York, Pergamon, 187-204.

Perkell, J. and Klatt, D.(eds.)(1986), *Invariance and Variability in Speech Processes*, Hillsdale NJ, Lawrence Erlbaum.

Pierrehumbert, J.(1980), *The Phonology and Phonetics of English Intonation*, PhD Diss., Dept. Linguistics & Philosophy, MIT, Cambridge MA.

Pierrehumbert, J. and Beckman, M.(in press), Japanese Tone Structure, *Linguistic Inquiry*.

Port, R.F. and O'Dell, M.L.(1985), Neutralization of Syllable-Final Voicing in German, *J. Phonetics 13*, 455-471.

Poser, J.P.(1984), *The Phonetics and Phonology of Tone and Intonation in Japanese*, PhD Diss., Dept. Linguistics & Philosophy, MIT, Cambridge MA.

Rome, J.A.(1964), An Artificial Pellet for Continuous Analysis of Speech, *MIT Res. Lab. Eletronics, Quart. Prog. Rep. 74*, 190-191.

Rosenberg, E.A., Rabiner, R.L., Wilpon, G.J. and Kahn, D.(1983), Demisyllable-Based Isolated Word Recognition System, *IEEE Trans. Acoust. Speech & Signal Process 31*, 713.

Rossi, M. and Autesserre, D. (1981), Movements of the Hyoid and the Larynx and the Intrinsic Frequency of Vowels, *J. Phonetics 9*, 233-249.

Rothenberg, M.(1981), Acoustic Interaction between the Glottal Source and the Vocal Tract, in *Vocal Fold Physiology*, K. N. Stevens and M. Hirano(eds.), Tokyo, U. Tokyo Press, 305-328.

Saito, S., Fukuda, H., Isogai, Y. and Ono, H.(1981), X-ray Stroboscopy, in *Vocal Fold Physiology*, K. N. Stevens and M. Hirano(eds.), Tokyo, U. Tokyo Press, 95-106.

Sawashima, M. and Hirose, H.(1968), New Laryngoscopic Technique by Use of Fiber Optics, *J. Acoust. Soc. Am. 43*, 168-169.

Sawashima, M.(1976), Fiberoptic Observation of the Larynx and other Speech Organs, *Proc. US-JAP Sem. Dynamic Asp. Speech Production*, Tokyo, U. Tokyo Press, 31-46.

Sawashima, M. and Kiritani, S.(1985), Electro-Palatographic Patterns of Japanese /d/ and /r/ in Intervocalic Position, *Ann. Bull. U. Tokyo RILP 19*, 1-6.

Schönle, P.W., Wenig, P., Schrader, J., Grabe, K., Brockmann, E. and Conrad, B. (1983), Ein Elektromagnetisches Verfahren zur Simultanen Registrierung von Bewegungen im Bereich des Lippen-, Unterkiefer- und Zungensystems, *Biomed. Technik 28*, 263-267.

Scully, C. and Allwood, E. (1985), Production and Perception of an Articulatory Continuum for Fricatives in English, *Speech Communications 4*, 237-246.

Silverman, K.(1987), *The Structure and Processing of Fundamental Frequency Contours*, PhD Diss., Cambridge Univ.

Sivertsen, E.(1961), Segment Inventories for Speech Synthesis, *Language & Speech 4*, 27-89.

Shadle, C.(1985), *The Acoustics of Fricative Consonants*, PhD Diss., MIT.

Shibata, S., Ino, A. and Yamashita, S. (1979), *Teaching Articulation by Use of Electro-Palatography* (English translation 1982 available), Kokubunji, Tokyo, Rion Co., Ltd.

Smith, S.(1981), Research on the Principle of Electroglottography, *Folia Phoniatrica 33*, 105-114.

Sonies, B.C., Shawker, T.H., Hall, T.E., Gerber, L.H. and Leighton, S.B.(1981), Ultrasonic Visualization of Tongue Motion during Speech, *J. Acoust. Soc. Am. 70*, 683-686

Sonoda, Y. and Kiritani, S.(1976), Analysis of Tongue Point Movements by a Linear Second-Order System Model, *U. Tokyo, Ann. Bul. RILP 10*, 29-36.

Sproat, R. and Borowsky, T.(1987), On the Resyllabification of /l/ in English, *J. Acoust. Soc. Am. 81, Suppl. 1*, S67.

Stevens, K.N.(1960), Towards a Model for Speech Recognition, *J. Acoust. Soc. Am. 32*, 47-55.

Stevens, K.N.(1972), The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data, in *Human Communication: A Unified View*, E. E. David and P. B. Denes(eds.), New york, McGraw-Hill, 51-66.

Stevens, K.N.(1975), Physics of Laryngeal Behavior and Larynx Modes, *Phonetica 34*, 264-279.

Stevens, K.N.(1983), Design Features of Speech Sound Systems, in *The Production of Speech*, P. F. MacNeilage(ed.), New York, Springer Verlag,247-262.

Stevens, K.N.(to appear), Modes of Vocal Fold Vibration Based on a Two-Section Model, in *Voice Production*, O.Fujimura(ed.) New York, Raven Press.

Stevens, K.N. and Hirano, M.(eds.)(1981), *Vocal Fold Physiology* Tokyo, U. Tokyo Press.

Sternberg, S., Monsell, S., Knoll, R. and Wright, C.(1978), The Latency and Duration of Rapid Movement Sequences: Comparison of Speech and Typewriting, in *Information Processing in Motor Control and Learning*, G. Stelmach(ed.), New york, Academic Press.

Sternberg, S., Wright, C.E., Knowll, R.L. and Monsell, S.(1980), Motor Programs in Rapid Speech, Additional Evidence, in *The Perception and Production of Fluent Speech*, R. A. Cole(ed.), Hillsdale NJ, Lawrence Erlbaum, 507-534.

Teager, H.M.(1983), The Effects of Separated Air Flow on Vocalization, in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, D. M. Bless and J. H. Abbs(eds.), San Diego, College-Hill Press, 124-141.

Thorsen, N.(1984), Variability and Invariance in Danish Stress Group Patterns, *Phonetica 41*, 88-102.

Thomas, T.J.(1985), *An Articulatory Model of Speech Production Including Turbulence*, PhD Diss., Cambridge Univ.

Titze, I.R.(1985), Mechanisms of Sustained Oscillation of the Vocal Folds, in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, I. R. Titze and R. C. Scherer(eds.), Denver, The Denver Center for the Performing Arts, 349-357.

Titze, I.R. and Talkin, D.T.(1979), A Theoretical Study of the Effects of Various Laryngeal Configurations on the Acoustics of Phonation, *J. Acoust. Soc. Am. 66*, 60-74.

Titze, I. and Scherer, R.(eds.)(1983), *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, Denver, Denver Center of Perform Arts.

Titze, I.R. and Durham, P.L.(1987), Passive Mechanism Influencing Fundamental Frequency Control, in *Laryngeal Function in Phonation and Respiration*, T. Baer, C. Sasaki and K. S. Harris(eds.), San Diego, College Hill Press, 304-319.

Umeda, N.(1975), Vowel Duration in American English, *J. Acoust. Soc. Am. 58*, 434-445.

Ushijima, T. and Sawashima, M.(1972), Fiberoptic Observation of Velar Movements during Speech, *Ann. Bull. U. Tokyo RILP 6*, 25-38.

Vaissiere, J.(1977), Quelques Experiences d'Analyse Perceptuelle en Français, *VIIIèmes Journées d'Étude sur la Parole, Aix-en-Provence, 25-27 Mai 1977*, 183-189.

Van den Berg, J.(1957), Sub-Glottal Pressure and Vibrations of the Vocal Folds, *Folia Phoniatrica 9*, 65-71.

Wood, S.(1979), A Radiographic Analysis of Constriction Locations for Vowels, *J. Phonetics 7*, 24-44.

Zealear, D.(1987), The Brainstem Connections with the Laryngeal Region of the Motor Cortex in the Monkey, in *Laryngeal Function in Phonation and Respiration*, T. Baer, C. Sasaki, K. S. Harris(eds.), San Diego, College Hill Press, 168-177.

GLOTTAL CONTROL OF ASPIRATION AND OF VOICELESSNESS

AILBHE NI CHASAIDE

CENTRE FOR LANGUAGE AND COMMUNICATION STUDIES,
TRINITY COLLEGE, DUBLIN 2, EIRE

ABSTRACT

There are two different, though potentially complementary explanations of how glottal control of voicelessness and aspiration is effected; a) by directly controlling the degree of glottal aperture during stop closure and specifically at oral release; b) by controlling the precise timing of the abduction/adduction gesture.

Photoelectric glottographic data for aspirated (pre- and post-) and unaspirated stops in Icelandic and Irish uttered in differing stress conditions suggest that ; glottal aperture is directly controlled; this and other "strategies" observed may be necessary to maintain voice offset and onset targets under differing aerodynamic conditions; such strategies might be best regarded, not as correlates of stress as such, but rather as evidence of a more general laryngeal response to changes in aerodynamic conditions.

INTRODUCTION

Two rather different proposals have been made regarding the laryngeal mechanism which determines the presence and duration of aspiration. Kim /1/ has hypothesised that the degree of glottal opening during stop closure, and crucially at stop release determines the duration of aspiration. The wider the glottal aperture at the instant of release, the longer the aspiration that ensues; if the vocal folds are already adducted at that point, there will be no aspiration. Working from Kim's data, Catford /2/ gives an approximate graph showing the relationship between glottal aperture during stop closure and aspiration duration.

More recently, Löfqvist et al. /3/ have questioned an underlying assumption of the earlier work, namely, that speakers directly control the degree of glottal opening as a means of determining the duration of aspiration. These authors suggest that "voluntary control of the size of glottal opening is rather poor and that subjects are unable to make very fine graded adjustments along this dimension." Löfqvist /4/ also points out that the glottal gesture is a relatively fixed ballistic opening/closing cycle; once peak glottal opening has been attained, the closing gesture tends to start immediately rather than maintain a static open position. Thus, rather than direct aperture control, Löfqvist proposes that control is exerted on the timing of the laryngeal gesture, to which Peak glottal opening (PGO) is an important index. The later the PGO, the greater the glottal abduction at stop release, and the longer the aspiration. The converse should be true for preaspirated stops. It follows from Löfqvist's account that, all else being equal, one might expect greater peak glottal opening for aspirated than for an unaspirated stop: such a difference would however be a secondary consequence rather than the primary control parameter.

It was felt that voiceless stops in Icelandic and Irish (yielding pre-, post- and unasapirated types) across differing stress conditions might provide a testing ground for these two models of aspiration control. The durations of preaspiration /5/ and postaspiration /6/ can be much shorter in unstressed than in stressed syllables, and it would be of interest to consider which of the above two models might best account for those differences.

MATERIALS

Recordings of four short data sets in Icelandic and Irish (a single subject in either case) included the following signals: photo-electric glottograph (PEG), oral airflow, and audio. Further recordings of data sets 1,2, and 3 were made, where subglottal pressure (strictly speaking, oesophageal pressure) was substituted for PEG. For details on equipment used see /5/ and references therein. The first data set (Icelandic, 66 tokens) contained the three possible bilabial stops in VCV, as exemplified in the words [laʰpa], [la:pa] and [s̥ka:pa]. Each of these was inserted into carrier frames so that the word in alternate sentences did and did not receive the main sentence stress: "Hann sagði -- við mig." and "Hann sagði ekki -- við mig." The three further sets involved Irish utterances containing the voiceless dental stop in VCV and #CV. For set two (40 tokens), the word [baʰtʰa] was simply inserted into the frame: "Dúirt sé -- liom." which was repeated so that in alternate repetitions the word received either normal sentence stress or emphatic stress. In set three (48 tokens), a further frame was added: "Dúirt sé -- beag liom." As sentence stress in this last frame falls on beag, the word [baʰtʰa] is in the relatively unstressed (prenuclear) position. The intention here was to elicit three stress levels; emphatic, normal and a (relative) lack of stress. In the fourth set (24 tokens), the word [tʰa:] was repeated with alternating emphatic and normal sentence stress in the frame: "Tá, adúirt sé."

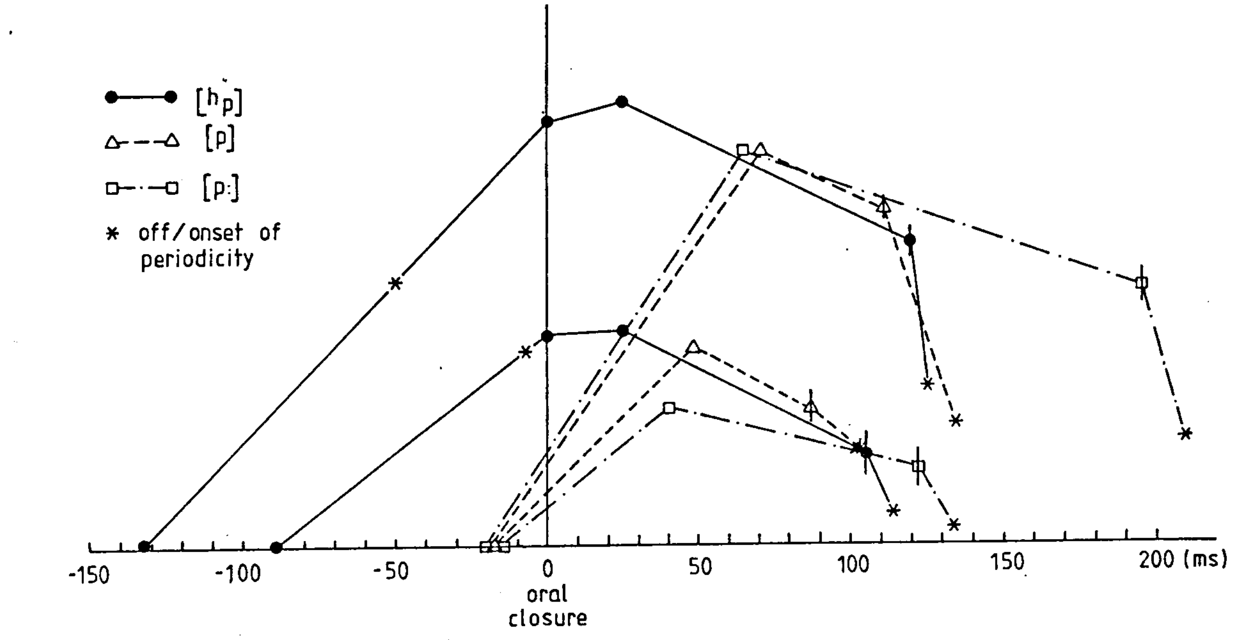


Fig 1. Averaged PEG traces for Icelandic medial stops for stressed (higher peaks) and unstressed (lower peaks) environments.

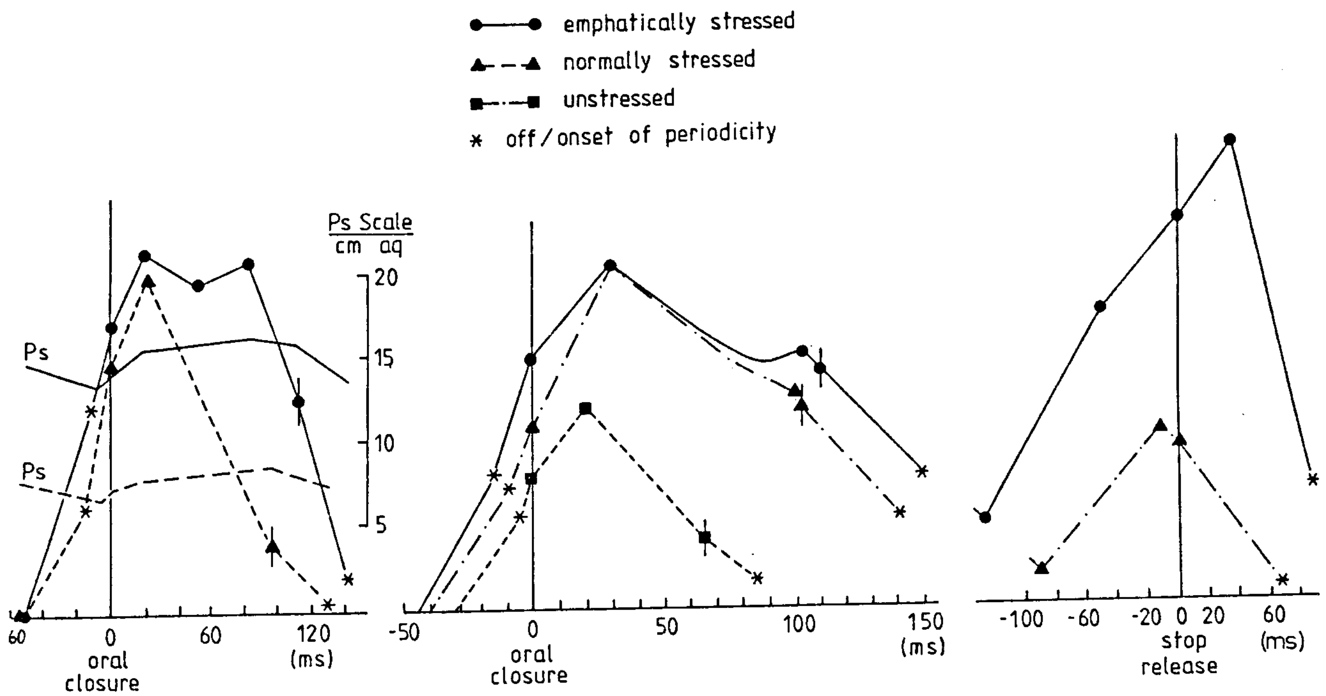


Fig 2. Averaged PEG for -ʰtʰ- in Irish, for emphatic v. normal stress. Ps values referred to in text.

Fig 3. Averaged PEG for Irish -ʰtʰ- in emphatic, normal and normal stress environments.

Fig 4. Averaged PEG for Irish #tʰ- in emphatic v. normal stress.

The comparative amplitude of the PEG waveform was measured at the following points: during the preceding vowel (except for stops in #CV) a point taken as the baseline from which the PEG amplitude was measured; at stop closure (for the preaspirated stops); at peak glottal opening; at stop release, and at voice onset. The time intervals between these points were measured, as were the durations of post- and preaspiration. Also measured for the latter, was the time taken to effect devoicing from the start of vocal fold abduction.

RESULTS

Figures 1,2,3 and 4 display the averaged PEG results. In Fig 1, where stressed and unstressed tokens of the three types of voiceless stops in Icelandic are compared, the amplitude of the peak glottal opening (PGO) would appear to be considerably greater for the stressed. And although the duration of preaspiration is a good deal shorter in unstressed than in stressed [ʰp], this is not reflected in the timing of PGO. Postaspiration values do not differ in the stressed and unstressed conditions, a fact which might appear surprising at first glance. Values are uniformly very short (ranging from 10ms to 35ms, and so, postaspiration is likely to be barely if at all perceptible). Also note that offset and onset of periodicity would seem to occur at greater amplitude of glottal opening for the stressed cases. In Fig 3, where Irish unstressed tokens are compared with normal and emphatic, PGO amplitude seems again much lower in the unstressed.

The difference between emphatically and normally stressed tokens (Irish) is not uniformly reflected by a single change in laryngeal behaviour. In #CV, there is not only a large differencein the amplitude of PGO but also in its timing. The peak occurs either before or after stop release, depending respectively on whether the stop occurs in the normally or emphatically stressed syllable. The difference in postaspiration duration is likely to be due to both timing and amplitude effects. The medial stops are rather different (note that these have both some pre- and postaspiration). When only emphatic and normal were contrasted (data set 2, Fig 2) there would appear to be little difference in the amplitude of glottal opening, but a striking difference in the glottal gesture itself. Emphatic tokens are characterized by a second glottal opening peak, the first of which corresponds in timing to that of the normally stressed tokens. The durations of pre- and postaspiration are virtually the same for both stress conditions. Note that the time taken to effect voicelessness is about the same also, but that the amplitude of glottal opening at which voice offsets is considerably greater with emphatic stress; the rate of glottal opening may therefore be quicker. As a consequence of the second opening peak, the amplitude of glottal opening at stop release also appears to be greater for these stops. In Fig 3 (data set 3, containing additionally unstressed tokens) the difference between the emphatic and normal is less striking, although there is again clear evidence of double peaking in the emphatic. It is likely that speakers were differentiating less consistantly between the emphatically and normally stressed tokens, when this was not the only thing they had to attend to.

INTERPRETATION

As the photoelectric glottograph is not calibratable there remains the possibility that results could be due to some artefact, e.g., a shift of the catheder with the light sensor in the pharynx might affect the amount of light picked up, and hence, the amplitude of the waveform. Long term catheder shifting could not however explain the systematic differences noted, given that the stress-varying utterances were read as alternating sentences and that recordings were short in any case. However, if there were to be some stress-related articulatory difference, e.g., laryngeal movement in the vertical dimension, it could conceivably yield the differences in PEG amplitude (though hardly the double peaks of the Irish emphatic stops). There is nevertheless some corroborative evidence for our interpretation. Andersen /7/ reports similar variation in PGO amplitude for voiceless stops spoken at different loudness levels. These findings were based on PEG data, but were to some extent backed up by fibre-optic and EMG data. More recently, on the basis of inverse filtered data, Fant /8/ and Gobl /9/ have reported for [h], a wider glottal opening in the stressed than in the unstressed syllable.

Even if our interpretation is correct, one must further ask whether these differences are actively controlled, or might simply be a passive consequence of some other stress correlate such as increased stop duration or higher subglottal pressure (Ps).

The Icelandic data shows that the degree of glottal opening can not be just a function of duration. Closure duration for the unstressed geminate is as long as for the stressed single stops, but PGO amplitude is much lower. Furthermore, within either stress condition, PGO is not geater for the geminate than for the single stops. Therefore, it can't be the case that the vocal folds will simply deflect more widely given extra time in which to do so.

A second possibility is that the vocal folds are blown wider apart as a consequence of increased Ps (a likely correlate of stress /10/, though not necessarily to be expected in every language: see, for example Welsh /11/). Ps values in our data were higher with increased stress. Peak Ps values in the vowel preceding the stops were on average 6cm Aq higher in stressed than in unstressed tokens for Icelandic (data set 1), and about 4cm Aq in Irish (data set 3). Emphatic tokens were 8 cm Aq higher again. Averaged Ps values for data set 2, are indicated in Fig 2. However, it is very unlikely that the differences in glottal opening degree are passive consequence of the Ps level, judging from an experiment by Löfqvist et al./3/. Sudden pressure changes, induced during PEG recordings by unexpected jabs in the subject's chest, made very little difference to the PEG

trace. Furthermore, the double peak glottal opening of the Irish emphatic stops can not be attributed to pressure variation. Although Ps is higher for the emphatic tokens (see Fig 2), there is no additional sudden increase during stop closure which could account for the second peak.

On balance therefore, it would seem that the observed differences in laryngeal behaviour are under active control. It may be the case that increased activity of the laryngeal musculature is a direct correlate of stress (typically perhaps an increase in abductive activity, but with an additional possibility of initiating a second abductive gesture when necessary). Thus, one could adopt the viewpoint that stress is potentially manifest by increased muscular activity at every level of production; the respiratory /10/, the laryngeal, and frequently, at the level of supralaryngeal articulation /12/.

These differences in laryngeal behaviour might not however be best regarded as correlates of stress as much as evidence of more general laryngeal strategies to ensure maximally equivalent output under differing aerodynamic conditions. To produce a voiceless segment there are potentially (depending on context) two crucial targets:

1) Sufficient glottal opening to ensure voicelessness. (Note: this excludes from present consideration glottalised stops, as occur in certain dialects of English.) The transition from voice -> voicelessness is not instantaneous (Westbury /13/ describes a "voice tail" of up to 40ms for voiceless stops in American English) and it can be very slow indeed when the vocal tract is not occluded, as for preaspirated stops /5/. If a greater degree of glottal abduction were not used at higher Ps levels, attainment of voicelessness might be delayed or prevented. Note that in the data presented, the higher the Ps, the greater appears to be the amplitude at which voice offsets.

2) The second target is the resumption of vocal fold vibration at the appropriate point in time subsequent to closure release. At voice onset, the initiation of vocal fold vibration results from glottal adduction and the Bernoulli effect; at a given stage of glottal narrowing, the vocal folds get sucked together. The point at which this happens depends on two factors working in an inverse relationship: the air flow rate through the glottis, and the degree of glottal narrowing. This may explain why, for the data in Figs 1 and 2, the actual duration of postaspiration is the same across different stress conditions, even though glottal aperture at stop release would seem to be quite different. Glottal closure in the higher stress tokens may simply be "stealing a ride", as it were, on the higer airflow and Bernoulli effect. Thus, at higer Ps and airflow rates, wider glottal opening may be not just tolerated, but actually necessary if VOT is to remain constant. At lower stress levels, too much glottal opening may be counter indicated as it would lead to unwanted aspiration.

Emphatic stress for the Irish medial stops may represent a particularly demanding articulation, given that they have both some pre- and postaspiration. Peak glottal opening occurs early during stop closure, as with preaspirating stops generally. If the ballistic opening/closing gesture were to proceed uninterrupted, the degree of glottal opening at stop release might not be sufficient to ensure the appropriate duration of postaspiration; hence the double opening. In initial postpausal position the situation is rather different, as only one of the above mentioned targets is relevant: appropriate voice onset. This may leave more freedom to use the additional strategy of delaying the peak glottal opening to prevent overshort aspiration at higher respiratory levels.

CONCLUSIONS

To conclude therefore, I would suggest that the degree of glottal opening is varied as a means of controlling voice onset and offset times across differing aerodynamic conditions. When necessary or possible, the additional strategies of a second opening peak or of a change in peak timing can also be brought into play. In other words, the larynx uses more than one control parameter to ensure that the crucial targets of voice offset and voice onset are maintained.

The fine interplay between laryngeal behavior and aerodynamic conditions suggests that there may be active monitoring of Ps at the laryngeal level.

REFERENCES

/1/ Kim, C., A theory of aspiration. Phonetica,21,107-116, 1970.
/2/ Catford, J.C., Fundamental Problems in Phonetics Edinburgh University Press, 1977.
/3/ Löfqvist, A., Baer, T., and Yoshioka, H., Scaling of glottal opening. Phonetica,38, 1981.
/4/ Löfqvist, A., Interarticulator programming in stop production. J. of Phonetics,8,475-490, 1980.
/5/ Ni Chasaide, A., Preaspiration in phonological stop contrasts. Unpublished Ph. D. thesis, University College of North Wales, Bangor, 1985.
/6/ Keating, P., Linker, W., and Huffman, M., Patterns in allophone distinction for voiced and voiceless stops, J. of Phonetics,11,277-290, 1983.
/7/ Andersen, P., The effect of increased speaking rate and (intended) loudness on glottal behaviour in stop consonant production as exemplified by Danish p. Ann. Rep. of the Inst. of Phonetics, Univ. of Copenhagen,15,103-146, 1981.
/8/ Fant, G., Interactive phenomena in speech production. Forthcoming in Proc. 11th Int. Cong. of Phonetic Sciences, Tallinn, 1987.
/9/ Gobl, C., Voice source dynanics in connected speech . Forthcoming in STL-QPSR, 1987.
/10/ Ladefoged, P., Three Areas of Experimental Phonetics. Oxford University Press, 1967.
/11/ Williams, B., Some problems in the description of stress in modern Welsh. Cambridge Papers in Phonetics and Experimental Linguistics,1, 1982.
/12/ Tuller, B., Harris, K.S., and Scott Kelso, J.A., Stress and rate: differential transformations of articulation. J.Ac.Soc.Am.,71,1534-1543,1982.
/13/ Westbury, J.F., Aspects of the Temporal Control of Voicing in Consonant Clusters in English. Texas Linguistic Forum, Texas, 1979.

REGULATION OF INTENSITY AND PITCH IN CHEST VOICE

HELMER STRIK and LOUIS BOVES

Institute of phonetics, Nijmegen University,
P.O. Box 9103, NL-6500 HD Nijmegen
The Netherlands

ABSTRACT

The simultaneous control of fundamental frequency and intensity of phonation was investigated.. Simultaneous measures were obtained of laryngeal muscles and the respiratory system. An inverse relation between intensity and activity of vocalis and cricithyroid was found at high pitch chest voice

INTRODUCTION

Studies of the simultaneous control of fundamental frequency (Fo) and intensity of phonation have dealt with either the intrinsic laryngeal muscles [1,2,3,4] or with the respiratory system [5,6]. In one study [7] subglottal pressure (Ps) was measured simultaneously with electromyographic (EMG) activity of intrinsic and extrinsic laryngeal muscles for singing. The purpose of the present study is to reexamine the simultaneous control of Fo and intensity in speech. To that end simultaneous recordings of speech, electroglottogram (EGG), Ps, lung volume, and EMG activity of cricothyroid (CT), vocalis (VOC) and sternohyoid (SH) were obtained while the subject sustained the vowels /e/ and /i/. These physiological signals were chosen because they are believed to be important in the control of both intensity and pitch. No open vowels were used because jaw opening may influence the intensity of the radiated sound and the EMG-activity of certain laryngeal muscles.

METHOD

Speech Material

The subject was a male native speaker of Dutch. He was instructed to breath deeply and to phonate a vowel as long as possible at a constant Fo. This task was repeated five times for each of twelve different conditions: 2 vowels (/e/ and /i/), 3 intensities (soft, normal and loud) and 2 frequencies (low and high), making a total of 60 phonations. The intensity level (IL) of the soft utterances was approximately 7 dB below the normal IL, and the IL of the loud utterances was approximately 5 dB above the normal IL.
The audio signal was led to a pitch-extractor, and the subject could see the Fo level on a oscilloscope. Both target levels for Fo, viz. a low pitch level of 116 Hz and a high pitch level of 160 Hz, were indicated on the oscilloscope screen. In this way the subject could control his Fo.

Recording and Processing of Data

Subglottal air pressure was recorded using a Millar pressure transducer, which was inserted pernasaly and fed through the glottis into the trachea. The pressure measurement was calibrated by recording the signal while the subject held lung pressures of up to 20 cm H2O against a water-filled U-tube manometer [8]. The catheter, situated in the posterior commissure of the glottis, did not have a noticeable effect on phonation. The perimeter of the thorax and abdomen were measured with a Respitrace inductive plethysmograph. Lung volume was calculated from the weighted sum of these two signals. Special calibration manoeuvres yielded the two weighing factors. The EMG activity was recorded with hooked-wire electrodes [9]. All electrodes were inserted percutaneously. Correct electrode placement was confirmed by monitoring an oscilloscope during various functional manoeuvres [9].
The physiological signals, the audio signal, an octal code and a timing pulse were recorded on a one inch, 14-channel instrumentation recorder [10]. The processing of the data was done with the Haskins Laboratories EMG data processing system [11]. The voice signal was sampled and digitized at a 10 kHz rate, Ps and EGG at a 5 kHz rate, and EMG, chest and abdomen signals at a 200 Hz rate. The mean glottal flow (Ig) was obtained by taking the derivative of the lung volume. The vocal intensity was evaluated from the audio signal. The intensity levels were calculated relative to the lowest measured IL. The Fo was derived from the EGG in order to verify whether the Fo of the utterances remained roughly at the target levels. The EGG signal was also used to obtain the open quotient (OQ), which is defined as the time during which the glottis is open divided by the time of one vibratory cycle. The signals of six physiological quantities (viz. Ps, Ig, IL and EMG activity of VOC, CT and SH) were used for further processing.
Using the voice onset as a line-up point the six physiological signals of the five repetitions were averaged, resulting in an averaged signal for each quantity and each condition. Because of an artefact in the recordings of the EMG activity of the SH some of the tokens had to be discarded, but of each utterance at least four tokens remained for processing and averaging. For reasons that will be mentioned below, the mean value between 2 and 10 seconds after voice onset was calculated for these average signals. So the final measures are mean values for each of the six quantities in each of the twelve conditions.

RESULTS AND DISCUSSION

The duration of the sustained vowels varied between 10 and 20 seconds. While phonating the EMG-activity of VOC and CT, the Ps, Ig, IL and Fo were approximately constant. The EMG-activity of SH had a peak value immediately before the onset of phonation. Collier [12] and Hirose and Sawashima [13] also observed SH activity just before voice onset and assumed that the SH helps in preparing the larynx for the "speech mode." The peak value of SH activity depended on the frequency and intensity of the vowel that had to be produced. The largest peak values were recorded in the low frequency - high intensity condition, while the peak was almost absent in the high frequency - low intensity condition. In all utterances the EMG activity of the SH had levelled off to a more or less constant value 2 seconds after voice onset. Therefore for each measured physiological quantity the mean value was calculated from 2 to 10 seconds after voice onset, as mentioned above.
The mean values of the average signals are shown in the Figures 1 to 4. For each repetition the mean value between 2 and 10 seconds after voice onset is also given. The results are analyzed by making three different comparisons for the relevant physiological quantities.

1. /e/ vs. /i/

A comparison is made between the data of the vowels /e/ and /i/. Both are closed vowels and therefore the jaw opening was roughly the same. The major distinction between the two vowels is a difference in their formants, caused by a different vocal tract shape. This did not result in big dissimilaraties between the recorded signals, but some differences did occur.

Ps. At equal intensity levels the Ps was always slightly higher when the vowel /i/ was produced.

CT. For the low Fo condition the activity of the CT was about the same for both vowels, but for high Fo the activity of the CT was less for the vowel /i/. This can be a compensation to keep Fo constant, because an increased Ps could raise the Fo

VOC and SH. These two muscles showed more activity for the vowel /i/ when phonating at low Fo, and approximately the same activity when phonating at high Fo.

2. Fo regulation

For the same phonetic condition and intensity, the signals recorded at low and high pitch voice are compared.

VOC and CT. From the Figures 1 to 4 it can be seen that the activity of VOC and CT was substantially higher in high-pitch chest voice than in low-pitch chest voice. This confirms previous findings that the VOC and the CT are the primary muscles in regulating Fo, especially in chest voice [3,4].

SH. The activity of the SH decreased with increasing frequency, a result also found by Ohala [14] and Collier [12]. The Figures 1 to 4 show that the decrease of SH activity was more obvious at the high IL.

Ps. Across different fundamental frequencies Ps was almost the same. This is contradictory to the general believe that Fo and Ps are positively related [12,15]. In this case the Fo is not raised by increasing Ps, but probably by an appropriate adjustment of the activity of CT, VOC and SH.

3. intensity regulation

Since the vocal intensity is also a function of the acoustic impedance of the vocal tract, comparison of intensity is only done between two states in which the shape of the vocal tract is approximately the same, i.e. when the subject produced the same vowel.

Ps. Intensity was always positively related to Ps. This is consistent with the results obtained by Isshiki [5] and Baer [7].

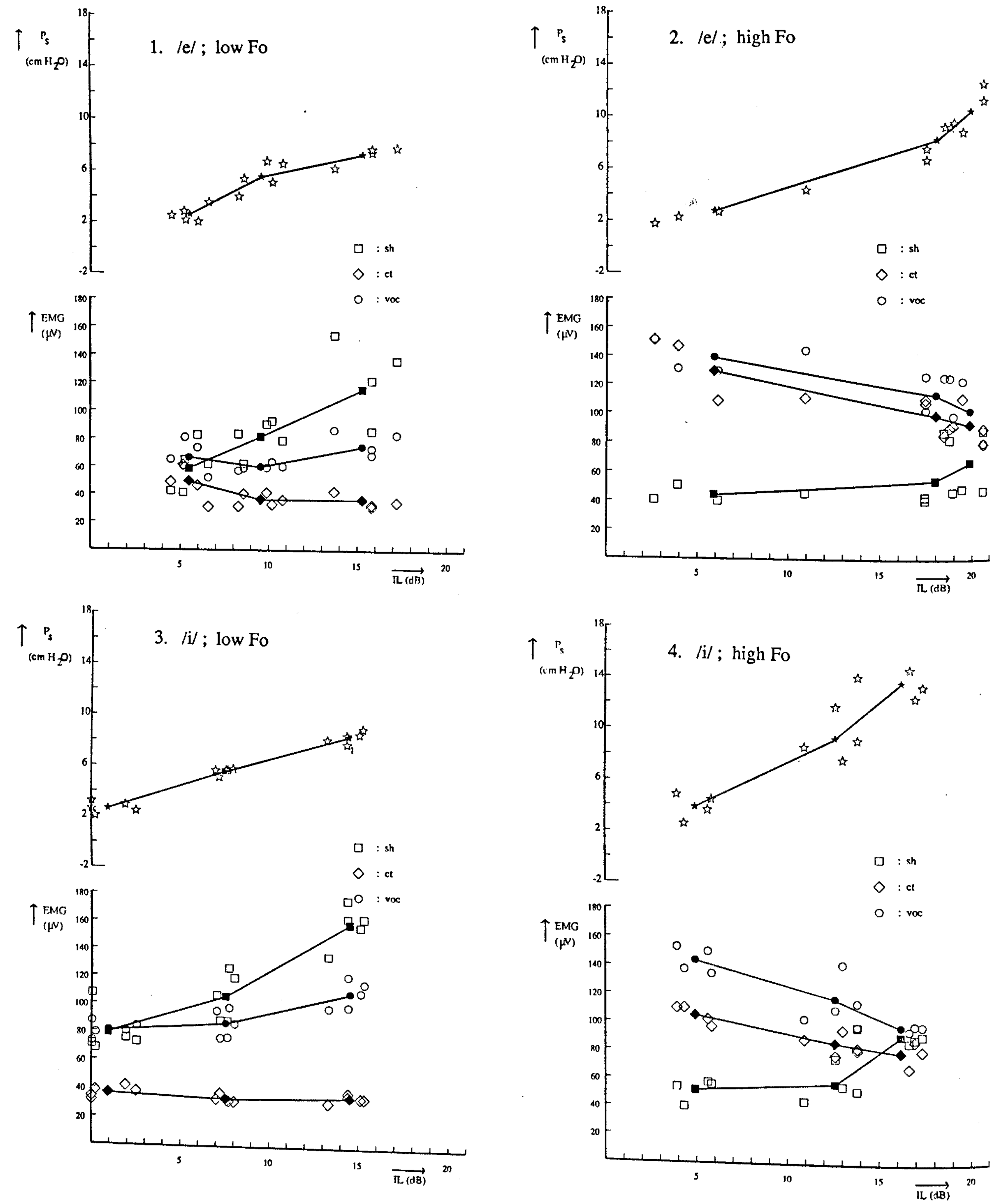
Ig. The glottal flow was more or less constant for different intensities. The EGG recordings revealed that the OQ decreased with increasing intensity. Therefore, although Ps increased with increasing intensity, Ig could remain fairly constant [5]. In the chest register glottal flow is not dominant in controlling intensity; apparently, the form and spectral content of the flow pulses are more important.

SH. A positive relation between intensity and the EMG activity of the SH was found.

VOC and CT. At low chest voice no significant change in the EMG-activity of the VOC and CT as a function of intensity were found. Gay et al. [4] also found that muscle activity, of all five intrinsic laryngeal muscles, remained relatively steady across changes in vocal intensity.
At high pitch chest voice a negative relation between intensity and the EMG activity of VOC and CT was found. The compensatory mechanism is necessary to keep Fo constant, because some of the factors that increase intensity also increase Fo.

CONCLUSIONS

First of all, it appears to be possible to maintain a constant subglottal pressure during a prolonged utterance, regardless of the decreasing volume of the lungs. Also Fo and IL can be kept constant without the need of apparent actions of CT, VOC or SH. Thus there seems to be no reason to assume that the often observed declination in speech is an involuntary effect of the decreasing lung supply.
The findings of this study are in agreement with those of previous studies: VOC and CT are the primary muscles in regulating Fo, IL is positively related to Ps, and Ig is more or less constant for different intensities in the chest register.
Müller found that an increase in vocal intensity without an associated rise in Fo had to be accompanied by a decrease in CT activity [16]. Rubin also speculated on a decrease in intensity of contraction of the CT with increasing loudness, if Fo is to remain constant [17]. Hirano actually measured that CT activity changes often varied inversely with the vocal intensity [1,2]. In the present study it was found that the increasing Ps can be compensated, not only by a decrease of CT activity, but also by a decrease of VOC activity. Further study is needed to explain why this compensation mechanism only occured at high pitch voice.



Figures 1, 2, 3 and 4. In these figures the Ps (cm H_2O) and the EMG activity (μV) of SH (□), CT (◊) and VOC (O) are plotted as a function of IL (dB). The open symbols represent the mean values of the tokens, and the closed symbols represent the mean values of the averaged signals (for further explanation see text).

REFERENCES

[1] Hirano, M. and Ohala, J. (1969). The function of laryngeal muscles in regulating fundamental frequency and intensity in phonation. Journal of Speech and Hearing Research (3): 616-627.

[2] Hirano, M.; Vennard, W. and Ohala, J. (1970). Regulation of register, pitch and intensity of voice. Folia Phoniatrica 22: 1-20.

[3] Sawashima, M.; Gay, T.J. and Harris, K.S. (1969). Laryngeal muscle activity during vocal pitch and intensity changes. Haskins Laboratories Status Report on Speech Research 19/20: 211-220.

[4] Gay, T.; Hirose, H.; Strome, M. and Sawashima, M. (1972). Electromyography of the intrinsic laryngeal muscles during phonation. Annals of Otology, Rhinilogy and Laryngology 81 (8): 401-409.

[5] Isshiki, N. (1964). Regulatory mechanism of voice intensity variation. Journal of Speech and Hearing Research 7: 17-29.

[6] Isshiki, N. (1965). Vocal intensity and air flow rate. Folia Phoniatrica 17: 92-104.

[7] Baer, T.; Gay, T. and Niimi, S. (1976). Control of fundamental frequency, intensity and register of phonation. Haskins Laboratories Status Report on Speech Research 45/46: 175-185.

[8] Cranen, B. and Boves, L. (1983). A set-up for testing the validity of the two mass model of the vocal folds. In: Titze, I.R. and Scheren, R.C. (eds.), Vocal Fold Physiology.

[9] Hirose, H. (1971). Electromyography of the articulatory muscles: Current instrumentation and techniques. Haskins Laboratories Status Report on Speech Research 25/26: 73-86.

[10] Port, D.K. (1971). The EMG data system. Haskins Laboratories Status Report on Speech Research 25/26: 67-72.

[11] Kewley-Port, D. (1973). Computer processing of EMG signals at Haskins Laboratories. Haskins Laboratories Status Report on Speech Research 33: 173-183.

[12] Collier, R. (1975). Physiological correlates of intonation patterns. Journal of the Acoustical Society of America 58: 249-255.

[13] Hirose, H. and Sawashima, M. (1981). Functions of the laryngeal muscles in speech. In: Stevens, K.N. and Hirano, M. (eds.), Vocal Fold Physiology.

[14] Ohala, J. (1972). How is pitch lowered? Journal of the Acoustical Society of America 52: 124.

[15] Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. Journal of the Acoustical Society of America 65: 1271-1275.

[16] Müller, J. (1843). Elements of physiology. Trans. by Wm. Baly, Philadelphia.

[17] Rubin, H.J. (1963). Experimental studies on vocal pitch and intensity in phonation. The Laryngoscope 8: 973-1015.

INVESTIGATION OF THE VOICE SOURCE MODELS

A.M.Smetanin

Department of Technical Cybernetics Mechanical Institute
Ustinov, USSR, 426069

ABSTRACT

An experimental research of conformity of the forming voice source impulse models to real process is carried out. The technique of the research is based on the analysis of power spectra of the actual and synthesized voice signal with using the linear prediction. The synthesis model of the vocalized speech signal accounting the influence of the voice source upon the voice canal is proposed.

INTRODUCTION

Before now the accepted notions about the speech forming process of the vocalized speech sounds proposed the excitation of the voice canal by quasiperiodic impulses of air stream such as the smooth unimodal time function. Proceeding from these notes it is obvious that the excitation of the voice canal must start near the moment of glottis opening. However, at present it is known that the start of the excitation of the voice canal coincides with the vocal chords closing moment [I]. The causes of this obvious contradiction weren't analised in literature about speech forming. In this paper the experimental researches were carried out in order to analyse real processes of the excitation of the voice canal by the voice source and build the model of synthesis of vocalized voice wave based on received results that is adequate to real speech signal forming. The carried out researches concerned only the interaction of the voice source with the voice canal and didn't discuss those processes as the loss in the canal, nasalization, radiation etc.

THE TECHNIQUE OF RESEARCH

The research method of the processes of the vocalized speech signal forming is based on analysis of power envelope spectrum of signal. Because of problem statement the speech wave analysis synchronous with the pitch was to be used, that involves application of the methods of increasing the resolution of the frequency. And at last the unstationary variant of the linear prediction was used [2]. The part of apparatus of the research complex contains mini-computer SM-4 supplied with the device of the analogous signal input. Special research software based on the program complex for processing and signal modelling included the following main modules: "visible speech" forming, calculation of the linear prediction coefficients, calculation of power spectrum by means of fast Fourier transformation, the impulses of different form excitation forming, the speech canal modelling, speech wave synthesis. The programming language is FORTRAN, the operational system is RAFOS.

A. REAL SPEECH SIGNAL

A speech signal put into computer (frequency range - 5kHz, quantification frequency - IOkHz, the number of quantification levels by amplitude - 256, signal-noisy relation hot worse than 40dB) was accumulated on the magnetic carrier of the computer. Speech material consisted of the words pronounced isolatedly by three announcer. Quasistationary parts of stressed vowels were subjected to the analysis. In calculation the prediction order was p=I2, coefficients of the linear prediction were averaged by three-four samples taken within the interval of analysis. Typical example of the evaluation of power envelope spectra is shown in Fig.I, where I - measured along the whole period of the pitch; 2 - during the closing interval; 3 - during the interval of vocal chords opening (the fragment of sound /i/ in the word "electrichestvo").

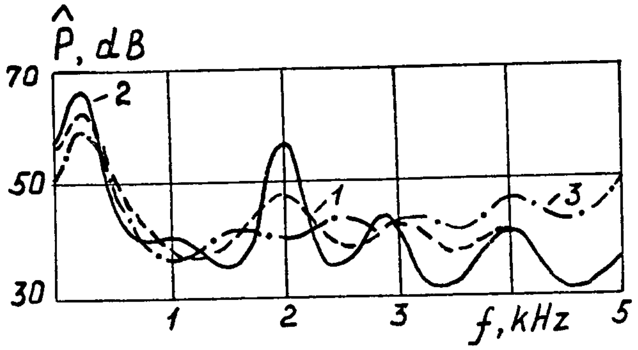


Fig.I. The power spectrum of the real signal

The experiments have shown the following [3]; I) envelope spectra during the intervals of vocal chords closing and opening are distinguished essentially by frequency and width of the spectral maximums band (formants); 2) the spectrum during the whole period of the pitch is more close to the spectrum during the interval of vocal chords closing. Thus there is sufficient variations of the voice wave parameters during the period of the pitch that is explained in general by the voice source influence on these parameters in the phase of glottis opening. The carried out calculations by the numeral evaluation of influence of the voice source on the speech canal parameters have shown that with normal conditions of pronunciation: I) the absolute changes of frequency of the formants achieve the quantity about IOOHz; 2) the absolute changes of the band width of formant-quantity about 300Hz [4].

B. THE SYNTHESIZED SIGNAL

Let us consider the model work of the synthesis of the vocalized voice signal with the excitation of the voice canal by smooth unimodal impulses. The raised cosinusoid with duration of 0,3...0,7 from the pitch period was used as the impulse. The voice canal was presented in the form of cascadely connected digital resonators corresponding to the formants [I]. In table I the parameters of the first of five formants are given, used at the synthesis of vowal /i/.

Table I

Frequency and width of the formant band

Formant	F_i, Hz	B_i, Hz
I	440	90
2	I800	50
3	2550	300
4	34I0	300
5	4400	3I0

The character of the time function of the synthesized speech signal has already shown the sufficient quality difference from real signal. The typical analysis of vowal /i/ are shown in Fig.2. The notations correspond to Fig.I. The duration of the pitch is I20 counts, the duration of

the excitation impulse for Fig.2a - 40 counts, for Fig.2b - 80 counts.
Let us compare Fig.2 and Fig.3 where the spectrum of impulse reaction of voice canal with parameters of table I is given. The calculation of this spectrum has been made also by means of written technology under the excitation of canal model by the single impulse. The comparing results illustrate that cosinusoidal impulse of excitation distort the spectrum of frequency of the synthesized signal. The typical results have been obtained under the other forms of the excitation impulse triangle, in particular.
The carried out experiments and modelling say about nonadequancy of model excitation of the canal by smooth unimodal impulse. Besides, you may conclude that excitation of digital model of the voice canal by the single impulse is more close to real process of speech formation.

THE SYNTHESIS MODEL OF SPEECH SIGNAL

The carried out investigations allow to formulate two main demands to the synthesis model of vocalized speech signal:
I) the excitation of the voice canal should be produced by short at the moments of vocal chords closing; 2) during the interval of the vocal chords opening the change of the voice canal parameters must be carried out, modelling, the influence of the voice source upon the speech wave parameters.
The model illustrated in Fig.4 meets these requirements. It is the development of the model examined in [I].
According to the given model the voice canal in the form of a filter

$$V(z) = \frac{G}{1-\sum_{k=1}^{p} a_k z^{-k}}$$

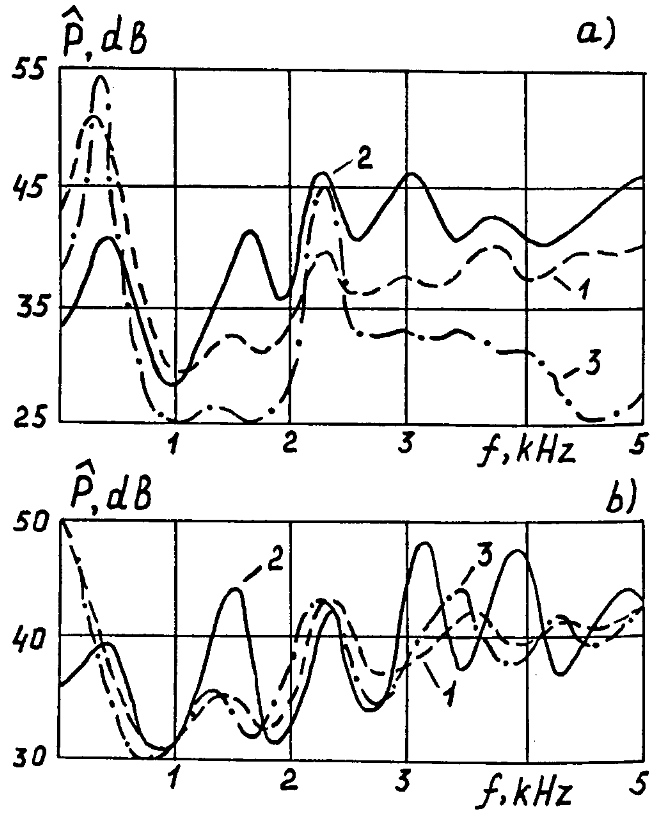


Fig.2. The power spectra of a synthesized signal with cosinusoidal source of the excitation

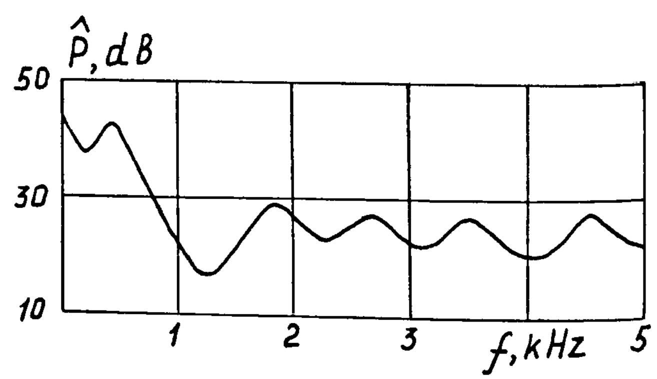


Fig.3. The impulse reaction spectrum of the voice canal

is excited by $\delta(n)$ signal at time moments t_1, t_3 etc. The parameters of the canal are changed in accordance with the function $f(n)$ during the interval of opening $[t_2, t_3]$. For the formant model, in particular, frequency F_i and width of strip B_i are these parameters counted in $V(z)$.

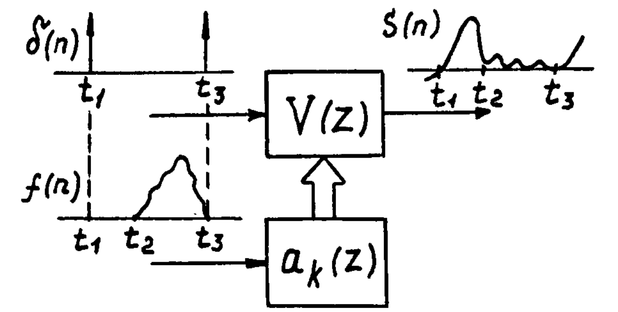


Fig.4. The synthesis model

Fig.5 illustrates the power envelope spectra synthesized by means of the suggested model signal (Fig.I). The parameters of formant during the interval of the vocal chords closing are given in table I. The duration of the pitch period was equal to I20 counts; the duration of the opening period was equal to 80 counts; the absolute changes of frequency of 5 formants during the opening interval - I00Hz; the absolute changes of the band width - 200Hz. Comparing figures I, 2, 3 and 5 we see that a given model at the spectral level is much closer to the real process of speech formation than a traditional model.

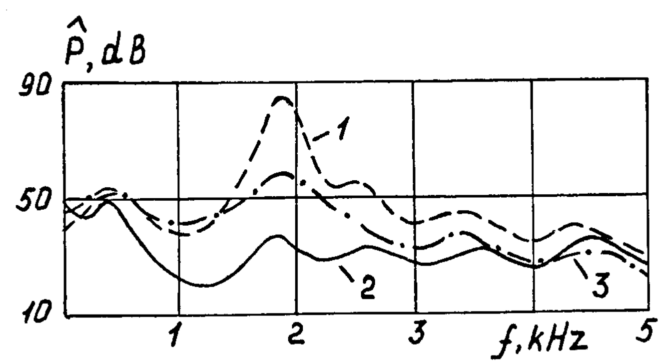


Fig.5. The power spectra synthesized by means of the suggested model signal

In the time domain the character of the synthesized speech wave is also close to real signal. It may be supposed that such model of synthesis will allow to raise naturality of synthesized speech, to model some individual peculiarities of voice. The examination of these assumptions is not included in this paper.

CONCLUSION

In consequence of carried out experimental investigations it is shown that the imagination about the excitation of the voice canal by smooth unimodal impulses of air stream are not adequate. The excitation of the voice canal by single impulses during the period of time corresponding to vocal chords closing and modelling of change of speech wave parameters as the smooth unimodal function during the phase of vocal chords opening is more close to real process of speech formation. In our opinion well known, so called, impulses of the voice source are the form of power accumulation for the next excitation of the voice canal at the moment of vocal chords closing.

REFERENCES

[I] L.R.Rabiner, R.W.Schater, "Digital processing of speech signals", New Jersey.

[2] Y.Makhoul, "Spectral analysis of speech by linear prediction", IEEE Trans. Audio and Electroacount, I973, vol.AU-2I, N3, p.I40-I48.

[3] В.Б.Гитлин, А.М.Сметанин и др., "Выбор интервала измерений частоты и ширины формант", материалы Всесоюзной школы-семинара АРСО-I0, Тбилиси, октябрь 1978, с. 20-22.

[4] В.Б.Гитлин, А.М.Сметанин, "О повышении точности измерения параметров формант", "Проблемы построения систем понимания речи", М.: Наука, I980, с.I09-II5.

LARYNX - DOUBLE-SOUND GENERATOR

Boris Chernov
Dozent
Conservatoire
Novosibirsk, USSR 630099

Valentin Maslov
Phoniatric
Policlinic No 1
Novosibirsk, USSR 630099

ABSTRACT

Only aural transmission from man to man of the oldest form of creative activity brought to us the method of phonation of the genetic period of an inarticulated speech. Millions of years have passed till the entrance to larynx was excluded from the process of phonation. The beginning tone of vocal folds got the possibility to be formed into vowel and consonant sounds of speech. From whistle to voice-whistle and articulate speech - that's the way of evolution.

INTRODUCTION

Nowadays on vast territories of the Centre and the South of Asia among Turkish and Mongolian peoples there remained the most ancient forms of the organisation of sound formation in normal larynx that demonstrate an active participation of vestibular folds at the expense of sharp contraction of the larynx entrance, such as larynx whistle and phonation simultaneously through the two barriers of vocal and vestibular folds.
Larynx shows the capacity of a double-sound generator and clearly demonstrates the mechanism through which the formation of an articulated speech is impossible.
The peoples of Mongolia, Touva, Bashkiria managed to preserve the ways of phonation peculiar for the genetic period of an inarticulated speech in the form of traditional singing folklore that have passed through centuries thanks to living transmission.
The most stable are the methods of double-voice singing in four styles: Syghyt, Ezingeler, Kargyraa and Barbbannadyr of the Touvinians.
The first notes of the Bashkirian style Usllau date to 1897. Folklorist Rybakov S.G. has characterized it as Forest Wildeness. The forms of Touvinian double-voice singing were studied by musicologists and folklorists beginning with 1900. The analysis of Touvinian guttural singing was made by the Soviet composer A.N. Aksionov during the 60-th. The acoustic analysis of Triple-voice singing of Tibet lamas was made by English scientists H. Smith, K. Stevens, R. Tomlinson in 1967. In 1973 at the YIII all-union acoustic conference of AS of the USSR A.A. Banin and V.N. Lozhkin reported the results of acoustic analysis of Touvinian larynx singing made with Sona-Grap-7029 A apparatus in diapason of 40 - 4000 Hz. They found the characteristics of low tone from 60 to 220 Hz and high pitch from 2000 to 3000 Hz. But it appeared to be impossible to explain the physiological mechanism of larynx with the help of acoustic analysis.

Special research group of the authors of this report was made at the initiative of Ministry of Culture of Touvinian ASSR and rectorate of the Novosibirsk Conservatoire.
The first examination of vocal apparatus of Touvinian singers showed that they have no abnormal deviations of anatomo-physiological character. When the singer began double-voice regime of phonation the before seen vocal folds entirely disappeared off the investigator's sight. The source of the second sound appeared to be a round whistle hole of D=1,5-2 mm formed by (false) vestibular folds.
In 1975 - 76 unique photographs of the sources of the high-frequency whistle of 2000 - 4000 Hz were made (Fig. 1).
The following methods were used: filming of singers; indirect laryngoscopy; filming of the functioning of the larynx in indirect laryngoscopy; tomography of the larynx; tele-X-ray cinematography; video-magnetic recording from TV screen; recording of various styles of double-voice singing.
In the Syghyt style, a singer begins the first phrase of the song with words in an ordinary manner with his face relaxed and his articulation and breath having no visible signs of effort. After finishing the phrase sung naturally the singer takes a new breath and begins double-voice singing which excludes the possibility of using words. So the vocal organs start working like a peculiar double-voice musical instrument (fig. 1). To the ear double-voice singing presents itself as two melodies the lower of which is of ostinato character and keeps the pitch

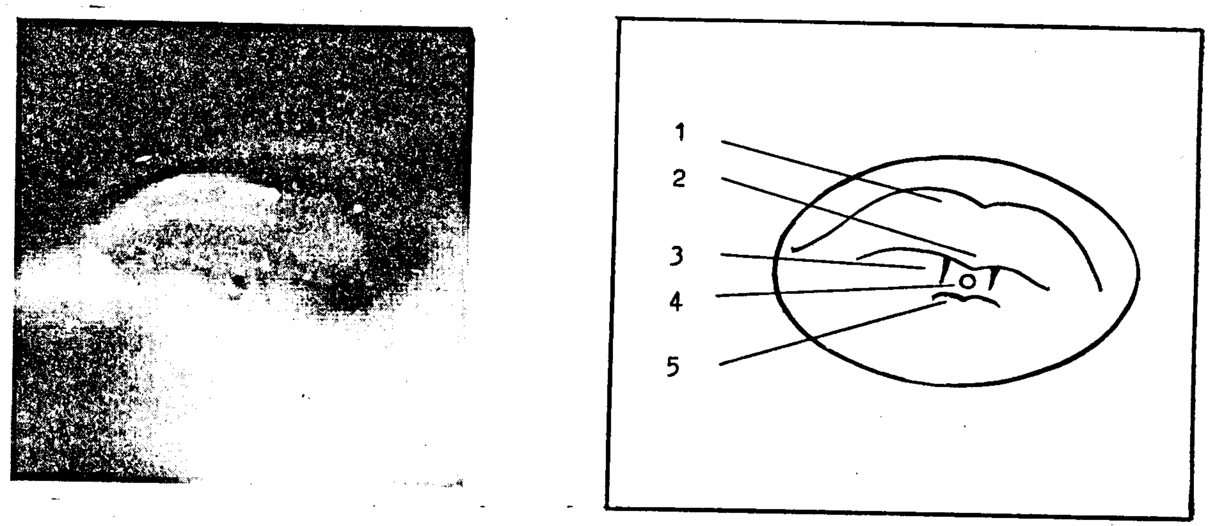


Fig. 1. Opening to the larynx in Touvinian double-voice singing (in laryngoscopy). One can see a tuberculum epiglotticum the tops of the arytenoid cartilages and the margins of false vocal folds in action in forming a narrow opening to the larynx like a small passage or nozzle. 1 = Lead of epiglottis; 2 - tuberculum epiglotticum; 3 - margin of ventricular vocal fold; 4 - opening of 'nozzle' surrounded by foaming mucus; 5 - top of arytenoid cartilage.

of the octave. The second melody is heard as a kind of whistle complicated by flowery decoration and lies in the sphere of the third and fourth octave.
The transition from usual single-voice phonation to double-voice singing is followed by abrupt changes in the functioning of the larynx. The larynx quickly pulls up and the loose margin of the epiglottis becomes visible deep in the mouth cavity, with the tongue not being stretched at all.
Indirect laryngoscopy shows that in this position the vocal folds become invisible as the upper opening of the larynx narrows to 1.5-2 mm because of all the formations arranged at this level. The tuberculum epiglotticum draws near the apex of the cartilagines arytenoideae. From the sides, muscles of the ventricular folds and fibres of the musculus aryepiglotticus participate in narrowing.
The upper opening of the larynx begins to function according to the 'nozzle' principle producing a whistling tone resounding in the pharyngeal cavities.
This process of changing the larynx into a double sound generator was clearly observed in sagittal X-ray cinematography that shows the rise of the larynx as well as its narrowing and the sharp exact movements of the tongue which are synchronous with the changes of pitch of the whistling tones leading the ornamental melody.
The formation of two narrow passages in the larynx in accordance with the sounds produced by it in double-voice singing can also be observed in tomograms (fig. 2). In the frontal tomograms two narrow passages are seen: the first is due to the closure of the plicae vocales, and the second is formed by closing the

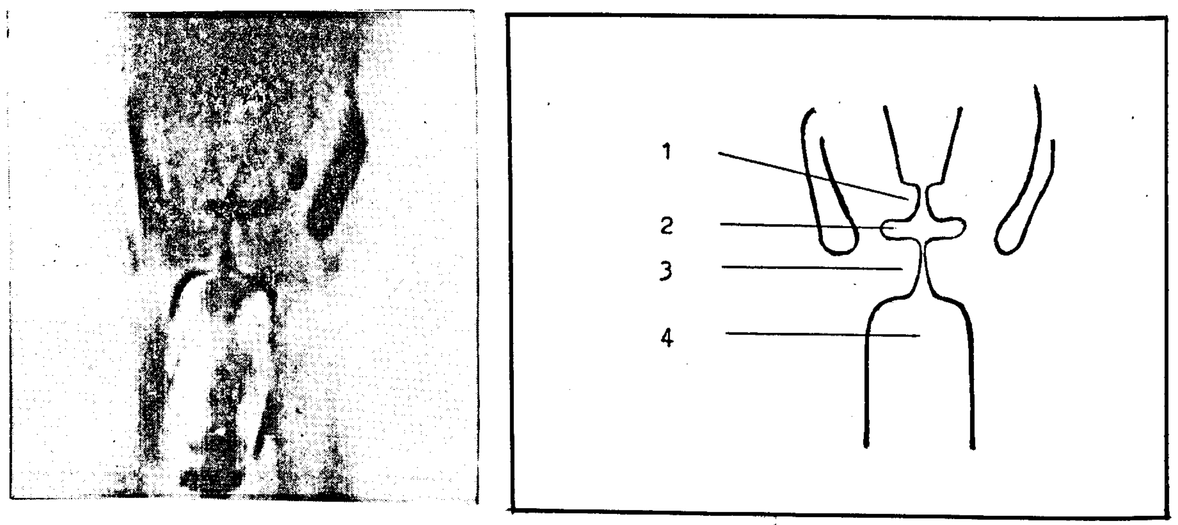


Fig. 2. Tomogram of the larynx in double-voice singing showing the presence of two narrow passages, one on the level of the true vocal folds and the other on that of the ventricular folds. Ventricles of Morgagni remain relaxed. 1 = Closed ventricular folds participating in the formation of a narrow opening to the larynx; 2 = ventricles of Morgagni relaxed; 3 = true vocal folds closed; 4 = aperture of trachea.

ventricular folds and musculature in general and other muscle elements of the upper opening of the larynx, with the ventricles of Morgagni being relaxed.
The second passage works like a nozzle or a whistle. It creates high-frequency vibrations which then resound in the pharyngeal cavities, forming sounds of various pitch.
The larynx of a Touvinian singer who sings in the double-voice manner is a two-sound generator in which the pitch of the low tone is created by the vocal folds, while the pitch of the whistling tone is a result of the activities of the narrowed entry to the larynx (nozzle) and of the resonator cavities of mouth and pharynx. These uses of the larynx and of the mouth and pharyngeal cavities demonstrate amazing functional possibilities of the vocal organs for creating sounds and exclusive abilities to govern their pitch and duration.
As a result of long and thorough examination of physiological mechanism of larynx singing of Touvinians, Khakassians and Bashkirs a new capacity of larynx unknown to science was opened - the capacity to form mechanism of aerodynamic whistle.

Literatur.

I. Wendler, W. Saidner. Lehrbuch der Phoniatrie. Leipzig, 1977.
A. Banin, V. Lozhkin. VIII th All-Union Acoust. Conf. Reports. Ser.G. (AS SSSR, MOSCOW 1973)
V. Maslov, B. Tchernov. Le secret du solo en duo "Sov. Ethnographia". No. 1 M., 1980.
B. Tchernov, V. Maslov. Fenomen of Touvinian double-voice singing. "Natur" No. 6. M., 1978

HIGH-SPEED DIGITAL IMAGE RECORDING FOR THE OBSERVATION OF THE VOCAL CORD VIBRATION

SHIGERU KIRITANI HIROSHI IMAGAWA HAJIME HIROSE

Research Institute of Logopedics and Phoniatrics
University of Tokyo, Bunkyoku 7-3-1, Tokyo, Japan

Abstract

A new method for the high-speed digital image recording of the vocal fold vibration was developped using a solid endoscope and an image sensor. Video signals from the image sensor are digitized and stored in a digital image memory. Stored images are displayed on a monitor oscilloscope. Frame rates of 2000/sec and 4000/sec are realized for the images with 100x37 and 100x17 picture elements, respectively. Compared to the ordinary high-speed motion picture system, the present system is compact and enables flexible data collection.

A similar system using a fiberscop was also developped. Although images by the fiberscope are darker, a frame rate of 2000 per second was achieved for the images with 100 x 17 picture elements. The system makes it possible to observe vocal cord vibrations during the consonants. Preliminary observations on the transitional characteristics of the vocal cord vibrations during the implosion and explosion of the consonants are performed.

Introduction

This paper presents a new technique for high-speed digital image recording for the observation of vocal cord vibration.

For the study of the voice source characteristics, it is essential to record the vocal cord vibration simultaneously with the speech signal and to analyze the relationship between the pattern of the vocal cord vibration and the acoustic characteristics of speech signal. Observation of the vocal cord vibration has generally been performed by using a high-speed motion picture. However, that method requires special equipment and is not suited for flexible data collection under various modes of phonation. For the simultaneous recording of speech signal, special considerations are necessary on the acoustic shielding of the mechanical noises from the high-speed camera.

In the present study, a system of high-speed digital image recording was developed. The sytem is small and compact and, thus, enables flexible data collection. Simultaneous recording of speech and other physiological signals can be performed easily.

Solid endoscope system

Fig.1 shows a block diagram of the system. The system consists of an oblique-angled solid endoscope, a camera body containing an image sensor and an image processor. The output video signal from the image sensor is fed into the image processor through a high-speed A/D converter. Stored images are then displayed on a CRT monitor as an array of small images which represent sequential time frames. It is also possible to display stored images as a slow motion picture. The image processor in the present system contains about 750k byte of image memory. Generally, for one shot of image recording, about 100 frames of image data are sampled and stored by the image processor.

Maximum frame rate that can be realized by the present system is determined by the brightness of the image obtained through the endoscope and the speed of scanning the picture elements in the image sensor.

In order to get a brighter image, a new model of the solid endoscope was constructed. The diameter of the scope was larger than that of the ordinary scope for the clinical use. The cross

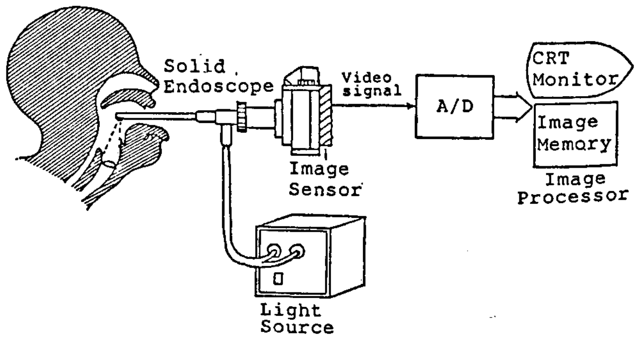


Fig. 1 Blockdiagram of the solid endoscope system.

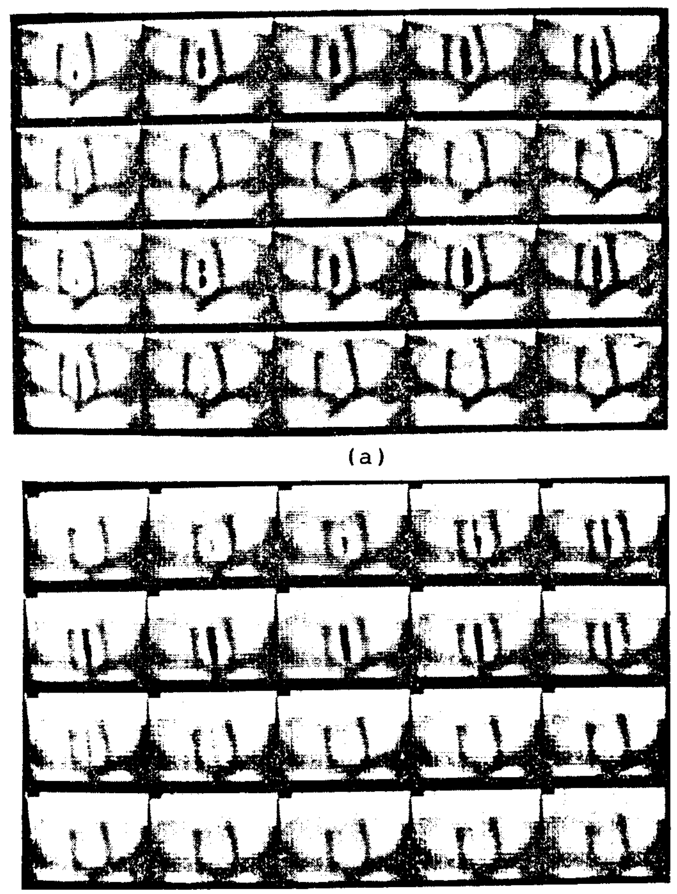
section of the tube of the scope is an ellipse. The image guide is located at the center of the tube and on both sides of the image guide are light guides. The two light guide cables are connected to the separate light sources.

Table 1 summarizes basic characteristics of the present system. Light sources are the two 250W halogen lamps. Number of the picture elements in the image sensor is 100x100 and the sampling rate is 10MHz. When entire picture elements are scanned, the resulting frame rate is about 1000 per second. In order to achieve a higher frame rate, a special scan method was devised in which only the selected scan lines were sampled. When 37 scan lines are sampled out of 100 scan lines, the frame rate is 2000 per second. A higher frame rate of 4000 frames per second is achieved by sampling only 17 scan lines. As far as the brightness of the image is concerned, this frame rate appears to be nearly the maximum that can be achieved for the most subjects. Naturally, the brightness of the image varies, depending on the laryngeal view of the individual subjects. The image memory can store the image data for the period of about 200msecond.

Table 1 Basic characteristics of the endoscope system.

Light source	250W halogen lamp x 2
Image sensor	MOS type
	100x100 picture elements
	Clock 10MHz
Image memory	768k byte

Scan elements	Frame/sec	Storage
100x37	2000	200
100x17	4000	450 (frames)



(a)

(b)

Fig.2 Examples of the recorded image of the vocal cord vibration. (a) 2000 frames /sec (b) 4000 frames /sec

Fig. 2 shows examples of the image recorded at a rate of 200 and 4000 frames per second. The pictures shows the vocal cord vibration of a male subject during the sustained phonation of the vowel /e:/. Fundamental frequency of the voice was about 200Hz. The images of the maximum glottal opening are observed at about every 20 40 frames.

Fiberscope system

It is very valuable if a similar system can be constructed using a fiberscope. Such a system makes it possible to observe the vocal cord vibrations during consonants. In the present study, a pilot system of high-speed image recording using a fiberscope was also developed.

Fig. 3 shows a block diagram of the system. In this system, a video camera is connected to the finder of a single-lens reflex camera to monitor and record the glottal view preceding and following the short period of high speed imaging. The image is sent to the image sensor for high-speed image recording only when the shutter of the camera is open. This monitoring is necessary beacause the fiberscope system mainly aims at the observation of the glottis during the running speech and the period of high-speed imaging is very short.

In order to obtain a brighter image, a new fiberscope was also constructed the

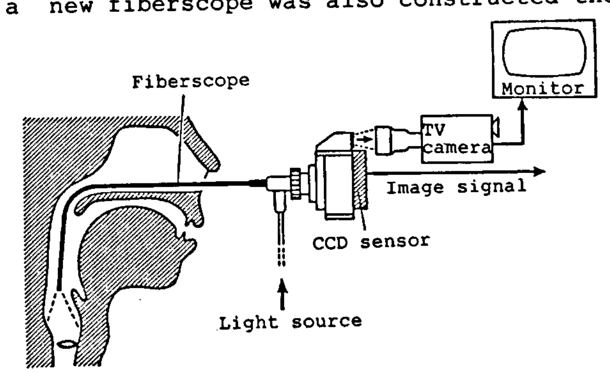


Fig. 2 Blockdiagram of the fiberscope system.

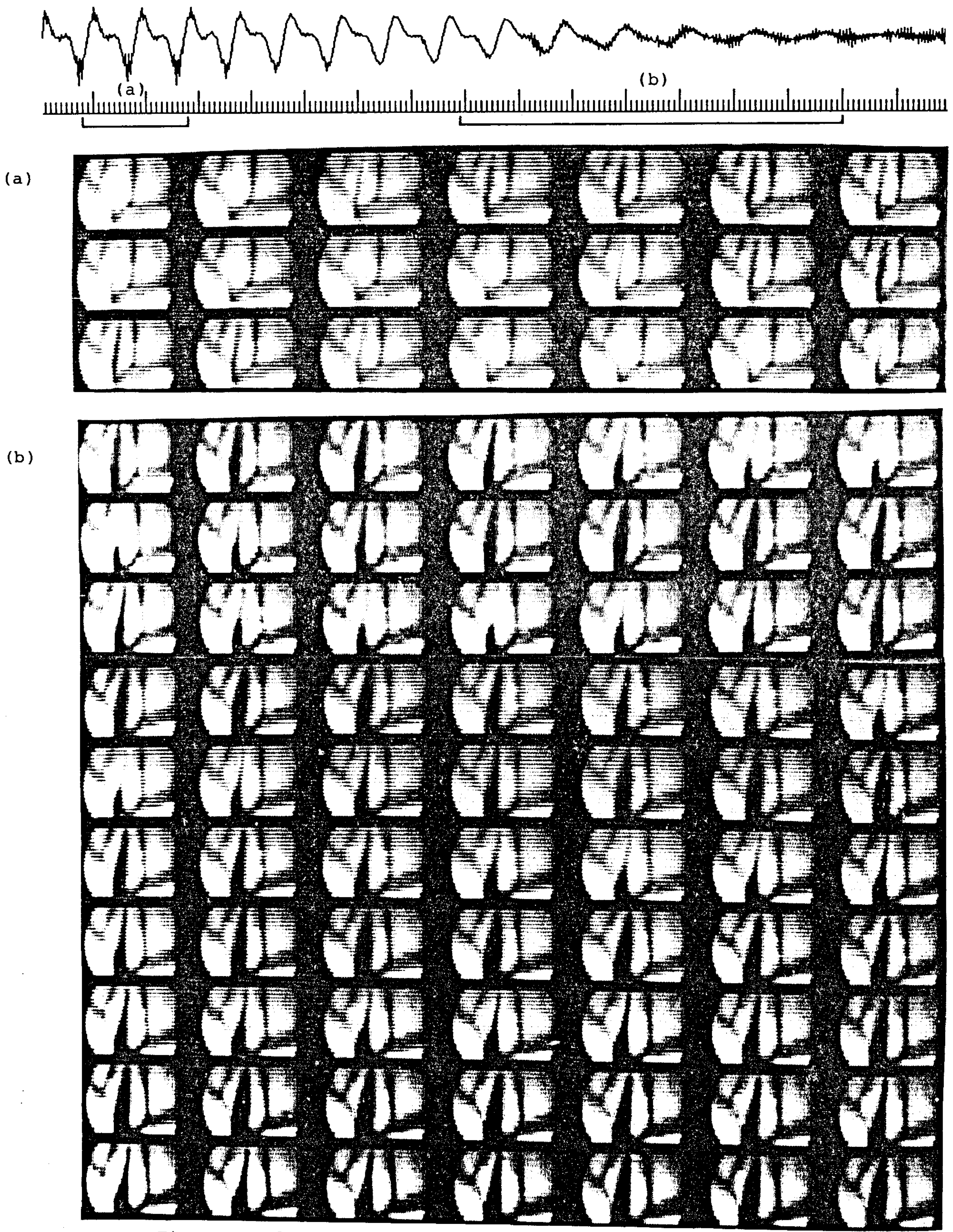


Fig. 4 Vocal cord vibration during transition from vowel to /h/ in the utterance /pi:hi:/.

Table 2 Basic characteristics of the fiberscope system.

Fiberscope	diameter 4.8mm view angle 43_ distance 7-70mm
Light source	500W xenon lamp
Image sensor	CCD type Clock 10MHz
Image memory	768k byte

Scan elements	Frame/sec	Storage
200x34	1000	110
200x14	2000	270 (frames)

diameter of which was slightly larger than that of the ordinary scope. At the same time, a CCD type image sensor was employed in this system, because the image by the new fiberscope was still darker than that by the solid endoscope,. The sensitivity of the CCD image sensor is generally higher than that of the MOS type imagr sensor which was used in the solid endoscope system. However, the comercially available CCD sensors generally contain a large number of picture elememts, 500x500 for example. Thus, in order to realize a high frame rate, it was necessary to develop a special scan method to sample only a very limited portion of the image sensor and to reduce the number of sampled elements.

Table 2 summarizes basic characteristics of the system. The light source is a 300W xenon lamp. The sampling rate of the picture elements is 10Mhz. The period of image recording is about 100ms long. For the picture elemnts of 200x34, the frame rate is 1000 per second. A frame rate of 2000 per second can be achieved with the picture elements of 200x14

By using this system, preliminary tests on the recroding of vocal cord vibrations during consonants were performed. A special triggering method was employed to record the glottal images for the selected period of the consonant in the VCV utterances. First a pulse from the camera shutter sets the entire recording circuit ready. Then, the subject start the utterance. When the beginning of the speech envelope is detected, a trigger pulse is generated. Actual sampling of image signal is started with a delay of the specified interval. By using an appropriate delay time, it is possible to record the glottal images during the selected consonantal period in the VCV utterances.

Fig. 4 shows an example of images recorded by the fiberscope system at a rate of 2000 frames per second. Vocal cord vibration during the transition from the vowel [i] to the consonant [h] in the utterance [pi:hi:] is shown. Fig. 4 (a) represents the stationary vibration during [i:]. During the transition period shown in Fig. 4 (b), the arytenoid cartlidges gradually seperate and the maximum glottal opening is getting larger. At the later period in the figure, the vocal cords are still vibrating but the right and left vocal cords do not contact. It can be seen in the speech wave that, corrsponding to the observed vocal cord vibration, there is a modulation of the amplitude of the noise in the speech signal.

Summary

We have developed a new method of high-speed digital image recording system using laryngeal endoscopes and image sensors. The system is compact and simultaneous recording of speech and other physiological signals can be performed very easily. The system using a fiberscope realized observation of the vocal cord vibrations during consonants. We believe that the system is useful for the study of the voice source characteristics in various mode of phonation.

References

1) K. Honda, S. Kiritani, H. Imagawa and H. Hirose: "High-speed digital recording of vocal fold vibration using a solid-state image sensor.", in Laryngeal Function in Phonation and Respiration, Acollege-Hill Publication, 485 (1987).

2) S. Kiratani, K. Honda, H. Imagawa and H. Hirose: "Simultaneous high-speed digital recording of vocal fold vibration and speech signal.",Proc. ICASSP, Tokyo, 1633 (1986).

MICROPHONEMICS - HIGH QUALITY SPEECH SYNTHESIS BY WAVEFORM CONCATENATION

Konrad Lukaszewicz

Inst. of Biocybernetics and
Biomed. Eng., PAN, Warsaw
Poland

Matti Karjalainen

Helsinki Univ. of Technology
Acoust. Lab., Otakaari 5 A, Espoo
Finland

ABSTRACT

Speech synthesis by waveform concatenation has been the subject of many attempts with fairly low quality results. We have reformulated the approach and found its potential to natural and personal-sounding speech by rule-based synthesis. Our study in Finnish and Polish shows that the method called *microphonemics* could be implemented by standard micro-processors and D/A-converters without any expensive signal processing hardware.

The main problems to be solved in the microphonemic method were the interpolation of pitch-sized phoneme and allophone units in wide formant transitions, the synthesis of fricatives and some other consonant classes, and the control of pitch and intonation. We found that the waveform interpolation works if the formant transitions are narrower than 2 Barks (critical bands), which implies the use of intermediate units in wide transitions. Fricatives are realized by time-randomized selection of 10 ms signal units from 50 ms unvoiced prototypes. Pitch and intonation problems can be solved by several windowing techniques in the formation and concatenation of pitch-sized units. The paper describes our experiments and proposes synthesis-by-rule strategies for implementation.

INTRODUCTION

The methods of speech signal generation in speech synthesis are often divided into two main classes: *model-based* source-filter models (formant and LPC-synthesis) and *waveform-based* time-domain synthesis methods. The advantage of model-based synthesis is the flexibility of generating an infinite number of signals according to parametric controls that can be computed by rules, tables etc. This has become the major method especially in speech synthesis by rule.

Time-domain synthesis can be based on a collection of varying sized speech signal units like waveform cycles, pitch periods, sound segments, phonemes, diphones, syllables etc., taken from real speech. Concatenation of speech samples is a simple method that has been used in synthesis experiments of low to moderate quality. In principle the sound quality could be very high if it were possible for enough samples of natural speech to be stored and carefully combined. This method takes more memory than the model-based synthesis but otherwise it is not as complex and arithmetically intensive.

A well known example of time domain speech synthesis is the *Mozer* method /1/, where pitch-period-sized prototype units of real speech are manipulated to take as little memory as possible but are still able to be reconstructed in an intelligible form. This moderate quality, low bit rate method is used in some limited vocabulary synthesizers. Our experiments show that the tricks like zero-phasing the signal to lower the bit rate tend to remarkably reduce the quality and speaker identity. The phase properties are important to be retained for very high quality, natural sounding speech in a similar way as in multipulse LPC-coding /2/.

The term **"microphonemic method"** that is used in our study was adopted from early experiments of similar principles in Poland. *Patryn* /3/ synthesized with phonemic units without transitions and pitch changes. His work was continued by *Kielczewski* in his doctoral thesis (1979). This microphonemic method applied pitch changes for intonation and transitions by mixing parts of neighbouring phoneme prototypes. *Lukaszewicz* et al. have worked on the method at the Institute of Biocybernetics, Warsaw, since 1980. Their synthesizer has found applications in a talking typewriter and a talking calculator.

The quality of speech in all of these synthesizers has been low to moderate. The objective of our study was to find methods to overcome the inherent difficulties in concatenating speech waveforms. Our experiments show that it is feasible to develop simple and inexpensive synthesizers with natural and high-quality human-like characteristics. This concerns also speech synthesis by rule with unlimited vocabulary.

PROBLEMS TO BE SOLVED IN THE USE OF SPEECH WAVEFORM CONCATENATION

The microphonemic method is based on modelling the time domain signal by using a dictionary of prototypes. These are derived from natural speech utterances and their size can be of different lengths. It is possible to store whole words, syllables, phonemes (allophones) or shorter segments. Using a dictionary of microphonemes and several rules it is possible to generate synthetic speech by concatenating prototypes one after another. Waveform interpolation and concatenation are applied to realize the transitions between consecutive units. There are several problems that need to be solved in order to obtain high quality synthetic voice, e.g.:

* realizing dynamic and static variations of the units, especially in the generation of smooth and natural transitions between consecutive segments and phonemes,
* synthesizing consonants, like tremulants (Finnish /r/), etc.,
* modifing parameters to control intonation, stress and rythm,
* determining the prototype set which is needed for a good representation of natural speech,
* extracting these prototypes and their positions in the uttered speech examples,
* formulating a good strategy when using waveform concatenation for synthesis by rule.

Some of these problems were studied by us at the Helsinki University of Technology, Acoustics Laboratory, by using the following experimental techniques.

WIDE FORMANT TRANSITIONS

The first problem to be solved in high quality waveform concatenation is to realize formant transitions e.g. in diphthongs (like /ui/ in Finnish) and in glides. The original idea of the microphonemic method was to apply simple linear interpolation from one pitch prototype to another by amplitude mixing (see Fig. 1). In our experiments we found that this works satisfactorily only if the glide in formant frequencies is less than 2 Barks (critical bands). In wider transitions, the amplitude-based interpolation is not sufficient to introduce a perceptually acceptable formant movement effect. For highest quality speech even 1 Bark transitions may be needed.

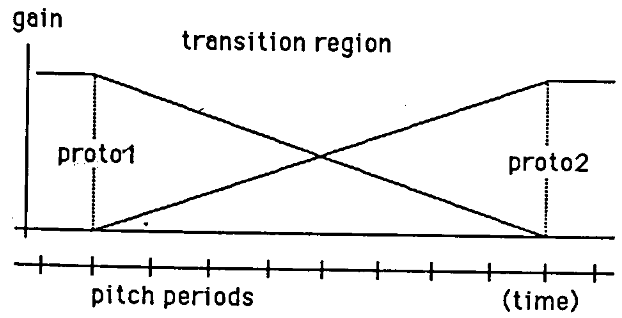


Fig. 1. *Linear amplitude-based interpolation between two pitch-sized prototypes to simulate formant transitions.*

If the formant distances between sound segments larger than 2 Barks are needed, some intermediate prototypes should be used to interpolate through (see Fig. 2.). It was possible for all transitions found in Finnish and Polish to be synthesized in this way.

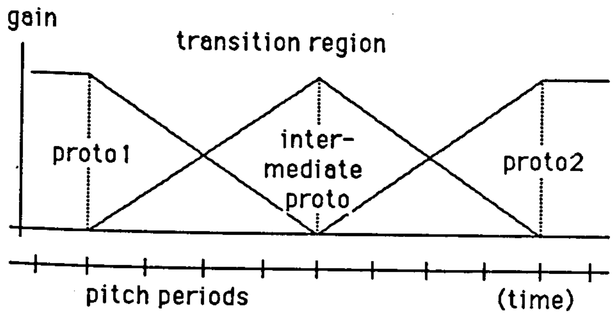


Fig. 2. *Linear amplitude-based interpolation between two pitch-sized prototypes with an intermediate prototype.*

SYNTHESIS OF CONSONANTS

Many consonants need special processing. Short non-repetitive units like bursts in stop consonants can be stored as direct waveform segments and as several variants in the context of different vowels or vowel groups. Sometimes the effect of neighbouring consonants must also be analyzed and the context stored for synthesis.

Fricatives need special treatment, too. Prototypes of about 50 ms in total length were found to be suitable and 10 ms units from them were randomly taken for concatenation. The same interpolation rule as in vowels can be applied. Most voiced consonants behave in the same manner as vowels except that the variability according to the context is only higher.

PITCH AND INTONATION CONTROL

Prosodic features reveal some difficulties in concatenation. A simple and fairly successful method to control intonation is the use of minimum-pitch-period-sized prototypes and insertion of zero-signal segments to obtain the desired effective pitch for each moment (Fig. 3). A suitable windowing technique and the overlapping mixing of pitch periods could improve the results still further (Fig. 4). Timing is controlled by counting a proper number of pitch periods.

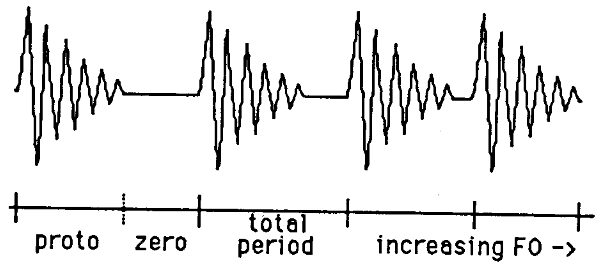


Fig. 3. *Zero signal insertion as a method of controlling pitch in concatenation.*

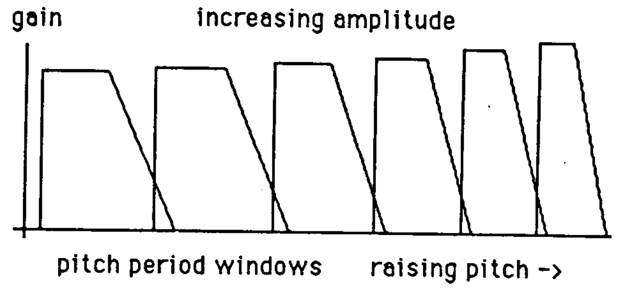


Fig. 4. *Overlapping window summation in pitch control.*

EXPERIMENTAL STUDY

About 70 Finnish and Polish phoneme pairs, concentrating on the synthesis of diphone-like transition segments, were studied experimentally by the microphonemic method. Some other larger units (syllables, words) were also modeled.

A microprocessor-based signal editor (SPS-02) was used to extract the prototype units from real speech. The same editor system was further applied to scale the amplitude, adjust the pitch period and to mix the prototypes for concatenation and synthesis experiments. Another analysis system, ISA /4/, with auditory spectrum and spectrogram display was used to pick up the best positions of the prototypes and to compare the original against the synthetic speech examples. The principle of the auditory model for this analysis is presented in /5/.

Prototypes from the original speech were used to model the phoneme pair transitions with two different principles of prototype selection. The first one was for producing intelligible, moderate quality speech with a minimum number of prototypes which were located in the middle of the quasiperiodic steady-state phonemes and one prototype in the middle of the transition.

The other method was to produce higher quality speech with a larger number of prototypes. This was accomplished by choosing the prototypes at each point where the formant frequencies started to change. If the change was larger than 2 Barks an extra prototype between the starting and ending points was taken. The maximum difference in any formant transition to be interpolated was always less than 2 Barks. A prototype was selected also at the points where the formants changed their direction of movement. For a full synthesis system some of the intermediate prototypes may be selected so that they can be used in several contexts.

As an example, the number of prototypes in the Finnish diphthong /ia/ was three for intelligible and seven for high quality speech. The maximum number of prototypes was never larger than nine for any diphone-like unit. The size of a prototype was usually equal to one pitch period. However, in the case of stop consonants the length of a prototype was two to five times longer and for fricatives five times longer.

Fig. 5. shows the auditory spectrogram of the original diphthong utterance /ia/ with the related loudness function. Vertical lines with the capital letters A to C mark the positions of the prototypes in the lower-quality experiment. The auditory spectrogram of the synthesized version is shown in Fig. 6. Lines related to digits 1 through 7 in Fig. 5 indicate the places of the prototypes in the case of the highest quality reconstruction. The corresponding auditory spectrogram is in Fig. 7.

The pitch-sized prototypes from real speech naturally inherit some speaker-specific features and personality of the voice. The time-domain signal carries the tone quality features related to the detailed amplitude and phase spectrum. Our experiments show that the phase, especially rapid phase transitions can be very important to the naturalness of some allophones (nasals, liquids etc.) and their combinations. The prototypes may also include inherent pitch and amplitude data of the allophones that will be modified according to the context during the resynthesis.

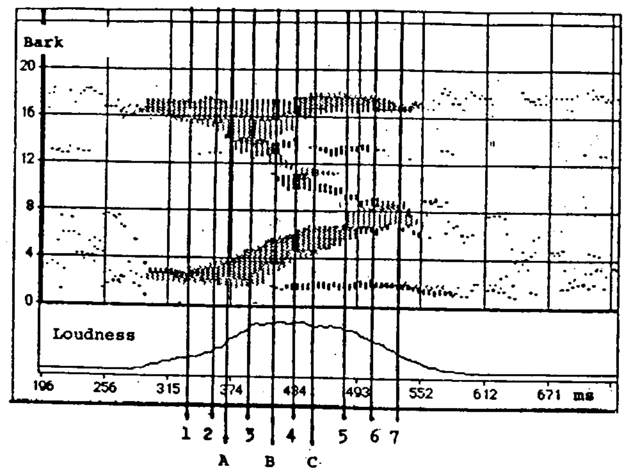


Fig. 5. *Auditory spectrogram of the original speech, (Finnish /ia/)*

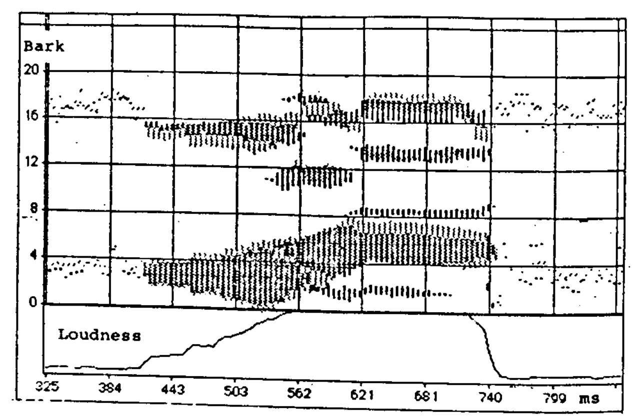


Fig. 6. *Auditory spectrogram of the lower-quality reconstruction by the microphonemic method with three prototypes (A, B, C in Fig. 5.)*

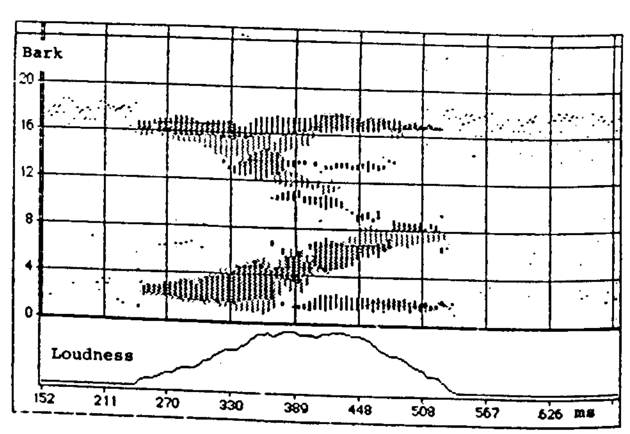


Fig. 7. *Auditory spectrogram of the higher-quality reconstruction by the microphonemic method with 7 prototypes (1 to 7 in Fig. 5.)*

MICROPHONEMIC SYNTHESIS BY RULE

Lukaszewicz et al. have implemented a low-to-moderate quality rule-based microphonemic syntesizer in Polish with some practical applications. The objective of the present study was to find the feasibility of the microphonemic method in high-quality synthesis by rule. Because of the relative high storage required and the tedious prototype preparation the method is not as well suited to limited vocabulary synthesis.

The synthesis process in microphonemics consists of the concatenation of precompiled prototype units with some context dependent modification rules applied to the prototypes. When compared to traditional model-based parametric synthesis this means more like operating with discrete symbol-like units instead of continuous-time control parameters. Some arithmetic and numeric computation can be avoided in this way.

The higher levels of text-to-speech synthesis transform the imput text to a string of phonemic symbols. This process is very language dependent. In Finnish it is almost a one-to-one mapping from grapheme string to phoneme string with some prosody control analysis /6/, while e.g. in English it is a much more complicated task. The assembly of microphonemes into speech signals by phonemic level control information follows the same guidelines in all languages. A set of rules defines how the prototypes are to be modified and concatenated and how the prosodic control information is taken into account.

In our semimanual experiment we used a special notation to describe the assembly of microphonemic units. The following forms were used:

x	means a phoneme
.	is a transition region
x_1x_2.	is one prototype (one pitch period) which was taken from the beginning of a transition between x_1 and x_2.
$x_1.x_2$	one prototype taken from the middle of the transition between x_1 and x_2
$.x_1x_2$	one prototype from the end of the transition between x_1 and x_2
n(...)	integer to show the number of repetitions of some units, e.g. $5(x_1.x_2)$
n(-)	n (integer) periods of linear interpolation of two neighbouring prototypes, e.g. 4(-), - equals to 1(-).

By using this notation we can express phoneme strings in the way of the following (Finnish) examples:

/ai/	->	12(.@a) 5(-) a.i 5(-) 10(i.@)
/anna/	->	17(.@a) 15(.an) 9(na.) - n.a - 15(a.@)
/olli/	->	@.o 5(-) ol. 5(-) 6(.ol) 5(-) 6(li.) 3(-) .li - 8(i.@) 3(-) 10(.i@)

where symbol @ denotes space (pause).

This notation could be developed towards a formal rule language to be used in the implementation of the rule-based synhesis. It should also be possible to express the prosody-related control information, durations of the concatenated units (instead of counting periods), relative pitch and amplitude, special effects etc. To do this, the basic unit in the language could be an event object that contains fields or slots for different properties and relations.

The automatic generation of speech from phonemic code could proceed as follows. A rule-based match of the phoneme code to a set of templates is carried out to give the best candidate string of allophonic units and corresponding microphonemic prototypes. Slot values related to prosodic features are filled based on context-dependent prosody rules. An experimental study of this kind is under development.

IMPLEMENTATION ASPECTS

An estimate of the memory capacity that is needed for prototypes in a moderate-quality synthesizer (Finnish) is: some 30 "phonemes", in average 8 variants (vowel contexts), and the same amount of intermediate prototypes. This results in a total number of about 500 units, each of 12ms in duration times 14 samples/sec (8 bits), which amounts to less than 100 kilobytes. At the level of present ROM-memory technology it is feasible to use up to 256 kbytes of memory for the prototype storage and synthesis rules, thus achieving high-quality synthetic speech with personal-sounding voice.

A single microprocessor like the Motorola 68000 is capable of doing this synthesis in real time. Serial and/or parallel ports are needed for input and a single D/A-converter (8 to 12 bits) with a reconstruction filter may be used to form the analog output. Another possibility is to design with multiplying D/A-converters so as to avoid software multiplications for amplitude scaling in the interpolation. The microphonemic method is also well suited to software-based speech synthesis in microcomputers with special D/A-hardware to support fast analog output. The software for the microphonemic synthesis by rule can be based on the manipulation of prototypes along the guidelines stated earlier.

The selection of the prototypes during the development of the system is a laborious and critical task that is difficult to be automated. A semiautomatic segmentation algorithm and pitch period detector could help if the voice of several speakers must be modeled. We are working to create two different development systems to continue the studies on the microphonemic method. One will be based on a personal computer, another in an artificial intelligence programming environment.

CONCLUSIONS

Our experiments showed clearly that the microphonemic method by waveform interpolation and concatenation has potential for high-quality speech synthesis by rule. Its main technical advantage is that no computationally intensive signal processing is required. To achieve the highest-quality results optimal extraction of prototype segments from real speech and a good strategy for rule-based concatenation is needed. Auditory spectra and spectrograms were found important in the extraction process to find the best segments that meet the requirements of human auditory perception.

REFERENCES

/1/ **Costello B.C., Mozer F.S.**, Time-Domain Synthesis Gives Good-Quality Speech at Very Low Data Rates. Speech Technology Sept/Oct. 1982, p. 62-68.

/2/ **Atal B.S., Remde J.R.**, A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates. Proc. of ICASSP-82, Paris, p. 614 - 617.

/3/ **Patryn R.**, Transitionless Synthesis of Speech. Acoustica Vol. 48 (1981) no. 4, p. 275-276.

/4/ ISA, Intelligent Speech Analyser, Instruction Manual, Vocal Systems, Finland, 1987.

/5/ **Karjalainen M.**, A New Auditory Model for the Evaluation of Sound Quality of Audio Systems. Proc. of ICASSP-85, Tampa 1985, p. 608-611.

/6/ **Karjalainen M.**, An Approach to Hierarchical Information Processes with an Application to Speech Synthesis by Rule. Ma 29, Acta Polytechnica Scandinavica, Helsinki 1978.

High-Quality Speech Synthesis Using Demisyllables and a Variable-Frame-Rate Vocoder

Wolfgang Hess

Institut für Kommunikationsforschung und Phonetik, Universität Bonn
Poppelsdorfer Allee 47, D-5300 Bonn 1, W. Germany

Abstract. *An experimental high-quality speech synthesis system is described. Demisyllables are used as phonetic units for concatenation; in a first step it is shown that 1665 demisyllables requiring about 0.5 MByte of memory at a data rate of 7.2 kbit/s are sufficient to synthesize a very large German vocabulary. To generate the output speech signal, a special variable-frame-rate vocoder synthesizer is implemented.*

1. Introduction

Text-to-speech synthesis systems principally consist of three major components: 1) an orthographic-to-phonetic transcription (including prosody control); 2) the concatenation block; and 3) a vocoder synthesizer. Usually the output of the orthographic-to-phonetic transcription block is a string of phonetic symbols plus a number of special characters for prosody control. The concatenation component converts this string into a data stream of vocoder parameters which are then transformed into synthetic speech by the vocoder synthesizer.

In the last years work on speech synthesis has concentrated upon higher-level tasks, i.e., orthographic-to-phonetic transcription and prosody control. Nevertheless, there are still a number of unsolved problems in connection with concatenation and even with the vocoder synthesizer; due to these problems, the quality of synthetic speech may still be unsatisfactory even for synthetic utterances with a well-modeled prosody. This paper deals with possibilities of improving the quality of synthetic speech by optimizing the concatenation block (Dettweiler, 1981, 1984) and by designing a vocoder that is particularly well adapted to a speech synthesis system by rule (Heiler, 1982, 1985).

2. Concatenation System for Demisyllable Elements

Concatenation is a central problem in any system for speech synthesis by rule. It provides the link between the phonetic level and the parametric level of the system. In practice concatenation is controlled by a set of rules that act upon a data base of speech data. This data base may contain experimental data, such as tables of formant frequencies; however, it may also consist of (parameterized) natural speech. The design of the concatenation component is determined by a tradeoff between the number and complexity of the concatenation rules on the one hand and the size of the memory required for the data base on the other hand. The crucial question in this respect is that of the phonetic units to be applied.

2.1 The Demisyllable Approach

Besides phonemes and diphonemes, syllabic units supply a viable data base for high-quality synthesis by rule. The influence of coarticulation strongly diminishes when a syllable boundary is crossed (Fujimura, 1981; Öhman, 1966). When syllabic units are used, the number of elements is minimized when the syllables are split up into *demisyllables* (DSs). Demisyllables as units of speech processing were first proposed by Fujimura both for speech recognition (1975) and for synthesis purposes (1976). For German DSs were taken up by Ruske and Schotola (1978) for a speech recognition system; for synthesis by rule they were first used by Dettweiler (1980, 1981).

Usually a syllable is defined to consist of the syllabic *nucleus* (in German this is always a vowel or a diphthong) which is preceded and followed by a number of consonants, the so-called *consonant clusters* (CCs). The consonants preceding the syllabic nucleus form the *initial consonant cluster*, and the consonants following the nucleus represent the *final consonant cluster*. A syllable is subdivided into demisyllables by cutting it within the syllabic nucleus. The initial CC and the beginning of the syllabic nucleus form the *initial demisyllable*, whereas the remainder of the nucleus and the final CC make up the *final demisyllable*.

2.2 The DS Inventory. Synthesizing Monosyllabic Words

A representative DS list for German was compiled by Ruske and Schotola (1978; cf. also Schotola, 1984). The initial CCs contain from 0 to 3 consonants, whereas up to 5 consonants may exist in a final CC. The number of CCs is rather limited due to linguistic constraints: we have to deal with only 51 initial and 159 final CCs (Dettweiler, 1984). Concerning the syllabic nuclei, 23 vowels and 3 diphthongs must be taken into account.

Contrary to speech recognition, where the syllabic nuclei and the CCs can be treated separately (Ruske and Schotola, 1978), the transitions between the syllabic nuclei and the CCs are essential for the quality of the synthesized speech; they cannot be generated by rule and must be available as stored data. For the complete DS inventory the number of elements thus becomes

$$N_c = 26 \cdot 51 \text{ initial DSs} + 26 \cdot 159 \text{ final DSs} = 5460 .$$

Since coarticulation has a strong tendency toward anticipating future articulatory gestures (Delattre, 1968; Fujimura, 1981), it is adequate to establish the DS boundary within the first part of the vowel. Fujimura's proposal (1976) to place the boundary 50 ms after the beginning of a vowel is also applied in our system (Dettweiler, 1981; cf. Fig.1).

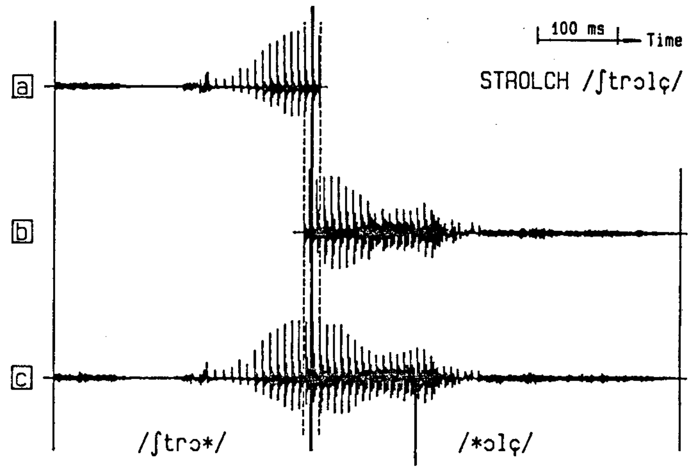


Fig.1a–c. Concatenation within the syllabic nucleus (rule CR1). (a) Initial DS, (b) final DS; (c) complete word after concatenation. The thick vertical line indicates the interconnection point; the smoothing interval is indicated by the dashed lines. The asterisk in the phonetic transcription refers to the position of the syllabic nucleus

2.3 Inventory Reduction

To reduce the number of DSs, two ways seem feasible: 1) vowel substitution, and 2) further splitting of CCs. Both these possibilities have been used in our system; the most important rule being the principle of rudiment and suffix (Dettweiler, 1981; Fig.2).

Certain consonants, when occurring in final position of a DS, may be split off from the DS and form separate units, the so-called affixes (Fujimura et al., 1977). As the experiments suggest, fricatives and stops in final position, like vowels in the syllabic nuclei, represent a natural coarticulation barrier; i.e., sounds following this barrier do not (substantially) affect previous sounds. A splitting scheme which is particularly efficient for German is the principle of *rudiment* and *suffix* (Dettweiler, 1981, cf. Fig.2). A suffix is defined to consist of any (existing) combination of the four consonants /f/, /s/, /ʃ/, and /t/, whereas the remainders of the final DSs form the rudiments. The linguistic constraints of German state that once a suffix consonant, i.e., one of the 4 consonants named above, has occurred in a final CC, the following consonant(s) of that final CC, if existing at all, must be suffix consonants as well.

In practice the rudiment is formed by uttering a DS that contains the remainder of the consonantal cluster (without any suffix consonant) plus a final /t/ and then removing the /t/ together with the pertinent silence before the burst (Fig.2b). Since the rudiment contains all the coarticulatory influences by the following /t/, it is easy to see that the rudiment and the final DS containing an identical consonant cluster without the /t/ are different (cf. Fig.2a,b). Any rudiment and any suffix may be simply concatenated without any smoothing at the interconnection point.

Using all these possibilities of inventory reduction, the total number of elements now decreases to N_R = 1665. Note that these inventory reductions do not degrade the quality of the synthetic speech.

With an average duration of 0.3 s per element, the memory required for this inventory is less than 0.5 MByte if a vocoder at 7.2 kbits/s is used.

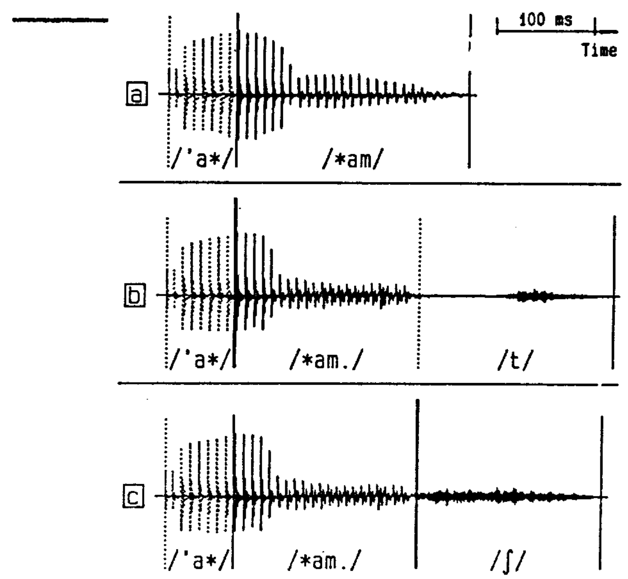


Fig.2a–c. The principle of rudiment and suffix. (a) 6 Ordinary consonant cluster: example /*am/. (b) Rudiment and suffix: the DS /*amt/ is split up into the rudiment /*am./ and the suffix /t/ (the dotted line represents the boundary). (c) Concatenation using rudiment and suffix: /*am./ ¦¦ /ʃ/ —> /*amʃ/. Signals drawn with dotted lines represent DSs that are needed to complete the word, but do not pertain to the DSs involved in rule CR2

2.4 Synthesizing Polysyllabic Words

Polysyllabic words contain *intervocalic consonant clusters* between subsequent syllabic nuclei. This requires additional rules for the concatenation of CCs (Dettweiler, 1984). The procedure is carried out in two steps. First an intervocalic CC is split up into a final CC followed by an initial CC, and the CCs are joined to the respective syllabic nuclei to form DSs. In the second step the DSs are concatenated.

The ICCs are split according to three rules. Firstly, an intervocalic CC must always be split up into a *valid* final CC and a *valid* initial CC. A CC is regarded as *valid* if it is contained in the DS inventory. If this rule does not yield a solution, the DS inventory must be enlarged. On the other hand, if this rule provides several solutions, a second rule states that the one solution is selected where as many consonants as possible are grouped within the initial CC. This "pragmatic" boundary takes into account the anticipatory effect of coarticulation; even when a DS boundary as established by this rule, differed from a given morph boundary. These two rules thus represent an adequate means to split up intervocalic CCs without requiring morphologic knowledge at this level.

When the intervocalic CC only contains one consonant, a third rule switches the system into a diphone mode by assigning this consonant to both the initial and the final DSs.

The way in which intervocalic CCs are concatenated strongly depends on the consonants involved. Due to lack of space, the individual rules cannot be discussed here. A flow diagram is depicted in Fig.3; the labeling of the concatenation rules (CR 3-12) corresponds to that in (Dettweiler and Hess, 1985); for an in-depth discussion, the reader is referred to that publication.

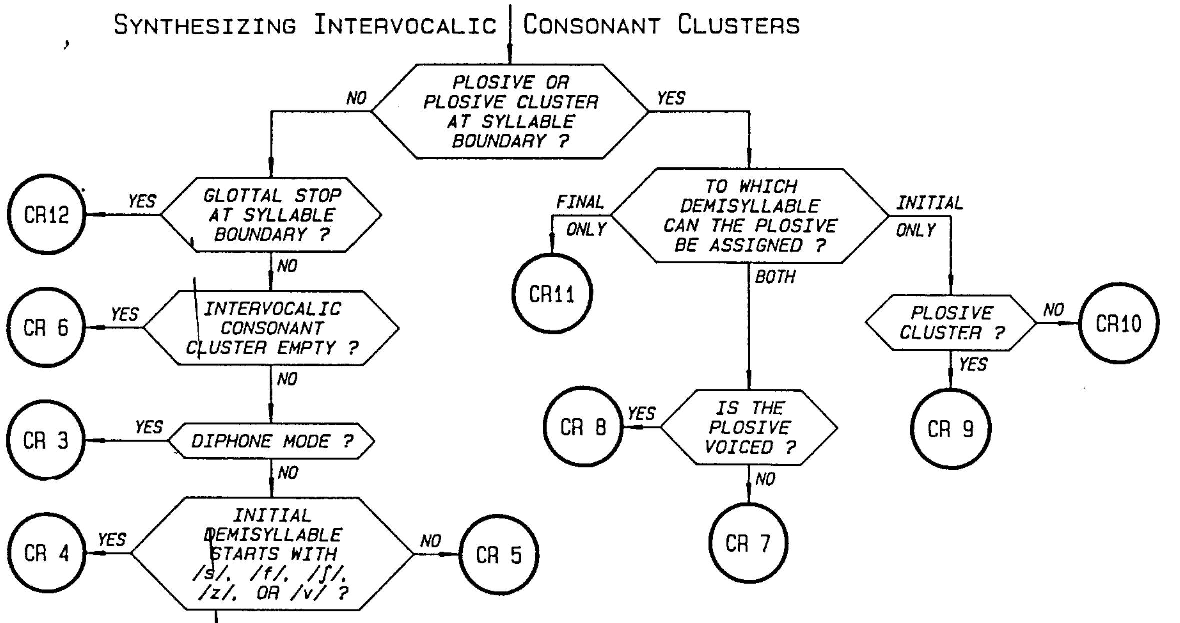


Fig.3. Block diagram of the concatenation step for intervocalic consonant clusters. The labeling of the concatenation rules (CR3 to CR12) corresponds to (Dettweiler and Hess, 1985)

2.5 Realization of an Experimental System

An experimental system was implemented using a 12th-order standard LPC vocoder with a constant frame interval of 10 ms and a signal sampling frequency of 10 kHz. For data acquisition the DSs were embedded in two-syllable meaningless words of the form /<initial DS>tər/ and /gət<final DS>/; the DSs were manually delimited using a display program and an interactive segmentation procedure.

Compared to systems that use phonemes as units for concatenation, the number of concatenation rules in this system is extremely low, and the quality is much better. In an intelligibility test, Dettweiler (1984) showed that the median word intelligibility dropped from 96.6% for vocoded speech to 92.1% for the DS synthesis system (using the same vocoder). The quality of the vocoder speech and the demisyllable synthesis system were judged to be almost the same. However, there were still a number of systematic confusions of fricatives, such as /s/ and /f/; some other errors were due to incorrect segmentation of the demisyllables. Systematic errors due to the concatenation rules were not encountered.

3. High-Quality Variable-Frame-Rate Vocoder

Since Dettweiler's experiments (1984) showed that even a rather high-quality vocoder may be a source of systematic intelligibility errors, it is useful to redesign the vocoder and to adapt it to the special requirements of the synthesis system by rule.

Usually vocoders are designed for the purpose of parametric speech transmission. Important criteria are robustness, good performance even in adverse environment, and real-time operation. If a vocoder, however, is to be optimized with respect to speech quality at a given transmission rate, the principle of variable frame rate (VFR) vocoding best fulfils this requirement (Huggins et al., 1977). In a VFR vocoder the frame rate is adapted to the speed of articulatory movement; i.e., frames are selected and transmitted in rather large time intervals during stationary segments such as vowels, whereas the frame rate during rapid transitions is rather high. This principle is not well suited for transmission purposes since it is rather complex in the analysis part; in addition, it may introduce a substantial processing delay which is intolerable in a dialog. In a speech synthesis system by rule, however, the requirements are substantially different. Analysis of the input data is done off line (even manual interaction may be permitted) and with high-quality data; thus the analysis algorithms may be sensitive and complex. In addition, the question of the processing delay is irrelevant here. Such a system is thus extremely well suited for this kind of application.

Heiler (1985) developed and optimized the so-called "evolution strategy" (Fig.5) for selecting the frames and approximating the parameters to be stored. In this algorithm an utterance (e.g., a DS) is regarded as one unit. Predetermined are 1) the number of frames to be selected, 2) the interpolation procedure for the frames which are not selected (a combination of linear interpolation and simple repeating of the most recent frame proved to give the best results), and 3) the approximation strategy for the parameters at the selected frames. The algorithms starts with the two frames at the beginning and end of the DS that must be selected. The third frame is positioned in such a way that the accumulated approximation error becomes a minimum over the whole utterance. Then another frame is looked for (with the frames kept constant that were already selected) according to the same criterion; this procedure continues until the desired number of frames have been obtained. Due to the successive approximation, however, the selection of frames is not yet optimal. For further optimization one more frame is now added to the selection. To keep the number of frames constant, the algorithm then removes the one frame whose removal contributes least to the global approximation error. If the removed frame is different from the one which was added in that step, this results in a frame shift; if it is the same, the optimization is terminated.

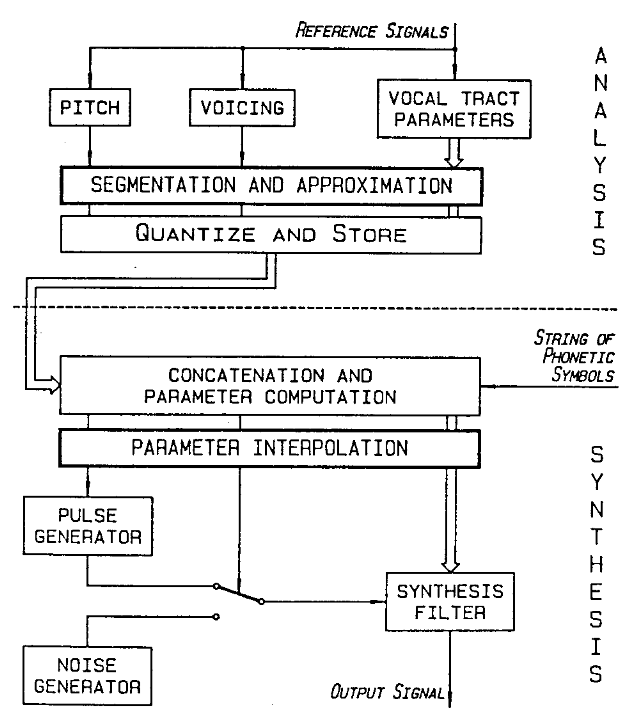


Fig.4. Vocoder configuration for speech synthesis by rule. The analysis (components above the dashed line) is done offline

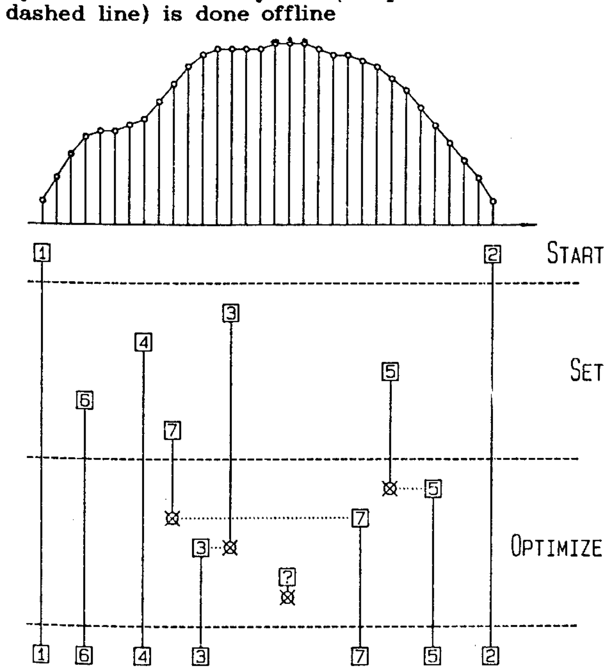


Fig.5. Example for the evolution strategy for a VFR vocoder system. After Heiler (1985)

In subjective listening experiments Heiler (1985) showed that, compared to a vocoder with constant frame rate and no parameter optimization, this VFR principle permits reducing the bit rate by a factor of 3 without a perceptible loss of quality.

4. Discussion and Conclusions

The work descibed in this paper concentrates on quality improvement of text-to speech synthesis by optimizing the front-end steps, i.e., the concatenation block and the vocoder synthesizer.

The use of demisyllables as phonetic units offers the great advantage that about 20 rules and 1650 DSs requiring a data memory of less than 0.5 MByte are sufficient to synthesize (nearly) unrestricted German text. A special variable-frame-rate vocoder synthesizer provides an optimal quality at a given data rate and helps minimizing the required amount of memory.

At the moment the synthesis system by rule and the VFR vocoder still exist as separate units. Efforts are under way to combine the two systems, thus improving the quality of the vocoder in connection with the stored data. A signal bandwidth of 7 kHz requiring a sampling frequency of 16 kHz will eliminate the systematic confusions between the fricatives /f/ and /s/ present in the actual 5-kHz system, and a VFR scheme permitting a minimum frame rate of less than 5 ms without increasing the overall amount of memory will particularly improve the quality of synthetic stop consonants.

Acknowledgement. The major part of this paper was extracted from the Dr.-Ing. dissertations by Dr. H. Dettweiler and Dr. J. Heiler.

References

Delattre P. (1968): "From acoustic cues to distinctive features." Phonetica 18, 198-230 703-706 (VDE-Verlag, Berlin)
Dettweiler H. (1981): "An approach to demisyllable synthesis of German words." Proc. IEEE ICASSP-81, 110-113
Dettweiler H. (1984): Automatic synthesis of German words by means of syllable-oriented segments. Dr.-Ing. dissertation, Technical University of Munich (in German)
Dettweiler H., Hess W. (1985): "Concatenation rules for demisyllable speech synthesis." Acustica 57, 268-283.
Fujimura O. (1975): "Syllable as a unit of speech recognition." IEEE Trans. ASSP-23, 82-87
Fujimura O. (1976): "Syllable as the unit of speech synthesis." Unpublished Bell memorandum (Bell Labs, Murray Hill, NJ)
Fujimura O. (1981): "Temporal organization of articulatory movements as a multidimensional phrasal structure." Phonetica 38, 66-83
Fujimura O., Macchi M.J., Lovins J.B. (1977): "Demisyllables and affixes for speech synthesis." Proc. 9th Int. Congr. on Acoustics, Madrid 1977, paper I107
Heiler J. (1982): "Optimized frame selection for variable frame rate synthesis." Proc. IEEE ICASSP-82, Paris 1982, 586-589
Heiler J. (1985): Minimization of the memory requirements of speech synthesis systems by optimizing the parameter approximation. Dr.-Ing. dissertation, Technical University of Munich (in German)
Huggins A.W.F., Viswanathan R., Makhoul J. (1977): "Speech-quality testing of some variable frame rate (VFR) linear predictive vocoders." J. Acoust. Soc. Am. 62, 430-434
Ruske G., Schotola Th. (1978): "An approach to speech recognition using syllabic decision units." Proc. IEEE ICASSP-78, Tulsa, OK, 722-725
Schotola Th. (1984): "On the use of demisyllables in automatic speech recognition." Speech Commun. 3, 63-87

HIGH QUALITY SYNTHESIS OF VOWELS

Leonid Chudnovsky, Vecheslav Ageyev

Institute for Information
Transmission Problem
Moscow, USSR 101447

ABSTRACT

A problem concerning synthesis of isolated Russian vowels is described. Approximation of excitation source functioning is at the centre of attension.

During vowel synthtsis attention is focused on vocal tract (modes) frequency values. Excitation source is approximated by a triangular function subjected to three jumps of a derivative in pitch period [1]. That approach doesn't provide a high quality synthesis and thus causes intelligibility degradation in additive noises. Somewhat better synthesis results are achieved for a more composite time function of the vowel excitation source [2].

Natural sounding and intelligibility of synthesized vowels can be improved due to taking into consideretion the real features of vowel excitation sources. One may get an idea of the excitation sources from vowel oscillograph traces using the inverse filtering techniques. To solve that inverse problem it is necessary to know such vocal tract parameters as a quality factor and moda frequency. A compensating method based on instant frequency measurement of a filtered speech signal has been used for moda frequency calculation [3]. It was continued by signal frequency filtering in order to extract formant oscillations. Low-pass filters have been used for the extraction of the first formant and band-pass filters for the extraction of other formants. The cut steepness of a filter respounse characteristic has accounted for no less than 48 dB/octave outside the transparence band. A quality factor of the extracted formant oscillations has been calculated using an analytical signal envelope [4]. Algoritm [4] has been modified to improve computing accuracy of a quality factor. After the extraction of formant oscillation $P_K(t)$ and the calculation of the quality factor Q_K and the vocal tract moda frequency ω_K it is possible to regenerate the moda excitation source from the following equation [2] :

$$P''_K(t) + \frac{\omega_K}{Q_K} P'_K(t) + \omega_K^2\left[1-\left(\frac{1}{2Q_K}\right)^2\right]P_K(t) = f_K(t) \quad (1)$$

The excitation source of the formant oscillation $f_K(t)$ is related to the vowel excitation source $f(t)$ in the following way :

$$f_K(t) = \int_{-\infty}^{\infty} L_K(t') f(t-t')\, dt' \quad (2)$$

where $L_K(t)$ - is a filter pulse response for extraction of the K -th formant oscillation.

Equation (1) may be used for speech synthesis as well.

Excitation sources of 5 Russion vowels "a", "э","o","y","и" have been experimentally studied. The extracted excitation sources of the first formant oscillation can be conventionally devided into two groups: the first group for the sounds "a" and "э", and the second group for the sounds "o","y" and "и". The first group of excitation sources represents two successive pulses with different amplitudes with the time interval 4-6 ms and each pulse duration 1-2 ms. The second pulse amplitude and its delay time with respect to the first pulse are related to the quality factor and the first moda frequency in such a way that the second pulse stops its free oscillations which appeared after the first pulse. The second group of excitation source is represented either by a single pulse with 1.5-2 ms duration or by two multy-or unidirectional pulses of the same duration with the second pulse time delay 1.5-2 ms, or by the three pulses of alternating direction with the duration 1.5-2 ms and the time delay 1.5-2 ms and 3-4 ms correspondingly. Excitation source of the sound "y" has one peculiarity. The regenerate excitation sources of the first moda and the extracted signal of the first formant oscillations are identical.

During vowel synthesis excitation pulses have been approximate by the following function :

$$\tilde{f}(t) = \eta(t+T)[1-\eta(t-T)]\cdot exp\{-[1-(\tfrac{t}{T})^2]^{-1}\} \quad (3)$$

where $2T$ -is a pulse duration;
$\eta(t)$ -is a unit function.

The synthesis resulted in high intelligibility of the vowels "a","э","o","y" when represented by single formant oscillation. Increase in the number of formant oscillation causes intelligibility improvement. For acceptable intelligibility of the synthesized vowel "и" it should be represented by two formants. The first mode excitation sources with the reduced pulse duration have been used for higher vocal tract moda frequencies (3). The duretion reduction factor for the K-th moda has been selected equal to the ratio ω_1/ω_K .

Fig.1 shows the excitation sources oscillograph traces of the first formant $\tilde{f}_{1a}(t)$, $\tilde{f}_{1э}(t)$, $\tilde{f}_{1o}(t)$, $\tilde{f}_{1y}(t)$, $\tilde{f}_{1и}(t)$ and of the second formant $\tilde{f}_{2и}(t)$ for the sounds "a","э","o","y", "и".

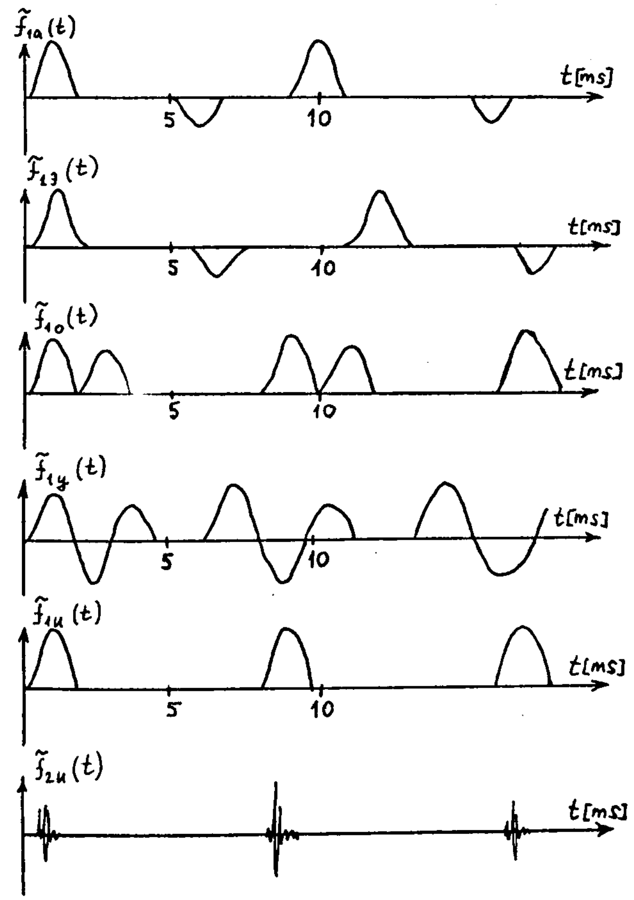


Fig. 1. The excation sources of the synthesized sounds "a","э", "o","y","и".

The oscillograph traces of the synthesized single-formant vowels "a","э","o","y" and the two-formant vowel "и" are shown in Fig.2.

Natural sounding improvemt of the synthesized vowels is achieved with due regard for time variation of the excitation source parameters of each moda $f_K(t)$. Test data analysis has shown that tne vowels excitation sources are subjected to different transformations , i.e. abrupt transformations with the time interval of 30-100 ms and slow period-by-period transformations. The vowels excitation sources which differ in their voice onset time with open or close vocal bands are well differentiated.

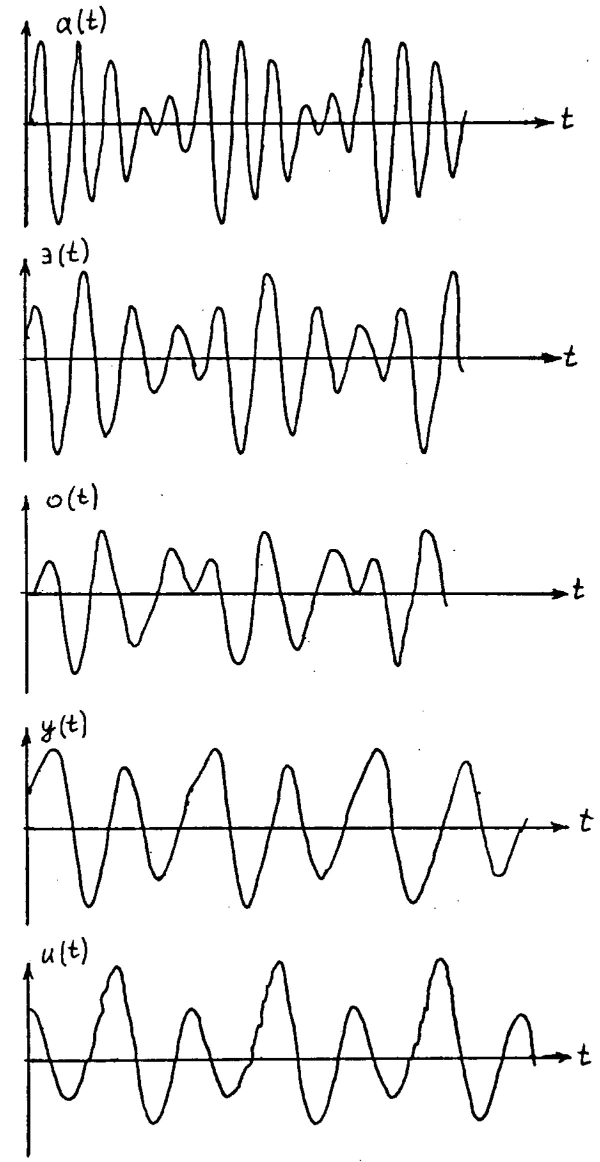


Fig. 2. The synthesized sounds "a", "э","o","y","и".

Fig.3. shows the extracted excitation sources of the first moda $f_{1a}(t)$ and $f_{1и}(t)$ of the vowls "a" and "и". The phonation of the vowel "a" initiates with close and of the vowel "и" with open vocal bands. Due to the extracted excitation source it has been found out that the voice onset time with open vocal bands and the cessation of phonation (Fig.3) have the same time structure and are practically speaker independent. To achieve the vowels high quality synthesis with due regard for the source signal variation the function has been approximated with the help of the tables.

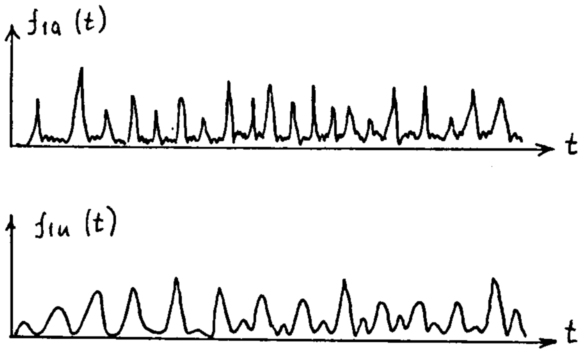


Fig. 3. The regenerated excitation sources of the first moda of the sounds "a" and "И".

The usage of excitation sources peculiarities and their relationship with vocal tract paramters gives an opportunity to achieve the high quality synthesis of vowels and speech as whole.

REFERENCES

[I] Дж.Фланаган "Анализ,синтез и восприятие речи",М.,Связь,I968.

[2] В.Н.Сорокин "Теория речеобразования",М., Радио и Связь,I985.

[3] Л.С.Чудновский, В.М.Агеев "Анализатор речевых сигналов", авторское свидетельство №I275527, МКИ GIO L 9/00,БИ №45 I986.

[4] В.С.Пичкур, А.Ф.Приставка "Определение текущих параметров частотных составляющих речевых сигналов", тезисы докладов и сообщений I3-ой Всесоюзной школы-семинара "Автоматическое распознавание слуховых образов", Новосибирск, I984.

A HARDWARE-SOFTWARE SYSTEM FOR DESIGNING HIGH-QUALITY SPEECH COMPILING SYNTHESIZERS

ALEXANDRE GORODNIKOV

All-Union Research Institute for TV and Radio Broadcasting Moscow, USSR 123298

MEELIS MIHKLA

TOOMAS TAGO

Computer Design Office, Institute of Cybernetics, Tallinn Estonia, USSR 200104

ABSTRACT

Nowadays, the highest possible phonetic quality of synthetic speech can be provided by compiling speech synthesizers. Proposed is an appropriate hardware-software system for their computer simulation and design. As an example of practical implementation, basic parameters of a high-quality speech synthesizer of the "speaking clock" type to be used in radio broadcasting are presented.

INTRODUCTION

Of different types of synthesizers available, it is the so-called compiling synthesizer which guarantees the highest possible quality of synthetic speech, and, consequently, boasts the greatest versatility. The synthesizer is based on a solid-state memory containing speech signals in a digital form. The set of signals consists of specially selected speech elements like phrases, words, syllables, or coarticulation units which, being read out from the memory in a preset order, permit to synthesize a certain number of utterances.

Designing a compiling synthesizer, the key problem is how to compromise among different and even somewhat antagonistic technical requirements, such as the quality of synthetic speech, the volume of the vocabulary, the complexity of the hardware part, dimensions, weight and cost. To provide an effective solution to the above problem, we have developed a hardware-software system that serves well for both research purposes and practical applications in creating compiling synthesizers. The system's hardware also includes a compiling synthesizer of the "speaking clock" type for high-quality speech synthesis.

THE HARDWARE-SOFTWARE SYSTEM

The system is based on a minicomputer EC-1010, operating together with 12-bit A/D and D/A converters, a bank of filters, a tape recorder, and other peripheral equipment. The system's features include: digital input of a speech signal, extraction of the synthesizer vocabulary units from a speech signal, computer simulation of the synthesizer operation algorithms, comparison of different methods of speech signal coding and redundancy reduction, objective analysis and comparison of prosodic characteristics and coarticulation joints of synthesized phrases. It is also possible to prepare and store in the solid-state memory bulks of labelled digital data and to check by listening the acoustic quality of synthetic speech. Sampling frequency of the speech signal input can be - depending on the application - 10,16 or 20 kHz. A segmentation program makes it possible to extract the wanted sentence, word, or syllable from a continuous speech signal. Thus derived speech elements are stored in a database on disks. The next step consists in the analysis and optimization of the vocabulary by means of synthesis. The prosody of a synthesized sentence and the intensity of the speech signal are compared to the corresponding parameters of an originally spoken sentence. According to the context of the sentence, the database is searched for speech elements whose main pitch contour and intensity most closely resemble those of the original sentence.

WORD SELECTION FOR THE SYNTHESIZER VOCABULARY

The highest possible quality of synthetic speech can be achieved in case the vocabulary consists of words and phrases. However, this requires a large-capacity solid-state memory, otherwise the synthesizer shall have a rather limited vocabulary. As an example, we may consider the vocabulary of a high-quality speech compiling synthesizer to be used for time announcement in radio broadcasting (the so-called "speaking clock"). The general structure of the Russian time announcement is as follows: "Moscow time is ... 10 x (hours), 1 x(hours)... 10 x (minutes), 1 x (minutes)" or "It is noon/ midnight in Moscow". Thus, in order to announce time with a minute's precision round the clock one would need 1440 announcements, each structured according to the above pattern and being 4-6 words long. The entire file of speech units used for time announcements would comprise 8592 words with a total duration of 183 min. Obviously, the vocabulary of a "speaking clock" should be considerably smaller in order to provide both a tolerable degree of complexity and a reasonable cost of the synthesizer.

A prosodic analysis of the original time announcemets carried out by means of our hardware-software system showed that abrupt changes in the pitch contour are observed mainly in the middle of the sentence in the words "time" and "(10 x) hours, (1 x)hours", where the pitch rises at the end of the word, and also in the sentence-final position in the words "(10 x)minutes, (1 x)minutes" where the pitch falls. The words carrying quantitative temporal information (i.e. numerals) can be divided into stressed and unstressed ones. In long words, however, changes in the pitch and signal intensity are relatively small, therefore the stressed vs. unstressed dichotomy is not worthwhile in this case. The above findings, alongside with the fact that most of the words display a high repetition rate across different announcements enabled us to considerably reduce and optimize the synthesizer vocabulary. The resulting vocabulary for round-the-clock time announcement service in Russian comprises 43 words with a total duration of 293 sec.

THE COMPILING SYNTHESIZER OF THE "SPEAKING CLOCK" TYPE

Prior to its practical implementation, the "speaking clock" design was simulated and optimized by means of our hardware-software system. The high acoustic quality of the announcements was achieved by 12-bit digital speech conversion with the 16 kHz sampling frequency. In order to economize the solid-state memory storage capacity, speech signals were DPCM-coded. The data-transmission rate was therewith 128 kbit/sec and speech signals were digitally encoded in the format of 8 bits per sample.

Figure 1 represents the block diagram of the compiling synthesizer. There are four main units: an electronic clock and a keyboard controller based on a one-chip microcomputer, a control and display panel, a CPU, and a solid-state memory. The overall dimensions are 475 x 280 x x 440 mm, the power consumption is 60 W.

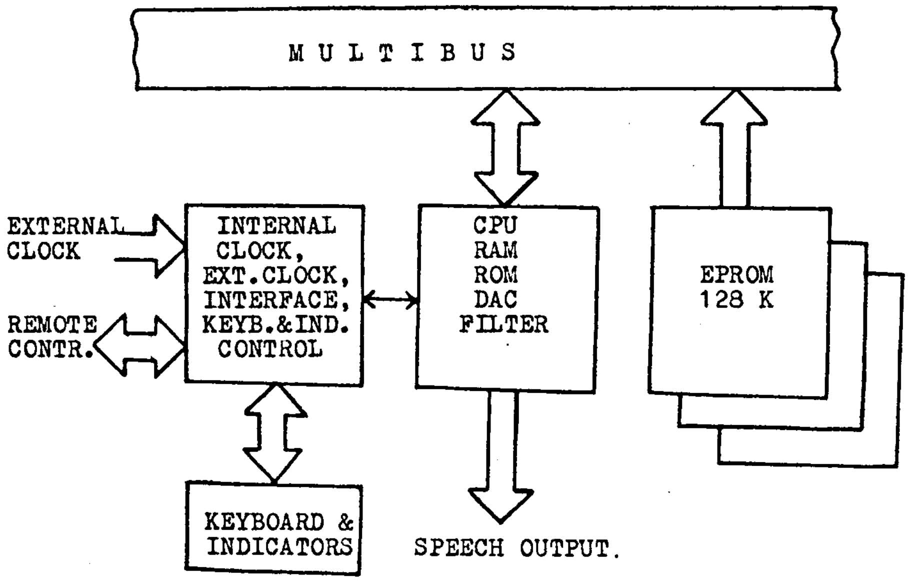


Fig. 1. The structured scheme of the compiling synthesizer.

COMPILATION SYNTHESIS OF SPEECH BASED ON CLIPPED SIGNALS

G.D.FROLOV

16th Parkovaya Str. 27-205
Moscow, USSR, 105484

I.M.YEFREMOVA

Zhukovskaya Ave. 13-75
Zheleznodorozhny 2
Moscow Region, USSR, 143980

The compilation synthesis of speech based on clipped signals is founded on a detailed analysis of the mutual position of zeros of a speech signal.The information used for the synthesis of speech is extracted from the speech signals of a definite speaker.Different minimal items are used for the compilation synthesis - words, syllabels,segments,phonemes.The best diversity of synthesized speech is achieved with tiny items,such as syllabels,segments, phonemes.To obtain high quality synthesized speech it is necessary to do the following preparatory work: define the stressed vowel of the word; define the prosodic features; define the syntactic stress.The preparatory work with the original text is carried out according to the syntactical, grammatical and phonetical rules of the Russian language.

As the pronunciation of a separate letter in the Russian language depends not only on the surroundings but also on its relational position to the stressed vowel, it is necessary to single out,in the original text,groups of words having similar stresses.Such a group of words corresponds in oral speech to a phonetic word.We call it a stressed group.To single out these stressed groups and to define the stressed vowel in them we must prepare the following starting data:

1.Multitude C of permanent components of a word form,where c(i) is an element of multitude C.A permanent component is understood as a most frequent beginning of the grammatical forms of the given word.Each element c(i) is put into correspondence with the number of the stressed vowel of the word (or with the number zero if the number of the stressed vowel changes at changes of the word form) and with a reference to the reciprocal element of the multitude V,described below.

2.Multitude V that consists of variable components of the form of a word.A variabl component of a word form is understood as a complex of all parts of a word form after depriving it of its permanent components. If in a word form there is no variable component,the sign "+" is inserted at its place in multitude V.If the stress in the word form changes,the element of the variable component is put into correspondence with a number that shows the stresse vowel number of the word form.

3.Multitude H of auxiliary words (prepositions,conjunctions,particles) that precede the significant word in the stressed group.

4.Multitude E of auxiliary words that stand after the significant word in the stressed group.This multitude includes only particles.

5.Multitude W of combinations of auxiliary words with significant words,where the sress is put on the auxiliary word.Each element of the multitude W is put into correspondence with the number of the stressed vowel.

For the description of the algorithm of automatic determination of stressed vowels we introduce the following designations: NS - empty word;G(N) - isolated stressed group.The original text is understood as an aggregation of words p(1)...p(i)...p(m).Words are separated by blanks or by marks of punctuation and blanks.Further we bring the algorithm metioned above:

1.Assume: L=1, N=1
2.Check: if L is bigger than m,go to 14
3.Check: if p(L) is an element of H,enter P(L) into G(N) and go to 4,otherwise go to 5
4.Check: if p(L),p(L+1) is an element of W, the number of the stressed vowel is taken from multitude W and go to 13
5.Check: if p(L) is an element of E,enter P(L) into G(N), p1:=p(L-1),go to 6.Otherwise check: if p(L+1) is an element of E, enter p(L), p(L+1) into G(N), L:=L+1, p1:=p(L), if p(L+1) is not an element of E, enter p(L) into G(N), p1:=p(L).
6.Check: if p1 is an element of C,go to 12 take number of stressed vowel from C. Otherwise perform: c1:=p(L)
7.Take the last letter of the word c1 from the right side and add it to the left side of the word v1.
8.Check: if c1 is an element of C,go to 7
9.Check: if v1 is an element of Q,where Q is the sequence of elements of the variable component corresponding with the permanent component c1,go to 7
10.Find the number of the stressed vowel in the word p1 as follows: if in the multitude C the number of the sressed vowel,which is not zero,corresponds with the element c1, then the number of the stressed vowel in the word p1 is found.If the number equals zero, take the number of the stressed vowel from the corresponding element of multitude V.
11.Compute the number of the stressed vowel in group G(N).Add to the number of the stressed vowel of the word p1 the quantity of letters in all auxiliary words that precede the word p1 in the stressed group G(N)
12.Compute: N:=N+1
13.Compute: L:=L+1 .Go to 2
14.End

After the automatic distributionof stresses in the original text has been accomplished,the automatic transcription of th this text is performed. The primary data for the algorithm of automatic transcription are:

1.Number of stressed vowel in the word.
2.Alphabet A of Russian letters and corresponding digital codes.The letters are coded in such a way that operations of substitution of symbols can be performed as arithmetical operations.
3.Alphabet T of transcriptional letters and corresponding digital codes.
4.Multitude S,containing words that form deviations from the rules of transcription,numerals in the form of numbers,special signs and symbols.Each word of this multitude is put into correspondence with its transcribed word.
5.Function F(a)=t,that transforms words written in letters of the alphabet A into transcribed words in letters of the alpha bet T. Further we bring the algorithm of rapid transcription:

1.From the original text a word is separated.If there remain no more words,go over to position 6.
2.If the separated word belongs to multitude S,it is replaced by the correspondigg transcribed word and we go over to 1.
3.Each letter of the separated word is substituted by the digital code that corresponds with the letters of alphabet A. The stressed vowel of this word is replaced by the corresponding digital code and after it we add the digital code 100. Now we designate the digital code of the replaced letter or the digital code of the replaced letter and the code determining the character of the stressed vowel with a(i),where i represents the sequence number of the letter in the given word.
4.In the coded word a(1)... a(i)...a(n) we replace a(i) or the aggregation a(i) in accordance with the value of function F,that is specified in the table for the code,or the aggregation of codes of the transcription alphabet,designating them by t(J),where j represents the sequence number of the letter in the transcribed word.
5.In the succession t(1)...t(j)..t(m) we replace t(j) by the corresponding letter of the alphabet T,regarding the character of the sressed vowel.
6.End.

The described algorithm allows to perform the transcription of any Russian text at random according to the rules of Russian phonetics.

The most important word in a syntagm or phrase,the stressed one, tends to occupy a place in the end,that is why the definition of the syntagmatic stress is mainly accomplished with an algorithm of derivation of syntagms in the original text. The algorithms developed for the derivation of syntagms and the definition of prosodic features are founded on the morphological,syntactical and semantic analysis of the text.The system of compilation synthesis of speech includes moduli of derivation of segments,estimationof main tone frequency and also a modulus for adaptive connection of segments and means of developing,storing and reflection of obtained information about the speech signal.

ORCHESTRATING ACOUSTIC CUES TO LINGUISTIC EFFECT

LEIGH LISKER

University of Pennsylvania
and
Haskins Laboratories

ABSTRACT

A most convincing way to demonstrate that an acoustic property is a cue for the listener would be to find speech events that constitute minimal pairs with respect to that property, but in nature such pairs are most unlikely. The English words rapid and rabid are a minimal pair at the level of the segmental phoneme, and are near minimal at the level of the phonetic feature, but as many as sixteen acoustic properties are candidate cues to the lexical distinction. Three properties lend themselves to simple waveform editing: the duration of the stressed vowel, the duration of the closure, and the glottal buzz vs silence of the closure signal. Listener responses to stimuli having natural values of these properties show that, with a single exception, there was no decisive effect on word identification produced by a shift in the value of any one property. At least two properties had to be changed to achieve any significant effect.

Phonetic research nowadays considers the processes involved in speech communication from a wide variety of perspectives, but a central concern remains that of identifying and characterizing those features of the speech processes that serve a message-differentiating function. The phonetic analysis of a speech signal into a temporal sequence of "sounds," as well as the decomposition of those sounds into features, provide a framework within which to specify the distinctive properties that determine a particular interpretation of the signal. A coherent account of a given speech event, considered as representative of a set of linguistically equivalent events, states the interrelations among physiological, anatomical and acoustic patterns, and the nature of their connection to the listener responses they elicit. By far the most attention has been given to finding the acoustic cues to the linguistic message conveyed by a vocal tract emission. The search has involved the acoustic analysis of signals, the selection of promising cue candidates, and the empirical assessment of their cue value by the methods of speech synthesis. Such evaluation of a feature's cue value typically has involved the use of sets of acoustic patterns designed to maximize the likelihood that the feature of interest will affect listeners' response behavior. The number of acoustic pattern features that have been found to have measurable cue value is uncertain, and presumably with continued research along established lines that number will only increase. Clearly it is easier to show that a feature has cue value than to justify a claim to the contrary (the famous unprovability of the null hypothesis).

Most of the acoustic cues so far uncovered are referred to as segmental cues, or even as cues to particular phonetic features of segments. The experimental data supporting their identification are derived via some variant of the linguist's "minimal pair" test. A most convincing way to show that an acoustic property is a cue for the listener would be to find speech events that constitute minimal pairs with respect to that property, but in nature such pairs are very unlikely. The English words rapid rabid make a minimal pair at the level of the segmental phoneme, and an almost minimal one at the level of the phonetic feature, but as many as sixteen acoustic properties are candidate cues to the lexical distinction. It is not certain, however, that any one of them is an independent cue, i.e. one that is capable of signaling a lexical distinction by itself. Even if a given acoustic property can be shone to have such power to affect perception, it need not be true that this property functions independently in nature.

Here I want to report some listener responses to sets of stimuli derived by waveform editing of some naturally produced tokens of rapid and rabid. Three properties served as experimental variables: the duration of the closure interval, the glottal buzz/silence difference during closure, and the duration of the pre-closure vowel. Unlike many tests of this kind, in which the values assigned a variable range over a span in steps of a size designed to establish category boundaries, in the tests reported on here each variable was given just two values, each chosen on the basis of naturalness.

A token of each of the sentences I think it's rapid and I think it's rabid was recorded by a speaker of a central eastcoast variety of American English and stored on computer. A waveform editing program was applied to produce a total of sixteen different acoustic patterns. The durations of the intervals corresponding to the

labial closures were set to values of 60 and 120 msecs, these being typical of /b/ and /p/ closures for the speaker. The closure intervals were either acoustically blank or filled with buzz derived from the original /b/ closure. The pre-closure intervals, from the cessation of the noise interval marking the /s/ preceding the target word to the beginning of closure, were set at the following values:-for derivatives of rabid: 270 msec, the original value, and 230 msecs; for rapid derivatives: 190 msecs, the original value, and 230 msec. The common value of 230 msecs was selected because it fell within the range of natural values for both words in the sentence context used. (Shortening the pre-closure span to a duration of 190 msecs effected a noticeable shift in vowel quality.) A test order in which each of the sixteen stimuli was presented five times, i.e. a random order of eighty items, was presented to twelve native American English speakers, all linguistically and phonetically naive. Each test item was composed of an acoustically invariant carrier I think it's followed by the target word to be identified. Listeners' responses were the following:

Source: rabid		Operation	% "rabid"
	1)	none	100
	2)	-voicing	95
	3)	+long closure	100
	4)	-long vowel	100
	5)	+long closure -long vowel	97
	6)	-voicing -long vowel	93
	7)	-voicing +long closure	15
	8)	-voicing +long closure -long vowel	8

Source: rapid		Operation	% "rapid"
	1)	none	100
	2)	+long vowel	100
	3)	-long closure	98
	4)	+voicing	12
	5)	+long vowel -long closure	62
	6)	+long vowel +voicing	10
	7)	-long closure +voicing	13
	8)	+long vowel -long closure +voicing	8

For each of the variables a change to a value not normally associated with the original stimulus type has, with one exception, no great effect on labeling behavior. Only when glottal buzz replaces the silence of the /p/ closure is there a decided shift to "rabid" judgments.

It does not follow, of course, that the three features are of negligible importance for the perception of the two words. Thus a combination of devoicing and lengthening of the /b/ closure elicited an overwhelmingly "rapid" response, a result in conformity with earlier findings. A shortening of the /p/ closure together with a lengthening of the preceding vocalic interval yielded mostly "rabid" responses. Original "rapid" was heard largely as "rabid," while "rabid" went to "rapid" when all three variable features were assigned values appropriate to the competing form.

The results summarized above indicate that an acoustic feature to which cue value has been attributed does not always effect a significant effect on linguistic labeling behavior; its effect is quite context-dependent. Indeed it may well be, in the case of certain properties, that the context in which it can be decisive can only (?) be contrived in the laboratory. The status of an acoustic feature of speech is therefore very different from that of a phonetic feature, which we generally suppose to possess the power, for at least some natural phonetic system, to mark differentially some words from others, and to do this independently of other phonetic features.

WORD-INITIAL CONSONANT LENGTH IN PATTANI MALAY

ARTHUR S. ABRAMSON

The University of Connecticut and Haskins Laboratories
Storrs, CT, USA New Haven, CT, USA

ABSTRACT

Pattani Malay has distinctive length in all word-initial consonants. Earlier work showed that variations in closure-duration yield perceptual shifts between "short" and "long" phonemes for all sentence-medial intervocalic consonants but only for sentence-initial consonants with acoustic excitation before the release. For words, however, with initial voiceless closures but no pre-release excitation, which are identified well in isolation, where are the cues to the "length" distinction? In the belief that the underlying mechanism is the temporal control of closure, two hypotheses are tested here acoustically: (1) For all consonants, the closure-durations differentiate the short and long categories. (2) The ratio of the amplitude of the first syllable to the second syllable is greater in disyllabic words with long plosives than in those with short plosives.

BACKGROUND

The use of time and timing [1, 2] for phonological distinctions is still an important topic for research. This study tries to shed further light on the acoustic bases of length contrasts in which the relative durations of vocalic and consonantal gestures seem to have a distinctive function. Insofar as it might be a phonetic matter rather than an abstract phonological one, the question of whether to treat long segments as "geminates" will not be handled here.

Treatments of phonemic consonant length usually discuss intervocalic consonants, as in Estonian and Italian, where it is easy to show the physical reliability and perceptual relevance of durational differences in closures and constrictions. A language with this distinction in word-initial, and thus potentially, utterance-initial position, is rare.

The Language

Pattani Malay, spoken by some 600,000 ethnic Malays in southeastern Thailand, has a length-distinction for all consonants in word-initial position [3]. (The language was first called to my attention by Christopher Court and Jimmy G. Harris.) Here are some word-pairs with the contrast:

/makɛ/ 'to eat' /mːakɛ/ 'to be eaten'
/lamaʔ/ 'late' /lːamaʔ/ 'to make late'
/siku/ 'elbow' /sːiku/ 'hand-tool'
/dʑalɛ/ 'way' /dʑːalɛ/ 'to walk'
/butɔ/ 'blind' /bːutɔ/ 'a kind of tree'

All of the foregoing examples have acoustic excitation during their closures or constrictions, but there is none in the voiceless unaspirated plosives, as in these examples:

/tɕuɣi/ 'to rob' /tɕːuɣi/ 'robber'
/tawa/ 'bland' /tːawa/ 'to show wares'

Recent work [4] has shown the power of closure-duration as an acoustic cue to the short-long distinction. Incremental shortening of acoustically excited closures yields perceptual shifts from long to short consonants. Voiceless plosives with their silent closures can be tested only in utterance-medial intervocalic slots; there, shortening or lengthening a silent gap induces shifts.

Goals

The justification for the perceptual experiments [4] was impressionistic observations of length and a small body of instrumental measurements. The first goal here was to determine the statistical reliability of closure-duration as a differentiator of the categories. The second goal was to explore the possible role of overall amplitude in the distinction. That is, for utterance-initial voiceless plosives, something other than audible differences in closure durations must convey the distinction. Although other acoustic features, such as fundamental-frequency shifts and formant-transition rates, are not ruled out, the hypothesis considered here was that the aerodynamic consequences of the apparent articulatory mechanism would cause a higher amplitude upon the release of a long plosive.

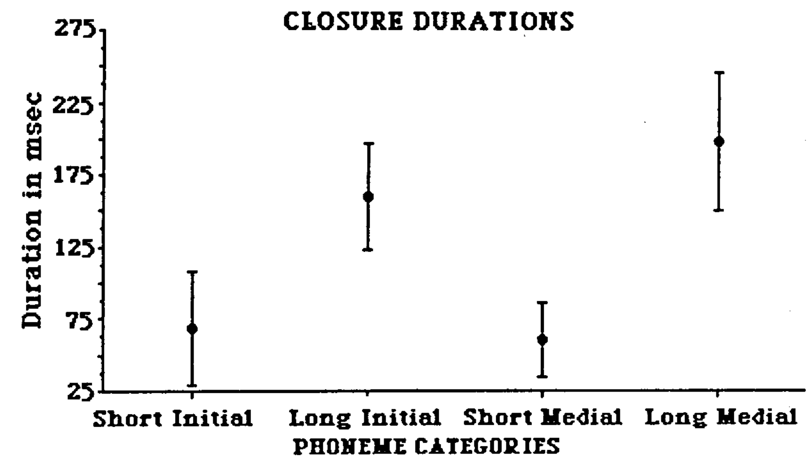


Fig. 1. Means and one-standard-deviation error bars for Speaker PCM. Initial: C, n = 28; Cː, n = 28. Medial: C, n = 44; Cː, n = 44.

Data

Recordings were made of several native speakers, but only those of one man, PMC, were analyzed for this report. Minimal pairs of disyllabic words, two tokens of each, were elicited in isolation and in a carrier sentence. These utterances were digitized for measurement in a waveform editing program and for spectral analysis.

DURATION

The durations of all closures and constrictions were measured for all utterance-initial consonants--except, of course, for the voiceless ones--and all utterance-medial consonants. This was done by examining the waveforms for acoustic signs of forming and releasing obstructions in the supraglottal vocal tract; these were mainly release bursts and sudden changes in amplitude. Occasional difficult cases were checked against spectrograms. The data are summarized in Figure 1.

An analysis of variance showed duration to be highly significant for initial consonants [F (1, 26) = 49.40, $p < 0.0001$] and medial consonants [F (1, 42) =185.19, $p < 0.0001$]. To measure durations of initial voiceless closures would require either a direct look at articulation or, perhaps, measurements of buccal air pressure. The robustness of the difference for medial voiceless plosives, in conformity with the graphs for the medials in Figure 1, and the data in both positions for all other consonants, suggest the high probability of a closure-duration difference for initial voiceless plosives too.

AMPLITUDE

Since the major concern was with initial voiceless plosives, measurements of amplitude were limited to isolated words. Pilot work with rise time, peak value, and average amplitude of the first syllable relative to the second gave useful results only with the third method.

A program with variable window-settings, designed by Richard S. McGowan, was used to derive the average root-mean-square (RMS) amplitude of each syllable in the disyllabic words recorded. (Apparently, monosyllabic words are rare.) The results are given in Table 1.

As expected, the most promising set of data in Table 1 is for the voiceless plosives (stops and affricates). In the analysis of variance of the underlying data, the interaction between consonant length and syllable approached significance: F (1, 14) = 4.36, p = 0.056. Indeed, post-hoc simple-effects tests showed that the difference between the short and long consonants with respect to amplitude-ratio is strongly significant: F (1, 14) = 11.037, p = 0.005. Although the continuants (nasals, laterals, and fricatives) showed a slight tendency in the same direction, the effect was not statistically significant. Compared with the continuants, the voiced plosives present a stronger case in the simple-effects test: F (1, 24) = 4.24, p = 0.05. With its greater number of degrees of freedom, however, this category underwent a more powerful test than the voiceless plosives and yielded a weaker although significant effect.

TABLE 1

Means and Standard Deviations for RMS Amplitudes in dB

Type	Syl. No.	Short Consonants n	M	SD	Long Consonants n	M	SD
Plosives							
Voiceless	1	16	**47.5**	3.0	16	**51.0**	2.2
	2	16	**45.0**	2.8	16	**45.0**	2.3
Voiced	1	14	**46.8**	3.9	14	**49.5**	3.5
	2	14	**43.3**	3.4	14	**44.4**	2.9
Continuants	1	14	**45.1**	3.9	14	**48.1**	2.8
	2	14	**46.1**	3.4	14	**46.9**	3.4

CONCLUSION

That the phonemic distinction between "short" and "long" Pattani Malay consonants is based on the quantitative feature of articulatory timing is abundantly clear from the data of Figure 1. Indeed, the perceptual efficacy of closure-durations has been demonstrated for medial position and for initial consonants with audible excitation [4]. (Of course, the value of this cue has been demonstrated for at least medial position in some other languages [e.g.,5].)

Even if, as seems likely, the underlying mechanism for this length distinction is articulatory timing, there may nevertheless be more than one acoustic cue involved. That is, temporal control of closures and constriction, intersecting with states of the glottis, may engender not only varying spans of silence or appropriate sound but also, perhaps, variations in air flow and pressure with certain acoustic consequences. The data in Table 1 show that for long voiceless initial plosives the average RMS amplitude is significantly higher in the first syllable than the second. There is also a significant but somewhat smaller effect for voiced plosives. We may speculate that although both categories involve complete momentary obstruction of the oral air flow, the presumed greater impedance at the larynx for the voiced plosives lessens the effect. For the continuants, however, which always have a by-pass for the air, there is no effect.

The amplitudes of PMC's embedded words remain to be measured. In the meantime, a cursory look at the productions of three other native speakers of the language seems to support the findings. Their utterances, too, will have to be measured. Finally, to round out the first experiments on perception [4], the plan is to produce stimuli with controlled variations in amplitude on disyllables.

ACKNOWLEDGMENTS

The work was supported by NICHD Grant HD - 01994 to Haskins Laboratories. I am grateful to Mr. Paitoon Masmintra Chaiyanara of The Prince of Songkhla University, Pattani and Dr. Theraphan L. Thongkum of Chulalongkorn University for their help and to their institutions for their warm hospitality. The advice of Dr. Richard S. McGowan and Professor Leonard Katz was most helpful.

REFERENCES

[1] I. Lehiste, *Suprasegmentals*. Cambridge, MA: MIT Press, 1970.

[2] L. Lisker, On time and timing in speech. In T.A. Sebeok *et al.* (eds.), *Current Trends in Linguistics*, **12** (pp. 2387-2418). The Hague: Mouton, 1974.

[3] P. Masmintra Chaiyanara, *Dialek Melayu Patani dan Bahasa Malaysia: Satu Kajian Perbandingan dari segi Fonologi, Morfologi dan Syntaksis.* Master's thesis, University of Malaya, 1983.

[4] A.S. Abramson, Distinctive length in initial consonants: Pattani Malay. *J. acoust. Soc. Am.*, **79**, S27 (abstract), 1986.

[5] A. Lahiri and J. Hankamer, Acoustic properties of geminate consonants. *J. acoust. Soc. Am.*, **80**, S62 (abstract), 1986.

A CROSS-LANGUAGE STUDY ON THE PERCEPTION OF SYNTHETIC SPEECH SOUNDS OF [r - l]

KATSUMASA SHIMIZU

Dept. of Languages
Nagoya Gakuin University
Seto city, Aichiken, Japan

MASATAKE DANTSUJI

Dept. of Linguistics
Kyoto University
Sakyoku, Kyoto 606 Japan

ABSTRACT

The present study is concerned with the perception of synthetic speech sounds of [r - l] continuum by speakers of different languages. Specifically, the study examines how the differences of the linguistic function of liquids in English, Spanish, Japanese, Hindi, Korean, and Chinese affect the perception of the synthetic continuum and reports the results of identification and discrimination tests. The results indicate that different modes of perception appeared depending on the phonemic functions of liquids in each language. The boundary between /r/ and /l/ differed systematically in each language and the speakers having a phonemic function of /r/ and /l/ showed a categorical mode of perception and differences of linguistic experience cause those of perceptual modes.

INTRODUCTION

It is generally known that speakers of different languages show some different characteristics in the perception of speech sounds. Among the cross-language studies on speech perception, the study on [r] and [l] has been of considerable interest among phoneticians since the /r/ - /l/ contrast has often been a choice to study the effect of linguistic experience and they have unique articulatory and acoustic features which can be defined as an intermediate between stop consonants and vowels. There have been several reports on the experiments of the perception of [r] and [l] using synthetic speech sounds. Miyawaki et al.[1] studied the effect of linguistic experience of English and Japanese in the perception of synthetic [r - l] continuum and mentioned that the difference of linguistic experience is specific to perception of speech mode. Furthermore, Mochizuki[2] and Shimizu and Dantsuji[3] carried out the experiments of speech perception to English and Japanese speakers by using natural and synthetic speech sounds and reported that English speakers perceive the [r - l] continuum categorically, while Japanese speakers do it continuously, and the difference of the perception mode can be attributed to the one of the linguistic function of the liquids in these languages. It is well known that the /r/ - /l/ contrast is functional in English but not in Japanese and different function of the liquids in these languages cause some learning problem for Japanese speakers.
Although the difference between English and Japanese speakers in the perception of [r - l] continuum has been accepted, the experimental data on other language speakers are very scarce, and it will be necessary to examine other language speakers in order to clarify the relationship between linguistic experience and the mode of speech perception. Viewing from these points, the present study aims at examining how the difference in the linguistic function of liquids in other languages affects the perception of [r - l] continuum and how linguistic experience affects the mode of perception.

EXPERIMENTAL PROCEDURE

Subjects

The subjects composed of speakers from six language groups: English, Spanish, Japanese, Hindi, Korean, and Chinese.

English: 7 native speakers of American English took part in the experiment. They had lived in Japan for a certain period, ranging from three months to three years.
Spanish: 4 native speakers of Spanish took part in the experiment. They were undergraduate students at UCLA.
Japanese: 23 native speakers of Japanese were tested in a classroom. They were undergraduate students in an introductory phonetics class at Sugiyama Joshi Univ.
Hindi: 2 native speakers of Hindi took part in the experiment. They were graduate students in physics and journalism at UCLA.

Korean: 3 native speakers of Korean took part in the experiment. They were graduate students at Kyoto University.
Chinese: 3 Chinese(1 from Hong Kong, 2 from Mainland China) took part in the experiment.

Stimulus Materials

The stimulus were prepared on the OVE III synthesizer at Haskins Laboratories. The 10 step [ra - la] stimuli differed in the frequency values of F2 and F3 within the initial state portions and the transition portions. F2 values varied in almost equal step from 951 to 1404 Hz and F3 values from 1488 to 3246 Hz. F1 values were kept constant for 10 stimuli. The stimulus with 1404 Hz of F2 and 3246 Hz of F3 was a good /la/, while the one with 951 Hz of F2 and 1488 Hz of F3 was a good /ra/. The total duration was 377 msec. Two types of test were prepared: an identification test and an oddity test. In the former test, each stimulus was repeated 10 times, making the total presentation 100, and the stimuli were randomly arranged. The interstimulus interval was 1 sec. and the block interval was 10 sec. The oddity discrimination test consisted of 18 repetitions of each of 7 stimulus pairs(1-4, 2-5, 3-6, 4-7, 5-8, 6-9, and 7-10), totalling 126 traids in all. Stimulus pairs were arranged such that members were three steps apart along the 10 step stimuli. For each pair, traids were constructed by dupulicating one member of the pair, and six permutations of each companion were included; i.e., for 1-4 pair, 1-4-4, 1-1-4, 4-1-1, 1-4-1, 4-1-4, and 4-4-1. All test materials were recorded on audio tape for presentation to subjects.

Results

Results of the identification and discrimination tests can be shown in Figures 1 - 3.

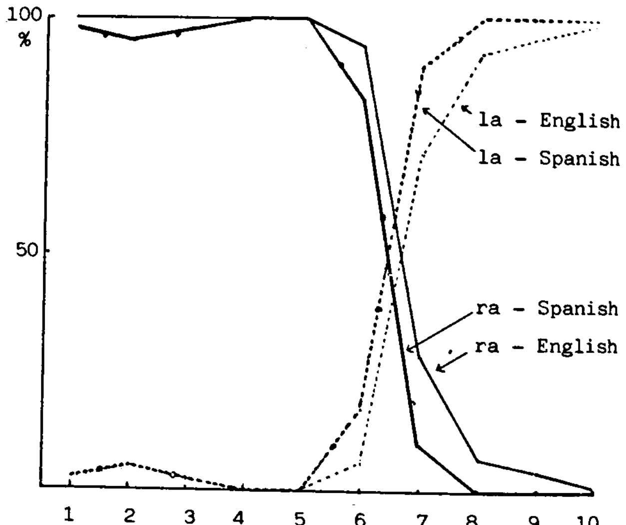


Fig. 1a Pooled identification of English and Spanish speakers

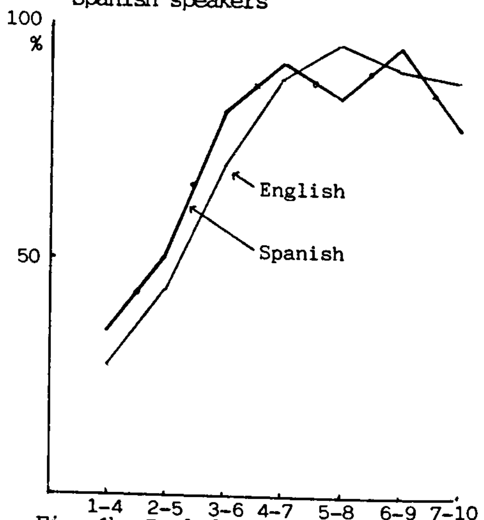


Fig. 1b Pooled discrimination of English Spanish speakers

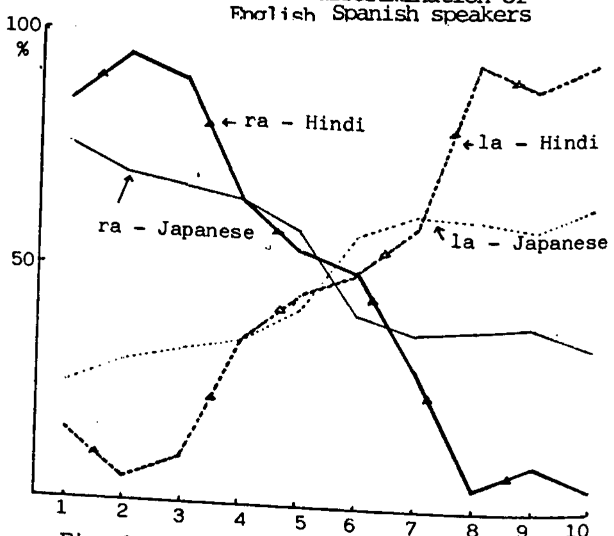


Fig. 2a Pooled identification of Hindi and Japanese speakers

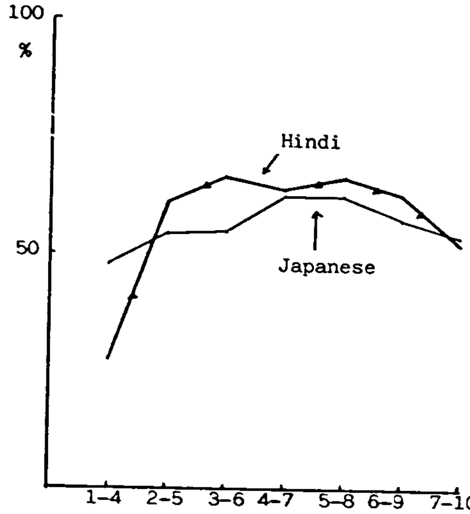


Fig. 2b Pooled discrimination of Hindi and Japanese speakers

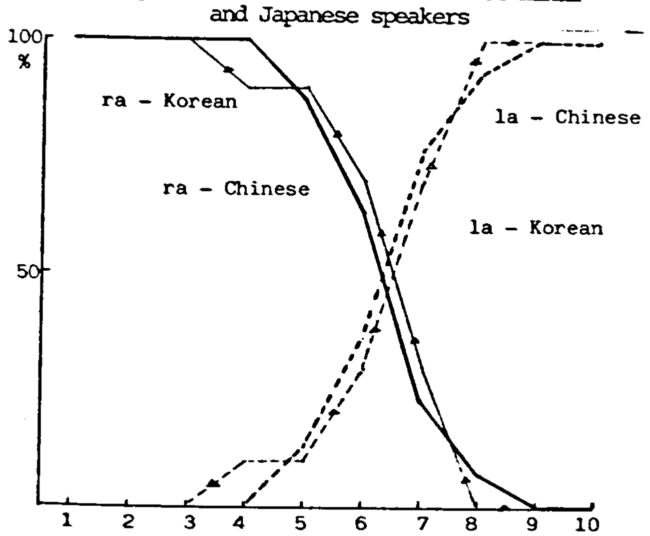


Fig. 3a Pooled identification of Korean and Chinese speakers

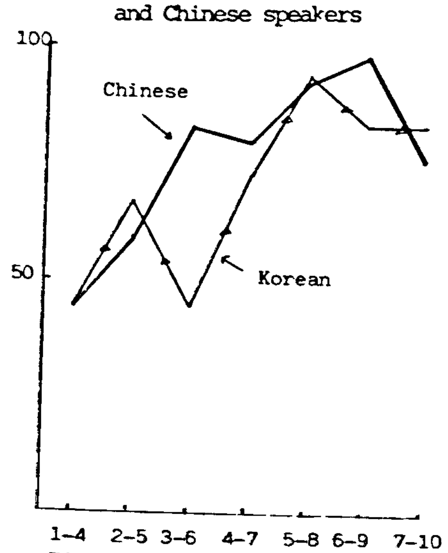


Fig. 3b Pooled discrimination of Korean and Chinese

Figures 1a shows the results of the pooled identification for English and Spanish speakers. Both subject groups show a similar pattern of identification curve. Subjects in both groups identified stimuli 1 to 5 as /ra/ and stimuli 8 to 10 as /la/, and showed an abrupt shift of the curve in the stimulus range from 6 to 8. The boundaries lie between stimuli 6 and 7, though there are slight discrepancies in both groups of subjects. Figure 1b shows the results of a pooled discrimination for both groups, and both subject groups show similar patterns in accuracy. The accuracy was below at 1-4 pair, but sharply rose to about 90 % accuracy at the pairs which are considered to be in the phonetic boundary between /r/ and /l/. That is, both groups of subjects discriminated very well between stimuli drawn from different phonetic categories but very poorly between stimuli from the same phonetic category. Both identification and discrimination curves show that both subject groups identify the [r - l]continuum categorically.

Figure 2a shows the identification curves for Japanese Hindi speakers. Unlike the results in figure 1a, both Japanese and Hindi speakers showed a gradual shift of identification curve as the stimulus shifts from 3 to 8. Japanese subjects identified stimuli 1,2 and 3 as /ra/ and stimuli 8, 9 and 10 as /la/ with 60 - 70% of identification rate, while Hindi speakers identified these stimuli with much higher rate than Japanese speakers. Figure 2b shows the discrimination curves for both language groups. Both subject groups discriminated the stimuli with an accuracy rate ranging from 50 to 70 % across the continuum of stimulus set from the 2-5 pair. Although there was a moderate increase in discrimination of stimulus pairs 3-6, 4-7 and 5-8, there is no noticeable change in the accuracy percentage in both groups within and between categories. It can be said that these results indicate that Japanese and Hindi speakers perceive the [r - l] continuum continuously.

Figure 3a shows the identification curves for Chinese and Korean speakers. The curves are similar to the ones of English and Spanish speakers, but with less abrupt shift. The boundary between /r/ and /l/ lies between stimuli 6 and 7 for both groups of subjects. Figure 3b shows a pooled discrimination for the two groups. Both groups show sharp rise at the pairs of 5-8 and 6-9. Examining these results, it can be said that Chinese and Korean speakers perceive the [r - l] continuum with near categorical manner.

DISCUSSION

We have examined how speakers of six languages perceive the [r - l] continuum and how the differences of linguistic experience affect the mode of speech perception. Examining the results of identification and discrimination tests, it has become clear that the speakers of six different languages show different patterns of performance. It can be said that familiarity with the [r - l] distinction has an impact on the perception of the continuum.

English and Spanish speakers show a peak of accuracy at the point where stimuli from different phonetic classes are being contrasted. Japanese and Hindi speakers are unable to discriminate [r] and [l] over the continuum. The results on Japanese subjects conform with the previous studies(Miyawaki et al.[1]) and are in harmony with what is known about linguistic function of the liquid in Japanese. The finding that Hindi speakers can not discriminate [r] and [l] over the synthetic continuum indicates that stimuli are not similar to the phonetically realized forms of the Hindi contrast of lateral and tap.

As shown in figures 3a,b, Korean and Chinese speakers discriminate the continuum in a near categorical manner. It is known that Korean has no phonemic contrast of /r/ and /l/ and has only one phonemic /l/, but /l/ has allophonic variations of [l] or tap in some phonetic environments. The finding that both groups of subjects can discriminate [r] and [l] indicates that the stimuli are similar to the allophonic variations of liquids in both languages.

These results of experiments indicate that different perceptual modes appeared depending on the phonemic functions of liquids in each language. The boundary between /r/ and /l/ differed systematically in languages, and the speakers having a phonemic contrast of /r/ and /l/ showed a clear categorical mode perception. It can be said, therefore, that the differences of linguistic experience cause those in the perceptual modes.

REFERENCES

[1] Miyawaki, K. et al. "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percep.& Psych., 1975,18(5).
[2] Mochizuki, M. "The identification of /r/ and /l/ in natural and synthesized speech," J. Phonetics, 1981, 9.
[3] Shimizu, K. and Dantsuji,M. "A study on the perception of /r/ and /l/ in natural and synthetic speech sounds," Studia Phonologica,1983, 18.

FORMANT TRANSITIONS AND RELEASE BURSTS AS PERCEPTUAL CUES FOR RUSSIAN VOICELESS PLOSIVES

V.N. LOZHKIN

Department of Applied Linguistics
Moscow State University
Moscow, USSR

ABSTRACT

The present study attempts to investigate the significance of the release burst and the formant transitions in the perception of Russian voiceless plosives by native listeners. The method involved deleting of releases in some consonants, it resulted in worsening recognition of these sounds - 34% for initial plosives, 60% and 70% for intervocalic and final plosives respectively. Thus, it is stated that release segments carry important information bearing on the place of articulation of Russian stops. The results of the study are in agreement with those obtained on the material of English and Hungarian stops and defy the prevailing significance of CV-transitions in voiceless plosives recognition.

INTRODUCTION

Although the history of experimental studies stimulated by the development of the 'Visible Speech' Sound Spectrograph dates back to the 40 ies, it is hardly possible to say that the distinctive features of phonemes are fully investigated. It is not surprising since the acoustic features of sounds in fluent speech can vary dramatically due to the context, the speaker's peculiarities, the mode of articulation, etc. Besides, the speech signal is highly redundant and possesses a great variety of distinctive features.

Quite a number of works have been dedicated to the distinctive features of plosives. It is a stated fact that information about the place of articulation of a stop can be found in the formant transition of adjacent vowels as well as in the stop burst. The relative significance of release and transition in the stop identification, however, is to be further investigated.

This problem is relevant for the systems of automatic speech recognition and high-quality speech synthesis. Another important problem concerns the search for invariant (i.e. independent of a context) features of the place of articulation/I/ /2/.

There exist at least four estimates of the relative significance of the release and transition cues.

On the one hand, under the influence of the studies carried out in Haskins Laboratories in the 50s on the material of synthesized syllables the view of the dominant role of the CV-transitions for the place of articulation identification has been adopted(The bulk of the results that became classical can be found in /3/). On the other hand, there are many indications that a formant transition might be of a smaller importance since the crucial information about the place is in the release. Such evidence has been obtained in some early studies of human speech/4-7/. One can argue that the transition and release cues are functionally equivalent, the former may be dominant in one case and the latter in another/8/. Finally, one can also argue that it is not correct to oppose transition and release since they may become inseparable in the case of a prevocalic position of a plosive/9/. It seems that all the approaches are sufficiently grounded(comparative analysis of various approaches is to be done elsewhere).

Keeping in mind that the results obtained in one language are not necessarily relevant for others and since much of the available information concerns English consonants we attempted to investigate the problem on the material of the Russian initial, intervocalic and final voiceless plosives. The relative significance of the transition and release cues in the perception of the place of articulation is discussed here. The problem is of particular interest since palatalized stops are very characteristic of the Russian language(we could mention only few articles dealing with the subject/IO-I2/)

METHOD

The model of a voiceless plosive is used according to which the four segments can bear information about the place of articulation of an intervocalic plosive: I)the segment of VC- or final transition; 2)the closure;3)the release after an abrupt closure breaks;4)the segment of a CV- or initial transition. The VOT was taken for a release end. In a more detailed model a release is further divided into: a) a starting impulse;b)frication;c)aspiration(cf./I3/). A certain amount of residual noise may add to the voiced beginnig of a following vowel/I0/. Since these peculiarities of the Russian language are not phonemically relevant they are not considered here.

To investigate the relative significance of the transition and release cues for prevocalic, intervocalic and postvocalic plosives some meaningless successions or 'non-words' (=nonsense words) of the $C_IVC_2VC_3$-type were chosen, the consonant beifig the same in the one case (e.g. 'papap', 'totot', etc) and different in the other (e.g. 'patak', 'kopot', etc). The vowel has been taken out of the set (a,o,u,i,e), the second syllable of each non-word was stressed.

30 non-words were tape recorded by two male speakers. The instruction to the speakers was to utter the stimuli distinctly without changing the quality of vowels. The interval between the stimuli was 5 seconds.

The tape rings were made of the copies of the original recordings that underwent segmentation by means of the low-noise electronic separator described elsewhere /I4/.

A release for one of the plosives (initial, central or final) has been deleted in every non-word by means of the separator. The fragments of non-words with a release deleted were used as test stimuli. Thus the relative significance of the CV-transition, VCV-transitions and VC-transitions in place identification was studied.

The deletion procedure has been controlled aurally and by means of the oscilloscope. The test stimuli have been recorded on the second tape-recorder. Each fragment was recorded three times on a tape ring. The triads of the identical stimuli were separated by the pure monotone markers. The presentation rate of the test stimuli which depended on the ring length and the tape speed was about 3.8 seconds.

Non-words with release deleted were mixed with undamaged non-words and thus presented to ten listeners (students, laboratory assistants etc.) without hearing loss. Most of the listeners were experienced in listening to articulatory tests and synthesized speech patterns. The signal was fed into the head-phones in a quiet room. Each listener could adjust the volume in his head-phones.

The instruction given to the listeners was as follows: "You will hear non-words of the CVCVC-type, where C is any of the plosives – /p/,/p'/,/t/,/t'/,/k/,/k'/. Each non-word is repeated 3 times. After listening to a triad you are to write it down in ordinary letters ('papap' for example). If you detect a distorted (damaged) consonant please underline it as shown below: 'patat', or 'kutuk', or 'kakak'*)

Notice: the soft /k'/ may occur in a final position that are not typical for the Russian speech, e.g. petek' "

The instruction was presented by the experimenter orally and then its printed text was distributed among the listeners. The nature of the damage was not revealed to the listeners as well as the consonants of a non-word being the same $C_I=C_2=C_3$ or different $C_I \neq C_2 \neq C_3$. Having learned the instruction the listeners began to listen to the test for a few minutes and then to listen and fix their judgements on special forms. Each listener listened to every test 3 times with a few days intervals.

A test consisted of I20 randomized non-words out of which 30 were not damaged, while 90 contained a stop with a deleted release. Thus a test contained in all 270 undamaged and 90 damaged stops. There were two non-word lists with different stimuli order, the first list was read by speaker L, and the second by speaker S.

RESULTS

The results of these tests are confusion matrices. A sample of such a matrix is given in the table below. The right and wrong judgements for the /p/, /t/,/k/ stimuli (with release deleted) preceeding /a/ vowel in the first hearing session (speaker L) are presented in the table.

Table

perceived

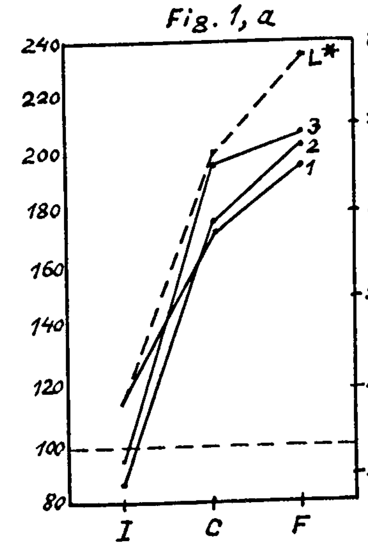
uttered	p	t	k	–	p	t	k	–	p	t	k	–
p	I4		4	2	I7	2		I	20			
t	II	4	3	2	6	I4			3	I5		2
k	I3	4	2	I	I2	2	4	2	5	5	9	I
	38	8	9	5	35	I8	4	3	28	20	9	3

The judgements are summerized for ten listeners in every matrix. The "–"-sign stands for refusal. The left matrix corresponds to the initial position of a stop, the second and the third – to the central and the final positions, respectively.

It can be seen that in two presentations of a damaged initial /p/ (non-words 'papap' and 'patak'), I4 out of 20 judgements were correct and 4 were substitutions of /k/ for the initial /p/. There were also 2 refusals.It follows from the same matrix that the initial /t/ (in 'tatat' and 'takap') was given only 4 correct judgements, II responses were misjudgements for /p/, 3 – for /k/ and there were also two refusals. The numerals in the bottom row of the first matrix show the number of judgements of the /p/, /t/,/k/ or "–" -type in all 60 presentations of the initial voiceless plosives preceding /a/ (speaker L). There were 38 /p/-judgements, 8 – /t/, 9 – /k/, 5 – "–"; there were also 20 correct judgements (the sum along the main diagonal).The second and the third matrices are constructed in a similar way. There are 30 tables of the kind (2 speakers x 5 vowels x 3 presentations).For the sake of brevity all tables are not presented here.

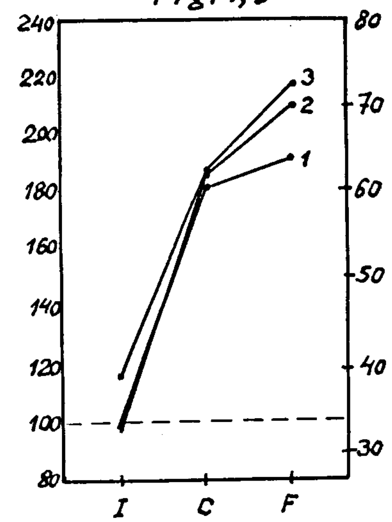
For getting more reliable results the data for every speaker (3 presentations) were combined. One can further combine the data for both speakers, data concerning different vowels, etc. to answer the question – in what way a release deletion may influence the plosive recognition in general, when it is independent of the vowel context and the place of articulation. Figure I(a,b) shows the major result of the experiment. The graphs show that the place of articulation of damaged stops for central plosives is recognized better than for initial ones and for final plosives better than for central ones. At first, we assumed that greater intelligibility of central and final plosives on the second syllable being stressed. The analysis of non-words with the first stressed syllable revealed a similar tendency (graph L^*, fig. I,a). (For the sake of brevity more detailed data of the relative significance of transition and release in stop recognition is omitted here).

Fig. 1, a



*) It is necessary to mention that listeners detected damaged stops very poorly, marking the right consonants and missing the damaged ones.

Fig. 1, b



Graphs a and b show the percentage of correct judgements as a function of the position of the damaged plosive. a – data for the speaker L; b – data for the speaker S. 'I' stands for the initial position of the plosive, 'C' – for its central, and 'F' – for its final position. The absolute numbers of correct judgements is plotted along the ordinate on the left, while the percentage of these judgements is to be found on the right. Curves I, 2, 3 correspond to the first, second and third listening to the non-words with the stress on the second syllable. Curve L^* shows the analogous results for the non-words with the first syllable stressed. Horisontal dash-line corresponds to guessing level.

Thus, the results of the tests show that in human speech a stop release carries very important information and its deletion may result in worsening recognizability of the place of articulation of voiceless plosives, especially of the initial ones. In the last case the recognition does not exceed the guessing level (~34%),*) while the information conveyed by VC-transition may provide higher recognizability for intervocalic plosives (60%) and for final plosives (~70%).

DISCUSSION

The results of the present study agree with the results of the other authors. Thus in /I5/ the relative significance of transition and release in the identificat-

*) It is not quite correct to mention the guessing level since the listeners judgements in this case were not mere guessig: most of them were judgements of /p/-type. The explanation might be that the release for /p/ is very short and faint so listeners may misjudge any voiceless plosive as /p/ when the release is absent.

ion of the place of English voiceless stops preceding /i/,/a/,/u/ vowels has been studied. It has been found that the voiced segment of the initial transition is neither a sufficient nor necessary cue for the place identification in the initial position, and it is the release that accounts for a correct consonant identification.

Similar conclusions were made for the Hungarian voiceless stops in VCV-syllables/I6/. The release was judged to be the most informative among the segments of closure, release and transitions due to /7/, where the distribution of stop cues has been studied.

A few studies of Russian voiceless stops have been described in /I0-I2/. It was found that in the final position of a stop the release is more important cue than the transition(the method consisted in transplanting bursts from one context into another)/I2/. The significance of bursts and final transitions in final positions of stops has been studied in detail in /I0/ on the CVC-syllables. It was shown that in most cases an isolated burst of final stops was sufficient for the identification of place. Final transitions can also carry information about the place of articulation.

In the article /II/, however, VC-transitions were not found significant in the perception of plosives.(A more detaled comparative analysis is required to explain the disparity between /II/ and the present study).

The question remains - which of the two transitions, CV or VC, is more informative in the place identification of the plosives? The results of the present study suggest that VC-transitions is more informative than CV-transitions.Similar conclusions can be found in /I7/, where the role of CV- and VC-transitions in the place of articulation of English stops and fricatives in syllables with neutral /ə/ in natural speech was determined.VC-transitions proved to be more informative than CV-transitions especially for voiceless stops: VC-transitions accounted for 92% of their intelligibility whereas CV-transitions accounted for only 32%. The corresponding values for voiced plosives were 92% and 7I%. According to the authors the reason for VC-transitions being more informative is their better physical manifestation. The conclusion has been substantiated by inverse listening data in particular.

SUMMARY

The results of the present study provide further evidence of the relative significance of transition and release as perceptual cues for Russian voiceless plosives. The regularities which have been found may prove useful in developing more efficient automatic recognition systems and high-quality speech synthesis.

REFERENCES

I.K.N.Stevens,S.E.Blumstein. Invariant cues for place of articulation in stop consonants. J.Acoust.Soc.Am. I978,v.64, N5,I358-I368.

2.D.Kewley-Port,D.B.Pisoni,M.Studdert-Kennedy. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. J.Acoust.Soc.Am., I983,v.73,N5,I779-I793.

3.J.L.Flanagan.Speech analysis,synthesis and perception.Springer Verlag,I965

4.M.Halle,G.W.Hughes,J.-P.A.Radley. Acoustic properties of stop consonants. J.Acoust.Soc.Am.,I957,v.29,NI,I07-II6.

5.A.Malècot.The role of releases in the identification of released final stops. Language,I958,v.34,N3(part I), 370-380.

6.W.S.-Y. Wang.Transition and release as perceptual cues for final plosives.J. Speech Hear.Res.,I959,v.2,NI,66-73.

7.P.W.Guelke,E.D.Smith. Distribution of information in stop consonants. Proc. IEE,I963,v.II0,N2,680-688.

8.M.F.Dorman,M.Studdert-Kennedy,L.J. Raphael. Stop consonant recognition:release bursts and formant transitions as a functionally equivalent, context-dependent cues. Perception and Psychophysics, I977,v.22,N2,I09-I23.

9.S.E.Blumstein,K.N.Stevens. Perceptual invariance and onset spectra for stop consonants in different vowel environment. J.Acoust.Soc.Am.,I980,v.67,N2,648-662.

I0. Дукельский Н.И. Принципы сегментации речевого потока. М.-Л. изд-во АН СССР. I962.

II. Люблинская В.В. Распознавание артикуляторных признаков смычных согласных по переходу от гласного к согласному. Акуст.журнал,I966,т.I2,вып.2,2I3-22I.

I2.Охарева Н. Г. Взрыв как основной показатель места образования конечных глухих взрывных согласных."Исследования моделей речеобразования и речевосприятия" Л.I98I,с.I40-I48.

I3.G.Fant.Stops in CV-syllables.Speech Transmission Laboratory.Quarterly Progress and Status Report.STL-QPSR 4/I969, I-25.

I4.Речь.Артикуляция и восприятие.М.-Л. Наука,I965,с.62.

I5.C.LaRivière,H.Winitz,E.Herriman. Vocalic transitions in the perception of the voiceless initial stops.J.Acoust.Soc. Am.,I975,v.57,N2,470-475.

I6.K.Vicsi.The most relevant acoustical micro segment and its duration necessary for the recognition of unvoiced stops.Acustica,I98I,v.48,NI,53-58.

I7.D.J.Sharf,T.Hemeyer.Identification of place of consonant articulation from vowel formant transitions.J.Acoust.Soc. Am.,I972,v.5I,N2(part 2),652-658.

PERCEPTUAL RELEVANCE OF ACOUSTICAL WORD BOUNDARY MARKERS

Hugo Quené

Institute of Phonetics, Utrecht University
the Netherlands

ABSTRACT

Measurements of segment durations in contrastive word boundary positions support the claim that acoustical boundary marking is realized differently among speakers. A tentative explanation of subjects' high accuracies in boundary detection experiments, viz. their ability to rapidly evaluate speaker-dependent boundary markers ("tuning in") was investigated. Ambiguous word pairs realized by 2 males and 2 females were presented either in lists of items realized by one speaker, or in a list of items of all 4 speakers randomized. From the results, it is concluded that subjects' simultaneous attention for multiple cues, rather than their "tuning in" to single cues, is responsible for the high boundary detection accuracy.

1. INTRODUCTION

Speech segmentation, the division of the continuous fluent speech signal into discrete words, is one of the most outstanding characteristics of human speech perception. Despite the numerous lexical ambiguities in the acoustic signal, listeners usually perform this task (as a preliminary for, or in interaction with word recognition) without much difficulty, although occasional errors do occur (2;7). Apparently, listeners are helped effectively by such top-down information as syntactic, semantic and contextual constraints, phonotactic restrictions of word structure, and sandhi phenomena. Besides, there are also bottom-up or acoustical phenomena related to word boundaries. Thus acoustically marked word boundaries in the fluent speech signal may help listeners in their segmentation. Among these phenomena, the following have been identified for English or Swedish by various researchers over the past decades as being functional in this respect: lengthening of pre-junctural consonant, aspiration of word-initial voiceless plosives, glottal stop or laryngealization of post-junctural vowels, and allophonic differences for /l,r/ in pre- or post-junctural positions (13;8;14).

In previous word boundary detection experiments (15), subjects were able to reach an overall accuracy of about 80% under conditions where no top-down information could have played a role (listening to ambiguous two-word sequences, not providing contextual or phonotactic cues). Contrary to e.g. (12;4), these results, as well as those by e.g. (10) show that listeners are able to make effective use of these acoustical word boundary markers as cues for speech segmentation, even without additional constraints based on top-down cues. Results of these previous experiments also suggest, that the following boundary marking phenomena had played a major perceptual role: (1) variation of word-initial vs. -final consonant allophone, (2) duration of ambiguous boundary consonant, (3) rise time of post-boundary vowel. Besides, (4) VOT of ambiguous plosives was observed to differ as a function of the intended boundary position.

Before establishing the perceptual relevance of these boundary markers more thoroughly, however, a rather unexpected finding from these experiments had to be further investigated, viz. the significant differences between speakers with regard to the produced acoustical (durational) boundary markers. Since such speaker effects (as well as language-specificity of these cues, as observed by (1)) have strong implications for the perceptual validity of the boundary markers mentioned above, we decided to investigate this matter first; the experiment reported here investigates subjects' ability to perceive speaker-dependent word boundary markers.

2. PRODUCTION

2.1. Material

Twenty-two word sequences were selected (2 word sequences with each of the 10 consonants /p,t,k,d,f,s,x,m,n,l,r/ which may occur word-initially as well as word-finally in Dutch, with this intervocalic ambiguous 'boundary consonant' in both word-final and word-initial position. (From (15) it was observed that word-final devoicing of /d/ did not affect boundary detection accuracy). The resulting 44 word sequences (11(consonants)x2 (sequences)x2(versions)) were embedded in sentences which disambiguated the word sequence.

2.2 Procedure

The 44 sentences were read aloud by 2 males and 2 females at a subjectively fast speech rate (to avoid pausing within sentences). Subsequently, the 4(speakers)x(44 word sequences)=176 ambiguous word sequences were spliced out of the original sentences by means of a computer programme (with visual and auditory feedback; sampling frequency 10 kHz; 12 bits resolution) and stored digitally.

Durations of the relevant speech portions (boundary consonant, VOT, and rise time of post-boundary V_2) were measured and analyzed.

2.3 Results

The data obtained show, that the four speakers produce the same durational difference between /C#/ and /#C/ boundary positions, for each of the three acoustical parameters under observation. However, the significance level of these differences is clearly speaker-dependent, as can be seen from Table I below.

Only one speaker, M2, produces significant differences for all three parameters in contrastive word boundary positions; the others produce some highly significant and some insignificant differences between /VC#V/ and /V#CV/ boundary positions.

Table I:
Resulting t-values (matched observations, pairwise deletion) of the durational differences between /C#/ and /#C/ boundary positions, for (1) duration of the ambiguous consonant, (2) VOT of ambiguous plosives, and (3) rise time of post-boundary vowel, for 4 speakers separately.

variable	M1	M2	F1	F2
(1)	- .132	-2.103*	-2.694*	-5.616***
(2)	2.436	2.865*	.108	1.899
(3)	4.806***	2.229*	5.483***	2.103*

*=p<.05; **=p<.01; ***=p<.001

In short, the durational observations suggest that acoustic marking of word boundaries is speaker-dependent. However, in (15) using a subset of these word sequences as stimuli, subjects obtained detection accuracies of over 75% with all four speakers. That is, although the acoustic marking of word boundaries is different among speakers, subjects were able to use these speaker-dependent markers to a considerable extent (as they were the only systematic cue). This high accuracy in boundary detection can therefore best be explained as a consequence of subjects' ability to evaluate these speaker-dependent acoustical cues rapidly, i.e. to "tune in" to them (analogously to feature adaptation in phoneme perception (5)).

In the following word boundary detection experiment, speaker-dependency in word boundary marking was further investigated. Presumably, it would be easier for listeners to "tune in" to speakers if they hear more stimuli realized by one speaker in a row, as compared to a listening situation in which they hear stimuli by several speakers in random order, and thus have to "tune in" to a different speaker for each new stimulus. This view constitutes the major hypothesis of the following detection experiment.

3. PERCEPTION EXPERIMENT

3.1. Design

In order to establish the effects of subjects' rapid evaluation of speaker-dependent word boundary marking, each subject had to perform boundary detection under two presentation conditions: (1) presentation of sub-lists of stimuli realized by the same speaker ('Sequences'), and (2) presentation of a randomized list of stimuli realized by all four speakers ('Randomized'). Thus, each subject had to perform boundary detection in all 176 word sequences (4(speakers)x22(word sequences)x2(boundary positions)).

The relative order of these two presentation conditions was co-varied between subjects with the intended boundary position in an ambiguous word sequence (/VC#V/ vs. /V#CV/), yielding 2 different test tapes.

Four sub-tests were designed, with different internal ordering of the 4 same-speaker sequences, in order to neutralize interactions between the different same-speaker sequences. Thus, the experiment yielded 8 different test tapes, all containing the same 176 stimuli (word sequences) but different with respect to presentation condition and internal sub-ordering.

3.2 Material

The 176 digitized stimuli were DA-converted (10kHz; 12 bits) and re-recorded onto 8 separate audio tapes. Test items were preceded by 4 trial items and 10 filler items, and followed by another 10 filler items; trials and fillers were identical for all tapes.

The inter-stimulus interval was 3 sec; total time of each test tape was about 14 minutes.

3.3 Subjects and Procedure

Six (native Dutch speaking) subjects listened to each tape, yielding a total of (8x6=) 48 subjects. Most of them were undergraduate students in various Linguistics and Language studies. Their participation was voluntary, but they were paid a small amount (Hfl. 5,=) for their services. Subjects were assigned at random to one of the test tapes.

Subjects received written instructions, as well as a response booklet. For each stimulus item, this booklet gave two possible responses from which a forced binary choice had to be made. It was emphasized that they should not allow themselves to be influenced by the sometimes contrastive orthographies of the two possible responses (e.g. "zeis om" vs. "zij som").

Subjects listened to the test tape with closed headphones binaurally. Nine of them listened in a none-to-quiet room, the other 39 in a sound-treated booth. After the 4 trial items, the experimenter checked whether the instructions had been understood and the playback volume was comfortable, and gave additional oral instructions if necessary.

Responses agreeing with the boundary position as intended by the speaker were scored as 'correct', alternative responses as 'wrong'.

3.4 Results

Mean accuracy percentages for the various conditions are given in Table II below.

Table II:
Observed mean word boundary detection accuracies in percentages. Means for each bottom cell are calculated over 11(stimuli)x48(subjects)= 528 observations.

presentation	speaker	/VC#V/	/V#CV/	mean
	M1	91.8	66.4	
Sequences	M2	88.1	73.0	
	F1	87.0	73.0	
	F2	83.5	71.8	
	mean	87.6	71.0	79.3
	M1	93.8	71.0	
Randomized	M2	86.2	78.0	
	F1	90.5	74.5	
	F2	84.0	74.5	
	mean	88.6	74.5	81.6
mean		88.1	72.8	80.5

The dependent variable in the present experiment, viz. correct or wrong response, establishes a discrete (h.l. binary) random variable, following the binomial distribution with p=.5 and N=24 (subjects). However, since N.p>10, this distribution approximates the normal distribution so that the latter may be used as well (11).

Separate three-way analyses of variance were carried out with Speaker, Presentation and (intended) Boundary Position as main factors, integrating over subjects and words, respectively. From the resulting F-ratios, the minF' was calculated (3).

The SPEAKER variable yields an insignificant effect with minF'(3,82)<1. The same applies to the main variable which was of prime interest in this experiment, viz. PRESENTATION with minF'(1,3)=1.367 (insignificant). Thus, no significant difference in the proportion of correct responses (detection accuracy) could be observed between the two presentation conditions. Besides, the observed difference tends to be opposite to the prediction: subjects' word boundary detection is slightly more accurate in the Randomized condition as compared to the Sequences condition.

The only main factor yielding significance was BOUNDARY POSITION: minF'(1,13)=8.899; p<.025. As can be seen from Table II above, boundary detection accuracies were considerably higher in the /VC#V/ context as compared to those in the /V#CV/ context.

Significant interaction occurred between the factors Speaker and Boundary Position: minF'(3,90)=2.919; p<.05. Thus, detection accuracy between the two boundary positions (or contexts) was significantly different for the 4 speakers; the lowest difference was found for female speaker F2 (10.6%) and the highest difference for male M1 (24.1%). Other interactions did not reach significance.

4. DISCUSSION

Results of the present experiment show no significant effects of either Speaker nor Presentation. Although each of the four speakers under investigation employed to some extent different acoustical (durational) means to mark word boundaries in his (her) speech, these differences are not reflected in subjects' accuracy in word boundary detection. Listeners do not yield higher accuracy when listening to stimuli realized by one speaker to whom they could "tune in", as compared to the "Randomized" condition in which stimuli realized by four different speakers were presented.

These results allow for two possible explanations:

(a) Although word boundaries may be marked differently by different speakers, listeners pay simultaneous attention to all phenomena that may provide cues to word boundary location. That is, they do not focus on one acoustic cue which marks word boundaries for one speaker, switching attention to a different cue when stimuli realized by a different speaker are presented. Instead, listeners simultaneously focus on several phenomena which may or may not function to mark word boundaries, depending on who is speaking. Thus, they are "sensitive" to any of the cues the current speaker might possibly use. When switching to another speaker, they simply discard information provided by phenomena which do not help them, and rely more heavily on the phenomena which for this speaker assume the function of boundary markers.
Since all possibly relevant acoustical information for word boundary detection is monitored and evaluated continuously, the switching to different speakers has no effect on subjects' detection accuracy.

(b) The acoustical phenomena under investigation bear no perceptual relevance at all for word boundary detection. Although the four speakers realize significant differences for these acoustical markers (between /C#/ and /#C/ positions) to a different degree, these differences are perceptually irrelevant.
This interpretation of the results implies, that there are other acoustical cues, consistent between speakers, that systematically mark word boundary locations in fluent (Dutch) speech. These cues, yet unknown (but non-durational in nature), then have to be further investigated.

Since it is a quite common phenomenon that different acoustical cues simultaneously contribute to speech perception (as e.g. vowel length, VOT and F_0 all contribute to the voiced-voiceless distinction (9)), we feel that explanation (a) is the most likely. In a broader view, people generally use multiple cues to perceive significant aspects of their environment; our evaluation of other people, for example, is based on simultaneous impressions about their face, physical posture, what they say and how, and on their further behaviour. Probably, as in word boundary detection, numerous other (yet unknown) cues bear relevance as well. However, in order to accept explanation (a), we must disprove (b), i.e. it must be shown that the durational differences observed (viz. duration of the ambiguous intervocalic boundary consonant, and rise time of the post-boundary vowel) are perceptually relevant. If manipulation of these two parameters can be demonstrated to influence subjects' boundary detection, then explanation (b) must be discarded and (a) gains plausibility. Preliminary results suggest that this indeed seems to be the case; a more extensive study will be reported in the near future.

References

(1) BARRY, W.J. (1984) Perception of Juncture in English. In: M.P.R. van den Broecke and A. Cohen (eds.), Proceedings of the Tenth International Congress of Phonetic Sciences. Dordrecht/Cinnaminson NJ: Foris.

(2) BROWMAN, C.P. (1980) Perceptual Processing: Evidence from slips of the ear. In: Fromkin (1980), pp. 213-30.

(3) CLARK, H.H. (1973) The language-as-fixed-effect fallacy: A critique of language statistics in psychological research, J. Verbal Learning and Verbal Behaviour, 12:335-59.

(4) COLE, R.A. and J. JAKIMIK (1980) Segmenting Speech into Words, J. Acoust. Soc. Am. 64(4):1323-32.

(5) COOPER, W.E. (1979) Speech Perception and Production: Studies in Selective Adaptation. Norwood NJ: Ablex.

(6) FROMKIN, V.A. (1980) Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand. London: Academic Press.

(7) GARNES, S. and Z.S. BOND (1980) A Slip of the Ear: A Snip of the Ear? A Slip of the Year?. In: Fromkin (1980), pp. 231-39.

(8) GÅRDING, E. (1967) Internal Juncture in Swedish. Traveaux de l'Institute de Phonetique de Lund, VI. Lund: Gleerup.

(9) HAGGARD, M., Q. SUMMERFIELD and M. ROBERTS (1981) Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading F_0-cues in the voiced-voiceless distinction, J.Phonetics 9:49-62.

(10) HARRIS, M.O., N. UMEDA and J. BOURNE (1981) Boundary perception in fluent speech, J. Phonetics 9:1-18.

(11) HAYS, W.L. (1973) Statistics for the social sciences. London: Holt, Rinehart and Winston. second edition.

(12) KLATT, D.H. (1980) Speech Perception: a model of acoustic-phonetic analysis and lexical access. In: R.A. Cole (ed.), Perception and Production of Fluent Speech. Hillsdale NJ: Lawrence Erlbaum Associates.

(13) LEHISTE, I. (1960) An acoustic-phonetic study of internal open juncture, Phonetica 5(suppl):5-54.

(14) NAKATANI, L.H. and K.D. DUKES (1977) Locus of segmental cues for word juncture, J. Acoust. Soc. Am. 62:714-19.

(15) QUENé, H. (1985) Word boundary perception in fluent speech: a listening experiment, Progress Report Inst. Phonetics Utrecht 10(2):69-85.

ON FOUNDATIONS OF HISTORICAL PHONOLOGY

TIIT-REIN VIITSO

Department of Finno-Ugric Languages
Institute of Language and Literature
Tallinn, Estonia, USSR 200106

ABSTRACT

Investigation of historical phonology of a language group may include several cycles each consisting of five stages. Main problems concern the correctness of phonological solutions for source languages and the typological reliability of reconstructed phoneme systems and phonological changes. Perspectives of predictive historical phonology are discussed.

INTRODUCTION

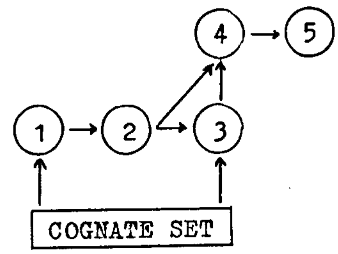
Historical phonology stems from the so-called historical phonetics. Despite its name, historical phonetics, actually, had to manage without any objective phonetical data about most languages whose history was dealt with. On the other hand, several historical phoneticians of the prephonological period possessed a remarkably good understanding of the possible directions of sound changes, of the conditioning role of the sound system in particular sound changes, and of the variant/invariant relationship. Hence it makes no sense to try to draw a strict borderline between historical phonetics and historical phonology on the basis of different authors' terminology. What is far more significant, is the existence of cycles and natural stages of investigating the historical phonology of related languages.

1. STAGES OF HISTORICAL INVESTIGATION

Ideally, any exhaustive study of the historical phonology of a set of related languages (e.g. of a language family) should begin from investigating groups of closely related languages (the first cycle) and then unite these groups and more remotely related languages step-by-step into bigger groups in order to repeat the procedure until all the related languages are included. Each such cycle consists at the utmost of five stages.

(1) Establishing for each positional (paradigmatic) class of phonemes (consonants or vowels) or phoneme sequences its set of correspondences on the basis of the cognate set of the language group.
(2) Reconstruction (a) of positional phoneme classes of the protolanguage *L of the language group on the basis of the correspondence sets and (b) of the consonant and vowel systems of *L on the basis of the reconstructed positional classes.
(3) Reconstruction of lexical items of the protolanguage *L in terms of the reconstructed phonemes on the basis of the cognate set.
(4) Reconstruction of the sets of ordered phonological changes necessary to derive all positional phoneme classes and all phoneme sequences of each language L of the language group from those of the protolanguage *L.
(5) Building a family tree or a family-tree-based net for the language group on the basis of ordered phonological changes.

The interrelations of a cognate set and of the five stages are presented on the following scheme:



Note, however, that any further stage of a cycle may cause corrections at some earlier stage.

2. PROBLEMS OF PHONOLOGICAL RECONSTRUCTION

The quality of an investigation depends on several factors:
(a) on the choice of the most realistic phonological solution for each related language (on a stage preliminary

to a historical study);
(b) on the exhaustiveness of the phoneme correspondence sets and on the quality of available distributional data for each correspondence (stage 1);
(c) on the amount of regular correspondences actually included in the reconstruction of positional phoneme classes of the protolanguage *L (stage 2);
(d) on the typological reliability of the reconstructed consonant and vowel systems (stages 2 and 4);
(e) on the typological reliability of reconstructed phoneme sequences (stage 3);
(f) on the typological reliability of all reconstructed changes (stage 4);
(g) on the choice of the most reliable historical solutions out of the set of competing solutions.

Note that the reliability of reconstructions does not guarantee their correctness (i.e. historical reality).
Nevertheless, only a reliable reconstruction can be correct although there are no correctness criteria for reliable reconstructions. Still, among several competing reliable reconstructions one may prove to be more reliable than the others.
2.1. There exist languages whose phonology presents no or few real problems at least within the scope of historical comparative studies. On the other hand, there are languages that create serious problems for any phonological school pretending to psychological reality of its solutions. E.g., there have been long-lasting discussions about the phonemic system of Estonian. One of its several nontrivial phonological problems is the solution of its monophthong + stop pattern series, cf. the series of minimal pairs (presented in the Finno-Ugric transcription) in Table 1.

Table 1

Q1		naGì	mak̆ki		màk̆kĭ	mak̄꜀kĭ
Q2			màk̆ki			
		māGi				
Q3		máGĭ		mákkĭ	mák̀꜀kĭ	
	mâG꜀kĭ	mâGĭ	mâk̆kĕ			

In Table 1 vowel and consonant length is indicated by means of upper diacritics, cf. (in the increasing order): (a) short vowels: ă a à; (b) long vowels: ȁ ā á â; (c) strong single consonants: k k̀ k̄; (d) homomorphemic geminate consonants: k̆k kk k̀k; (e) heteromorphemic stop clusters: G꜀k k꜀k k̄꜀k. Note that the second component in a heteromorphemic stop cluster begins in implosion. In addition to words with heteromorphemic stop clusters also the words mâGĭ and mâk̆kĕ belong to dimorphemic patterns with a root morpheme ending in a (super)long monophthong. All the words have stress in the initial syllable. The situation is complicated by the fact that Estonian has also a complicated system of diphthongs (26 types in my pronunciation), most of which pattern analogically to long vowels, and consonant clusters with there own problems; still most consonant clusters behave analogically to geminates when preceded by a short monophthong. All the monophthong + stop patterns participate in morphophonological alternations with each other. Native grammars of Estonian group all the patterns into three distinctive quantities, cf. Q1, Q2, and Q3 in Table 1; each quantity has, alongside the durational characteristics, also a characteristic pitch that has been, however, denied by some authors.
In any case, a complicated language like Estonian has many phonological solutions. Most of them are wrong and thus may serve as a source of a wrong history.
2.2. When reconstructing a phoneme system, any correspondence must be treated as a phone whose functional properties are unknown. Hence, an exhaustive set of phoneme correspondences should not include correspondences established on the basis of suspicious or ambiguous cognates. Inclusion of suspicious or ambiguous cognates and aberrant correspondences in the data body covered by reconstructions would probably falsify the history. There are, however, cases where most of the original environments of observable correspondences have been lost in the course of later changes. In such cases the reliability of reconstructions depends first of all on the number of correspondences covered by reconstructions on condition that no relatively big cognate set is ignored. Cf., e.g., the set of Permic (Finno-Ugric), i.e. Proto-Komi (the 1st row) and Proto-Udmurt (the 1st line) correspondences of the vowels of the first syllables in Table 2. i̮ e̮ e̮ and u̮ o̮ are, correspondingly, illabial and labial central vowels, o̦ e̦ e̦ are rised mid vowels. For each correspondence, the number of reliable cognates is indicated. The number is underlined if the correspondence occurs in the stem-final position.

Table 2

	*u	*o	*a	*i̮	*u̮	*o̮	*e	*i
*u	78	2	1	24	18	1	2	
*o̦	29	3	16	8	20		3	
*o	52	3	9	6	3	3	2	
*a	16	4	91		2	5		
*e̮	2	28	9			5	27	
*e̮̦	7	2	1	15		16	2	
*i̮	14	3	1	130		2		8
*e	4	24	3	2		2	16	2
*ê	14						7	3
*i	4	8		12	1	2		64

2.3. The problem of typologically reliable reconstructed phoneme systems usually results from the presence of a great number of attested correspondences whose former complementarity has been eliminated by some later innovations. In such cases usually either the actual number of protophonemes will be multiplied or the original structure of the phoneme system fantastically distorted in the course of reconstruction. E.g., for Proto-Permic vowel systems containing 12-15 short vowels have been proposed. Such reconstructions are clearly unreliable.
There has been a lasting discussion about a typologically reliable Proto-Indo-European stop system. The discussion has concerned the problems of markedness, voicedness, aspiration, glottalization, and murmur of stops both in Indo-European and in other, relatively badly investigated language families. Interesting enough, the discussion has not led to new phonetical investigations of different stop triads even in modern Indo-European dialects, e.g., the Armenian ones. In other words, even the typology of the historically most interesting dialects of the world's best investigated language family rests mostly on impressions and not on firm data.
The situation is still worse in the field of the typology of changes. Although there has been steady progress in the field of the theory of change, the typology of individual changes is still almost nonexistent. We know, e.g., that long non-high vowels tend to rise, and not vice versa. On the other hand we know that all long monophthongs can be diphthongized.
According to Z. Zinkevičius the Lithuanian correspondence series uo ~ o ~ o̦u ~ ū represents the Proto-East-Baltic *ō and not *ou from Proto-Indo-European *ōu [1]. Probably he has in view three independent changes *ō > uo, *ō > o̦u, *ō > ū.
The South Estonian long mid vowel rising *ō *ȫ *ē > ū ǖ ī (the South Estonian vowels are lowered high vowels, contrasted to high vowels ū ǖ ī) in quantity 3, cf. 2.1, and the North Estonian diphthongization of long mid vowels into uo üö ie both in quantity 2 and quantity 3 are usually considered to represent the same change chain with a specifically South Estonian monophthongization of uo üö ie into ū ǖ ī [2]. This contradicts the facts (1) that South Estonian has retained long mid vowels in quantity 2 whereas no Estonian dialect has diphthongized its long mid vowels only in quantity 3 and (2) that in quantity 3 the second components of the North Estonian uo üö ie seem to be somewhat longer than the first ones and tend to lower. Besides, in Kodavere (East Estonian) described by L. Kettunen [3], long mid vowels had risen and merged in ū ǖ ī in quantity 3 and diphtongized to ua üä iä in quantity 2; at the same time long low vowels *ä *ȁ had diphthongized to ua iä both in quantity 2 and quantity 3. Hence, both Kodavere and South Estonian indicate that rising and diphthongization of long mid vowels have different prerequisities: quantities 2 and 3 differ both in duration and in pitch. Likewise, it is possible that these prerequisities are to some extent universal, and probably there exist also conditions when long mid vowels reveal no tendency to change. Probably there are also several other changes or tendencies that can be actualized only under certain "hidden" conditions.
In view of that it is meaningful to look for such conditions comparing both dialects that have retained an old feature and those that have changed it. Apparently, often conditions are preserved after the change has taken place. E.g., in Livonian (a Baltic Finnic language spoken in the northern tip of Kurland, Latvia) all short vowels were lengthened (also in diphthongal nuclei) before a short vocalic or sonorant coda after the Livonian coda polarization had taken place in long syllables with plain tone and no long monophthong. Later u (except after ì) and i (after è) were dropped after such lengthened vowels. Now a new round of vowel lengthening is in progress. Short initial components in diphthongs that have no contrastive pair beginning in a long component tend to lengthen [4]. This tendency concerns alongside syllables with plain tone also those with stød in diphthongs whose second component u or i does not morphophonologically alternate with v or j. Cf. Table 3 with Livonian polyphthongs. The lengthening tendency concerns diphthongs of classes 2 (tend to shift to class 3) and 4 (tend to shift to class 5). As this lengthening still remains in the framework of subphonemic free variation, Livonian must have retained the conditions necessary for such a lengthening for a considerable stretch of time.

Table 3

	u					i				
	Plain			Stød		Plain			Stød	
	1	2	3	4	5	1	2	3	4	5
u							uì	ùi	u'i	ū'i
o	oŭ	où	(ō)	o'u				òi		ō'i
a						aĭ	aì	ài	a'i	
o̮							o̮ì	ò̮i	o̮'i	
e̮		e̮ù	(ē̮)	e̮'u			e̮ì	è̮i	e̮'i	
ä		äù	(ǟ)	ä'u						
e							eì	(ē)	e'i	
i		iù	ìu	i'u						
uo						ŭoĭ	ŭoì	ùoi	ŭo'i	u'oi
ie		ĭeù	(ìe)	ĭe'u						

Estonian, on the other hand, reveals no tendency to lengthen the initial components of diphthongs even when they have a relatively short final component (in quantity 2). North Estonian dialects rather represent different stages of lowering of the final components u and i under the influence of the quality of the following consonants.

3. PREDICTIVENESS IN HISTORICAL PHONOLOGY

Apparently, establishing the necessary conditions of changes like those discussed in 2.3 is a task of historical phonology. Such a task means that historical phonology must become predictive at least in the weak sense of predictiveness: it must be capable of estimating the possibility or impossibility of one or another change. Doubtless, predictive historical phonology has higher requirements for the quality of synchronical phonological studies than does the current synchronical phonology. Synchronical phonology can often well manipulate any data of a local dialect or a standard language having only an impressionistic knowledge of manner and place of articulation. Predictive historical phonology needs considerably more concrete knowledge. One must be able to satisfactorily characterize the differences of "the same phoneme" (a) in different positional classes of the same language or dialect and (b) in similar positional classes of different languages or dialects. Nevertheless, collecting the relevant data on different phonological changes and their prerequisities is a task of typology rather than historical phonology. Hence, phonological typology must change from a branch that eagerly deals with collection and classification of both correct and incorrect impressionistic data into one that carefully checks up the correctness of the data it manipulates.

The perspective of predictive historical phonology demands that the role of abduction in phonological changes must be reviewed. Abductive changes, singled out by H. Andersen, are claimed to be unpredictable [5]. Actually there are maybe only two classes of unpredictable phonological changes: (a) sporadic and (b) those conditioned by speech disorders of a prestigious member of a little language community. The most striking examples of abductive change are rather chains of entirely natural single changes. The output of other examples of abductive change still contains features known from the input stage of the change. Such changes result from the effect of a set of universal tendencies whose actual number, scope and structure is still unknown. E.g.:

(1) The number of phonological rules in a grammar tends to be minimal.
(2) The domain of a phonological rule tends to be minimal.
(3) The phonological complexity of a phoneme sequence (syllable, stress group, word) tends to vary periodically.
(4) The number of phonemes in a phoneme system tends to be minimal.
(5) The length of allomorphs of a language tends to be minimal.
(6) Phoneme mergers tend to follow the principle of minimal articulatory efforts.

The first three tendencies are, probably, consequences of the tendency to minimize the volume of brain work. Thus it is more economical to memorize frequent inflectional forms and phrases than to compose them again and again. Tendencies (4) and (5) have partially opposite effects: tendency (5) may cause an increase both in the number of phonemes and in homonymy. As these tendencies act persistently they must be considered both in historical and synchronical phonology.

REFERENCES

[1] Z. Zinkevičius, "Lietuvių kalbos dialektologija", Vilnius, 1978.
[2] L. Kettunen, "Eestin kielen äännehistoria", Helsinki, 1962.
[3] L. Kettunen, "Lautgeschichtliche Darstellung über den Vokalismus des kodaferschen Dialekts", Helsinki, 1914.
[4] H. Pajupuu, T.-R. Viitso, Livonian polyphthongs. "Estonian Papers in Phonetics 1984-1985", Tallinn, 1986. 96-134.
[5] H. Andersen, Abductive and deductive change. "Language" 49 (1973). 765-93.

ZUR BEWERTUNG VON PHONETISCH-PHONOLOGISCHER REKONSTRUKTION

HERBERT PENZL

Dept. of German
University of California
Berkeley, CA 94720

1. Realitätswert der Rekonstruktionen

In einem von der Forschung wenig beachteten Buch habe ich Typen und Methoden der Rekonstruktion ausführlich behandelt. Der Ausdruck Rekonstruktion "Wiederherstellung") wird meist für das Erfassen der schriftlich unbelegten sprachlichen Vorgeschichte, also von Lauten, Phonemen, Morphemen, Wörtern, Flexion, Syntax, Wortschatz aus späterem belegtem Material verwendet. Haupttypen wie innere (interne) Rekonstruktion, vergleichende Rekonstruktion, umgekehrte Rekonstruktion, Prärekonstruktion sind in der Gegenwart öfter beschrieben worden (Penzl 1972, S. 116ff.). Die früher allgemeine Praxis, Rekonstruktionen, ob es nun Einzelphoneme, Phonemverbindungen, Wortformen usw. sind, von den tatsächlich belegten durch ein Sternchen zu unterscheiden, hat man im allgemeinen beibehalten. Es bleibt oft unklar, ob Weglassen der Sternchen Hypostasierung aus gro3em Vertrauen auf die Methode oder nur Ungenauigkeit der Bezeichnung bedeute. Bei jeder Rekonstruktion ist die Frage des Realitätswerts, der geschichtlichen Wirklichkeit, wichtig. Die Einstellung der Forschung zu besternten Formen hat zwischen "blo3en Formeln", "Abstraktion ohne Realität", "formelhafter Zusammenfassung sprachlicher Entsprechungen", "Annäherung an die geschichtliche Wirklichkeit", "Art phonemischen Diagramms der Grundformen" geschwankt (Penzl 1972, S. 113). Hermann (1907, S. 62) schlug vor, ein Kreuz für "wirkliche Rekonstruktion" wie idg. +esmi 'bin', ein Sternchen für blo3e Formeln zu setzen. Aus der Summe all dieser Einzelrekonstruktionen ergibt sich die genetische Rekonstruktion einer historischen oder prähistorischen Sprachstufe, Ursprache oder Zwischenursprache wie Indogermanisch, Urgermanisch, Westgermanisch, Voralthochdeutsch aufgrund von belegten Formen in Tochtersprachen, Dialekten. Was den Begriff Ursprache (engl. proto-language) anbelangt, so dringt in letzter Zeit die Einsicht durch, da3 sie nur dann als natürliche Sprache anzusehen ist, wenn zumindest e i n Text, also das pragmatische Resultat einer Sprachhandlung eines Sprachträgers vorliegt. Das trifft wegen der einen Satz langen Inschrift auf dem Horn von Gallehus von 400 n. Chr. für das Nordisch-Westgermanische zu (Penzl 1975, S. 69ff.). Wir können mit unseren Rekonstruktionsmethoden wohl Grammatik und Sprachregeln, aber keine prähistorischen Texte (trotz A. Schleicher und H. Hirt) rekonstruieren, weil auch Sprachträger und Sprachakte weder vergleichend noch intern rekonstruierbar sind.

Wenn hier auch nur von phonetisch-phonologischer Rekonstruktion die Rede sein soll, so ist die morphologische Rekonstruktion eigentlich miteinbegriffen, weil phonologische vergleichende Rekonstruktion stets im morphologischen, sogar lexikalisch-semantischen Rahmen erfolgen mu3. Innere Rekonstruktion kann synchronisch und diachronisch sein und ist auch oft für die historische Zeit, d.h. die Zeit mit Textbelegen notwendig, wenn System, Variation, Wandel, Verbreitung nur lückenhaft schriftliche Bezeichnung gefunden haben.

2. Phonologische und phonetische Rekonstruktion

Die Unterscheidung zwischen phonologischer (phonemischer) und phonetischer Rekonstruktion ist wichtig. Phonologische Rekonstruktion bedeutet das Erfassen von Phonemen und Phonemsystem mit den wichtigsten distinktiven Merkmalen (Eigenschaften). Phonetische Rekonstruktion hat als Ziel die Beschreibung von Lautwert und Lautinventar, die akustisch, auditorisch, organogenetisch erfolgen kann. Wenn wir nach dem letztgenannten Typ verfahren, kann jeder Laut einer Sprache beschrieben werden, wenn wir (1) Artikulationsorgan, (2) Artikulationsort, (3) Artikulationsweise, (4) Artikulationsenergie, (5) Artikulationsdauer und (6) eventuelle Koartikulationen angeben können. Von mhd. /d/ in der, das mit dem lat. Zeichen d geschrieben wird, nehmen wir an, da3 es (1) apikal oder dorsal, (2) dental oder postdental oder alveolar, (3) Verschlu3laut, (4) lenis, (5) stimmhaft oder stimmlos (Artikulation der Stimmbänder) war. Schreibung, Vorgeschichte, nhd. Weiterentwicklung, Alternanz mit t im Auslaut (mhd. lant, Gen. landes) lassen keine weitergehende Bestimmung zu. Im Falle der ahd., mhd. Sibilanten, die <s> und <z> geschrieben werden, z.B. ahd. thaz, mhd. daz 'das', ahd. thes, mhd. des 'des' (Gen.) ist nur die phonemische Opposition bis zum frühnhd. Zusammenfall deutlich. Die Zeichenwahl mit häufigem <zz> (mhd. wazzer 'Wasser', germ. *t) deutet auf stimmlose Fortis gegenüber einer Lenis <s> (mhd. wesen 'sein'). Der Zusammenfall erfolgt nach der mhd. Entwicklung eines Schibilanten aus ahd. /sk/: mhd. waschen. Aus altslowenischer Orthographie und gegenseitiger Wiedergabe slawischer und ahd. Sibilanten in Namen geht nur hervor, da3 /s/ einem Schibilanten näher als /z/ war. Sonst versagen die weiteren Quellen für Lautbestimmung wie Reime und Assonanz, Weiterentwicklung, Herkunft; orthoëpische Beschreibungen fehlen überhaupt (vgl. Penzl 1971, S. 71f., Penzl 1986, S. 38f.). Im Altfranzösischen scheint es ein gleiches Sibilantenpaar gegeben zu haben. M. Joos sah in /s/ apikale und in /z/ dorsale Artikulation, die aber in der Gegenwart auditorisch schwer zu unterscheiden ist. Im Baskischen werden nach seiner Angabe zwei Sibilanten so unterschieden. Meine Vermutung von retroflexer Artikulation von /s/ geht eigentlich auf meine Kenntnis distinktiver retroflexer Schibilanten im Paschto Afghanistans zurück (Penzl 1955, S. 31f.). Im Grunde bin ich überzeugt, da3 das Beweismaterial in diesem und vielen anderen Fällen zu einer genauen phonetischen Identifizierung der historischen phonemischen Opposition nicht ausreicht. Diese Einstellung hat ein amerikanischer Linguist als linguistischen "Nihilismus" gerügt.

3. Prähistorische Lautanalyse: Allophonie

Für prähistorische Rekonstruktion von Lautwerten sind wir auf den Schlu3 aus Schriftzeichen und Schreibungen in Wortentsprechungen angewiesen. Aus lat. decem, griech. deka, got. taihun, ahd. zehan, ai. daśa usw. rekonstruieren wir eine idg. Grundform *dek'm̥, in der jeder Laut sich in einer Fülle von Einzelentsprechungen belegen lä3t und z.B. im Anlaut in umgekehrter Rekonstruktion als Probe sich die Laute der idg. Tochtersprachen durch natürliche und wahrscheinliche Lautwandlungen divergierend aus der idg. Media ergeben: *d zu lat., griech., ai. d, got. t durch "Grimms Gesetz", nach dem aus der idg. Media regelmä3ig die urgerm. Tenuis *t wurde. Um ahd. zehan mit Affrikate neben asächs. tehan, ae. tien zu erklären, nehmen wir Allophone mit starker Aspiration an, die wir z.B. für *t nach *s (ahd. stein, asächs., afries. stēn, ae. stān) weder urgerm. noch idg. ansetzen.
In daśa 'zehn' mit Sibilant wie auch in der aslaw. und der lit. (š) Entsprechung haben wir gegenüber lat., griech. [k], germ. [χ] (nach Grimms Gesetz) die berühmte centum/satem "Isoglosse". Wir schreiben palatales <k'> nach dem Vordervokal *e, um die Sibilanten zu erklären, sehen aber nicht mehr darin eines von drei Gaumenphonemen wie in der traditionellen Rekonstruktion, sondern ein Allophon von */k/. Die Verteilung lä3t den einfacheren Ansatz von Allophonie statt Gliedern einer Phonemreihe zu; das entspricht mehr den Prinzipien der Entfaltungstheorie Otto Höflers, ohne aber der alten Stammbaumtheorie zu widersprechen. Manche Forscher werden diese Allophone nur für ein frühes Entwicklungsstadium (**/k/) als richtig ansehen. In unseren germanischen und indogermanischen Entsprechungen in ihrer Beziehung zu den Grundformen der Ursprache zeigt sich deutlich der Rekonstruktionszirkel: aus den Lautwerten der Tochtersprachen rekonstruieren wir diejenigen Lautwerte der Grundsprache als Phoneme oder Allophone, die ihrerseits eine Motivierung für die Divergenz abgeben können. Der Konvergenz der vergleichenden Rekonstruktion entspricht die Divergenz der umgekehrten Rekonstruktion.

4. Prähistorische Phonemsysteme: Die indogermanischen Vokale

Als idg. Vokalphoneme hat man folgende rekonstruiert:

Kurzvokale: *i *e *u *o *a *ə, auch *ə$_2$ (schwa secundum)
Langvokale: *ī *ē *ū *ō *ā
Diphthonge: *e + *i, *u — *o + *i, *u
*ē + *i, *u — *ō + *i, *u
*a + *i, *u — *ā + *i, *u

Viele Forscher setzen *i̯ und *u̯ als Allophone der unsilbischen Halbvokale *j (englisch: *y) *w an. Ein Diagramm mit Anordnung der Vokale nach ihrer Artikulation im Mundraum, wie ich es anführe (Penzl 1975, S. 43), ist im allgemeinen nicht üblich. Man vermeidet gerne genauere phonetische Bestimmung. Ich kenne auch keine Tabelle mit drei idg. Quantitäten, die man nach dem rekonstruierten Dreimorengesetz für "Nebensilben" aufgestellt hat, und m.E. recht zweifelhaften Realitätswert besitzt (Penzl 1975, S. 44, 62).
Für Wechsel (Alternanz) zwischen Langvokal und entsprechendem Kurzvokal oder Kurzvokal und Schwund oder Schwa ergibt sich eine Erklärung aus den Akzentverhältnissen. Der qualitative Ablaut, d.h. der Wechsel *e/*o, auch bei Langvokal und Diphthong, lä3t sich phonotaktisch im rekonstruierten Idg. nicht erklären. Intonation als Faktor lä3t sich mit verfügbaren Methoden nicht erfassen. Sievers' Schallanalyse hat sich ja als utopisch, im strukturalistischen Sinne als "Hokuspokus" erwiesen (vgl. Penzl 1975, S. 48f.).

5. Die indogermanischen Konsonanten

Für das idg. Konsonantensystem hat man au3er den Sonorlauten *w *j *l *r *m *n (mit velaren und palatalen Allophonen) auch lange silbische Liquide und Nasale rekonstruiert: z.B. oben *m̥ in *dek'm̥, denen vielleicht etwas Formelhaftes anhaftet.
Das traditionell rekonstruierte Obstruentensystem weist folgende Geräuschlaute auf, für die man gerne die lateinischen Fachausdrücke statt der phonetischen verwendet (vgl. Penzl 1975, S. 47):

Tenuis *p *t *k (mit Allophonie oder Phonemdivergenz)
[Tenuis aspirata]
Media *b (selten) *d *g
Media aspirata *bh *dh *gh
Sibilant *s

Die aspirierte Tenuis des Sanskrit und Griechischen in der Ursprache mag universalistisch, wie z.B. R. Jakobson hervorhob, das System symmetrischer und realistischer gestalten, aber es macht es von Standpunkt der anderen Tochtersprachen noch "unindogermanischer" mit seinem Mangel an Reibelauten und der Aspiration als distinktivem Merkmal. Für eine "buchstäbliche" Media aspirata nach der traditionellen Phonetik, also [bh], [dh] usw. gibt es nicht einmal in den indischen Sprachen der Gegenwart Unterstützung, da neuere Forschung besonders von P. Ladefoged die "Aspiration" eher als Art Glottalisierung erwiesen hat. Vom Standpunkt des Germanischen z.B. (vgl. Polomé 1982) ist die distinktive phonologische Dreiheit der Obstruentenreihen in der rekonstruierten Ursprache das Wichtige, kaum irgendeine angenommene phonetische Realisierung der Media aspirata, deren Bezeichnung als formelhaft angesehen werden kann. Abgesehen von diesen spezifischen Lautwerten bestätigen die Kombinationsregeln wie etwa Media + Media, nie Media + Tenuis u. dgl. sowie Dissimilations- und Assimilationsregeln, wie sie Gra3manns Gesetz und Bartholomaes Gesetz beschreiben, die gegenseitige Opposition der drei Reihen, auch z.B. die Stimmhaftigkeit (in Media gegenüber Tenuis) als distinktives Merkmal.

6. Die Laryngale und idg. Prärekonstruktion

Kein traditionelles Diagramm der idg. Konsonanten enthält den einen Laryngal, der aufgrund der traditionellen Entsprechungsmethode des Rekonstruierens eigentlich anzusetzen wäre: im Hettitischen pahs- 'schützen', lat. pāsco 'lasse weiden', pāstor 'Hirte' (Lindeman 1970, S. 28). Die Entsprechungen sind nicht zahlreich und für den Anlaut z.B. von Heinz Kronasser angefochten worden. Ein einsames *H wäre im Konsonantensystem unintegriert und die Anhänger der Laryngaltheorie rechnen mit mindestens drei Laryngalen: *H$_1$ *H$_2$ *H$_3$. Die Theorie geht auf die Bemühung von Ferdinand de Saussure zurück, den idg. Ablaut, besonders den mit Langvokal als Grundstufe in "schweren Wurzeln", phonotaktisch zu erklären. Das führte zunächst zur Prärekonstruktion von voridg. Vorstufen der Langvokale *ē *ā *ō aus Kurzvokal plus Laryngalen, dann auch zur Prärekonstruktion der Kurzvokalvariation *e /*o und seltenem *o /*a im Ablaut durch Einwirken eines idg. geschwundenen vorhergehenden Laryngals.
Diese Prärekonstruktion (Penzl 1972, S. 135ff.), die als innere Rekonstruktion (**e) auf einer Rekonstruktion (*e *o usw.) aufgebaut ist, nimmt also eine Dekonstruktion nicht nur aller idg. Langvokale, auch eine Konvergenz aller mittleren und niedrigen Kurzvokale (*e *o *a) mit Divergenz einer allgemein idg. geschwundenen Laryngalreihe (*H$_1$ *H$_2$ *H$_3$) an. So ergibt sich für ein Urururidg. ein einsamer Kurzvokal ***e, für den man durchaus trotz typologisch-universalistischer Bedenken Realitätswert beansprucht hat, wobei man kaukasische Sprachen als Parallelen anführte. W. F. Lehmann (1952, S. 112) ging für sein Urururidg. ("pre-stress stage of pre-IE") noch weiter, indem er neben vier allerdings meist als Obstruenten phonetisch bezeichnete "laryngeals" ein überhaupt vokalloses System mit einem "non-segmental phoneme, syllabicity" prärekonstruierte. Diese Wendung ins Abstrakte sollte eigentlich nicht überraschen. Auch für die Laryngale war zuerst die Bestimmung nur phonemisch mit sehr sporadischer phonetischer Beschreibung. Die Gegner der Laryngaltheorie haben besonders in ihrer späteren erweiterten Form sie als reine Hokuspokuslinguistik ohne historischen Realitätswert charakterisiert (Penzl 1972, S. 136f.).
Es ist aber bemerkenswert, wie viele Anhänger unter namhaften Indogermanisten die Laryngaltheorie gefunden hat. Auch die Einvokalthese hat viele überzeugte Anhänger. Gilt etwa Vokalarmut als besonders primitiv und archaisch? Soll ein allmählicher Zuwachs an distinktiver Sonorität die sprachliche Urentwicklung widerspiegeln? Nimmt man an, da3 die menschliche Sprache zuerst Konsonanten zur Kommunikation verwendete? Kann etwa prähistorische Dekonstruktion der verschiedenen Vokalsysteme und Ansetzen neuer Konsonantentypen auch andere Sprachfamilien auf einen gemeinsamen Nenner mit dem Idg. zurückführbar machen? Die Geschichte der Laryngaltheorie zeigt, da3 tatsächlich mit dem Ansetzen von Laryngalen auch Beziehung zum Semitischen und Uraltaischen verknüpft wurde. Oder sah man in den phonetisch zuerst recht vage beschriebenen Laryngalen (wie auch einmal bei den Schnalzlauten) besonders "primitive", also zur Prärekonstruktion besonders geeignete Laute? Die Ansicht, da3 ein Einvokalsystem vom Standpunkt der idg. Sprachen als "exotisch", daher als besonders glaubwürdig primitiv erscheint, unterstreicht einen ikonischen Faktor in der Rekonstruktion, der m.E. kaum vertretbar ist.

7. "Exotische" Rekonstruktion?

Es war wohl vor allem die angesetzte Reihe der aspirierten Medien (siehe 5. oben) die Ursache, da3 in steigendem Ma3e seit den letzten Jahren phonetische Umdeutungen des idg. Konsonantensystems veröffentlicht werden. Phonologisch sind damit auch neue distinktive Merkmale verbunden. Der prinzipielle Einwand gegen die phonetische Realität der traditionellen Rekonstruktion des Konsonantensystems wird allerdings in keiner Umdeutung behoben, ja die Distanz zwischen System der Ursprache und System der Tochtersprachen scheint nur noch vergrö3ert zu werden, so da3 wir kaum noch einen Rekonstruktionszirkel ansetzen können. Polomé (1982, S. 54f.) erwähnt einschlägige Vorschläge von Joseph Emonds, Paul J. Hopper (1973), Jens E. Rasmussen, Allan R. Bomhard und als besonders wichtig T. V. Gamkrelidzes und V. V. Ivanovs Artikel (1973). Vennemann, der schon früher (1979) die Aspiraten des Sanskrit nach P. Ladefogeds Forschungen als glottalisiert analysierte, begrü3te enthusiastisch die neuen Rekonstruktionen (Vennemann 1986), in denen z.B. die Media durch glottalisierte Verschlu3laute ersetzt wurde. Phonetische und universalistische Forschung der letzten Jahre hat die Typen und weite Verbreitung der Glottallaute zeigen können (Greenberg 1970). Vennemann zitierte mit Genugtuung die "Inventare der glottalischen Rekonstruktion" in Ian Maddiesons Lautstatistiken (Maddieson 1984).
Wir setzen in unseren Rekonstruktionen voraus, da3 im Laufe von Tausenden von Jahren die physiologisch-akustischen Voraussetzungen der Lautproduktion die gleichen sind. Es gibt für die Zeit der menschlichen Sprache kein anatomisches oder sonstiges Beweismaterial, das dagegen spricht. Es gibt auch keinen Beweis dafür, da3 das Ansetzen von "exotischer", d.h. kaum mehr oder nur "peripheral" idg. belegter Artikulationsweise eine Auswirkung der Ansicht ist, da3 wir mit physiologisch-anatomischen Änderungen rechnen sollten. Vielleicht überschätzen aber manche Forscher die Wichtigkeit au3eridg., ja "exotischer" Parallelen für eine Rekonstruktion. Das mag für die erwähnten baskischen Sibilanten, die Schibilanten des Paschto, kaukasische Einvokalsysteme und die Glottallaute in Südasien und Afrika, sogar im Sindhi gelten.
Dies ist nicht die geeignete Gelegenheit dazu, also möchte ich hier nicht meine eigene Rekonstruktion der idg. Obstruenten vom Standpunkt des Germanischen an Stelle der oben besprochenen vorschlagen, schon deswegen nicht, weil ich genaue phonetische Bestimmung prähistorischer (und oft historischer) Laute für unmöglich und irrelevant halte. Zur Erklärung der germ. Lautverschiebung ("Grimms Gesetz") würde ich aber folgende Allophone ansetzen: für die Tenues, wie bereits erwähnt, stark aspirierte, aber noch nicht affrizierte Allophone, für Medien auch stimmlose Allophone, für die Media aspirata auch frikative Allophone. Das stimmt mit den Methoden der Entfaltungstheorie überein und schlie3t auch glottalisierte Allophone wegen des Indischen keineswegs aus. Aber das Indogermanische soll ja nicht nur für das Indische, sondern auch für das Germanische die Grundsprache sein.

Gamkrelidze, T. V. und V. V. Ivanov. 1973. Sprachtypologie und die Rekonstruktion der gemeinindogermanischen Verschlüsse. Phonetica 27, S. 150-6.

Greenberg, Joseph M. 1970. "Some generalizations concerning glottalic consonants, especially implosives." IJAL 36, S. 123-45.

Hermann, Eduard. 1907. "Über das Rekonstruieren." ZfvglSpf. 41, S. 1-64.

Hopper, Paul J. 1973. Glottalized and murmured occlusives in Indo-European. Glossa 7, S. 141-66.

Lehmann, W. F. 1952. Proto-Indo-European Phonology. Texas University Press.

Lindeman, F. O. 1970. Einführung in die Laryngaltheorie. Sammlung Göschen.

Maddieson, Ian. 1984. Patterns of sounds. Cambridge University Press.

Penzl, Herbert. 1955. A Grammar of Pashto. A descriptive study of the dialect of Kandahar, Afghanistan. Washington, D. C. ACLS.

____. 1971. Lautsystem und Lautwandel in den althochdeutschen Dialekten. München, Max Hueber.

____. 1972. Methoden der germanistischen Linguistik. Tübingen, Niemeyer.

____. 1975. Vom Urgermanischen zum Neuhochdeutschen. Eine historische Phonologie. Berlin, Erich Schmidt Verlag.

____. 1986. Althochdeutsch. Eine Einführung in Dialekte und Vorgeschichte. Bern, Peter Lang.

Polomé, E. C. 1982. Germanic as an archaic Indo-European language. In: Festschrift für Karl Schneider, S. 51ff. Amsterdam, Benjamins.

Vennemann, Theo. 1979. Grassman's Law, Bartholomae's Law and Linguistic Methodology. In: Linguistic Method, Essays in honor of Herbert Penzl, S. 557ff. I. Rauch und G. Carr, Hg. den Haag, Mouton Publishers.

____. 1986. Neuere Entwicklungen in der Phonologie. Berlin, Mouton de Gruyter.

SYNTAGMATIC VERSUS PARADIGMATIC APPROACH IN PHONOLOGICAL EVOLUTION

Herbert Galton

The University of Kansas
Lawrence, Kansas, U.S.A.

Abstract

In der Erklärung historischer Lautentwicklungen ist man letzthin zu sehr von den Beziehungen innerhalb des phonologischen Systems ausgegangen, d.i. der paradigmatischen auf Kosten der syntagmatischen Stellung der Laute. Es ist Zeit, die letztere wieder in ihre Rechte einzusetzen, und zwar in der Form der Silbenstruktur, innerhalb derer die Laute allein ihre Wirklichkeit haben. Dies wird am Gemeinslavischen illustriert, wo die Sonorität den gemeinsamen Grundsatz sowohl für die Einteilung der Rede in Silben wie auch für die Entwicklung des Silbenkerns, dem keine Coda folgte, abgab.

Phonetic explanations of sound changes have somewhat gone out of fashion of late. Perhaps this is a natural reaction to the fact that with the advent of structuralism, it was discovered that the speech sounds of a language hang together in a sort of system regulated by a set of internal relations. What more natural than to believe that these relations also preside over the evolution of the systems, thus endowing them with a sort of creative force of their own, working in the direction of a closer integration. It was also obvious that in language, the segmental phonemes would be the likeliest objects of such an approach, since they are farthest removed from the representation of our untidy thoughts on the one hand, and on the other are subject to the constraints of the vocal organs with their limited number of positions. With the further refinement of technical data, acoustic classes emerged beside the traditional physiological ones, seemingly capable of a much higher degree of abstraction from the actual phonic material (even though nobody has ever heard compactness or diffuseness). The substance seemed to fade beside the network of relations, the unilinear sequence of sounds receded as against the paradigmatic arrangements of the elements.

I may be permitted to point out that in natural science, taxonomy has never, to my knowledge, been credited with a driving force of its own. What it represents is very largely the record of the interaction of its elements - plants struggling against plants, animals struggling against animals, the whole subject to the varying conditions of the environment. Few people still believe that the whole of the natural kingdom rose into being by <u>fiat</u> and <u>then</u> came to fill in the environment. Rather on the contrary, we believe that the environment created the species or at least changed them into what they are now. There never was a stage in which they were not profoundly affected by their environment, which includes every other species of plants and animals as well. Underlying it all is the great will to live (Schopenhauer, though he was unaware of evolution).

Underlying all language is the will to communicate. But as is well known, the other great force in nature, including human, is inertia, which evolution, under the dire threat of necessity, has to overcome. We are well aware that inertia is a powerful force also in the evolution of language, where it constantly has to be overcome by the need to express oneself, and the expression must take place by means of discrete and distinctive elements. Inertia would merge them in one inarticulate primeval cry.

Here we are back to the distinctions which, as we know, can be arranged in a "meaningful" pattern, the parts of which hang together in certain (cor) relations. But all relations in the phonological system bear on sets of phonemes and their realizations; it is not as individual entities, though correlated, that they developed. Exactly as plants and animals, all the way up to man (who became man through the social use of language), developed in a particular habitat, did sounds develop in their natural environment, and this natural environment is the word, or more precisely the syllable within the word, if it has more. All phonemes are abstracted from the positions in which they occur, and it was those which have shaped them, unless we want to go all the way back to Brugmann and assume with him that "der Anlasz zur Änderung des Lautes in seiner Eigenart zu suchen ist." We might as well assume that the incentive for the evolution of natural species lay in their specific nature. When we compare stage B of a language with stage A, represented by their phonological systems, we are almost unavoidably subject to an optical illusion, and that is, that the system as such has changed somehow on its own account. Especially if we find so-called fuzzy points at one place, gaps in another, a more systemic relationship in a third, we are bound to credit the system itself with a driving force, forgetful of the fact that all the phonemes are abstracted from the concrete (phonetic) words where they occupy a specific position in the syllable. Such an approach would, therefore, overemphasize the paradigmatic aspect, against which it can be contended that the syntagmatic aspect, allowing for the concatenation of sounds in their natural sequence, should be asserted as an equal partner in evolution. Much of this has been worked into the history of linguistics and does not, therefore, amount to a basically novel discovery, but a caveat may seem in place all the same.

Thus, it has been suggested that e.g. the /r/ is articulated with greater care and precision in the Czech language as an apical trill, because there it is held in place, as it were, by two one-dimensional oppositions: $\frac{\check{r}}{r - l}$, opposing it to the fricative /ř/ and to the lateral /l/ within the system; on the other hand, the German (or English, for that matter) /r/ is said to have a weaker position in the network of relations, being largely characterized negatively - as a non-lateral liquid, hence a non-nasal resonant, and therefore not an occlusive (1). Yet there are Slavic languages in which the /r/ is in no better a systemic relationship than in German, while in Dutch e.g. the /r/ is regularly pronounced in final position as well as preceding a consonant either in the same or at the onset of the next syllable. Admittedly there are two kinds of /r/ in German taken as a whole, the tongue-tip trill and the uvular variety, but neither of them is slurred in intial position in the word or syllable. On the other hand, in Common Slavic as reflected in Old Church Slavic, the final /r/ did drop out e.g. in the word for the "mother", mati, and this although it occurred in all other cases, Gen.Sg. matere etc. Not only that, but in the place (not only, of course) of the Czech /ř/, Slavic had a palatal /ŕ/ which should have helped to keep the /r/ in position everywhere, as a member in a paradigmatic network. Indo-European certainly had the final -r (2). The same final -r was lost in other Slavic kinship terms like *bhrātēr or *dhughətēr, obviously because of its final position in the syllable. (If it still is there in the remodeled form of Czech bratr, one cannot help thinking that, ironically, what kept is there alone among all Slavic languages including Slovak was precisely the symbiosis in which the Czech and German languages lived in Bohemia.)

It would, of course, be perfectly true to say that even if the Germanic /r/ is still always there in syllable-initial position, its incidence as a clearly articulated trill has nevertheless been seriously impaired. Yet we surely cannot on the one hand blame the statistical recession of the /r/ in the inventory of some Germanic languages on its allegedly isolated place in the system and on the other proclaim the emergence of /ž/ in the phonemic system of English as well as marginally in German loanwords from Romance as being due to an empty slot for it in the system despite its low frequency in the text. Be it not denied that the English /ž/ might not have come into existence without the drag-chain (3) of its pre-existing voiceless counterpart /š/, but it exists, after all, only in a few words such as vision, leisure, azure etc. The incidence of a phoneme should, accordingly, not amount to a major criterion in the establishment of a phonological system, any more than that of a grammatical category in the morphological system. A certain tense may be actually quite rare (e.g. in Bulgarian), but nevertheless occupy an important place in the system and endure for many centuries.

We have seen that the weakening or even loss of the /r/ in the two I.-E. language groups discussed seems to be due ultimately to the same cause, i.e. the position in the syllable, and cannot be generalized at all as proceeding from the paradigmatic place in the system. If isolation within the system were a valid criterion, the English /h/ would have been subject to a much wider loss than merely in some Cockney and other dialects. But an /h/ even occasionally comes into existence at the expense of another phoneme much better integrated with the others, in particular /s/; this is what happened in ancient Greek in initial and intervocalic position, it has arisen in some Slavic languages in the place of a well-connected /g/ and we can see it spreading before our eyes in a widely prevalent variety of Latin American Spanish, here again only in certain syllabic positions; in Spanish itself. /h/ arose out of /f/ preserved in Judeo-Spanish (Ladino). Alas, the system does not seem to be working consistently in the direction of its closer integration; these features are not entirely absent, but we must never forget that, as de Saussure has pointed out (4), the phonemes are really abstracted from their concrete position in the syllable, and cover an explosive and an implosive species. Only these actually occur in the chain of speech.

If in the Slavic kinship terms referred to above the final /r/ disappeared, then it shared this fate with all other implosive consonants, and the result was a rising wave of sonority, not followed by any coda. The syllables thus created may not correspond to Stetson's chest pulses (5) effectively criticized by Ladefoged (6), but they certainly constituted the best syllabic division, and division is the raison d'être of the syllable much more than any intrinsic nature of its own, hence the difficulty phoneticians have experienced in defining it. With some phenomena, their delimitative function is more important than their substance (if any), of which perhaps the most telling example is the present tense, which, looked at more closely, fades into nothingness except precisely as a dividing line between past and future. Hence also its flexibility (not as a "non-past"!)

We can, therefore, unfortunately, not agree with Martinet that the opening of the Slavic syllable indicates some mysterious "affaiblissement général des articulations implosives" (7), because the reason for such a negative development seems entirely unclear. Rather, it was the positive effect of an effort to mark off the syllables maximally from each other, as is the case in a sequence V/C. If, on the other hand, the same author says, "la syllabe est le segment du discours où l'unité d'intensité trouve le plus naturellement sa place", then it would seem to follow that they were fairly even in intensity and rather dominated by a musical intonation. In Sievers' distinction, they would be Schall- rather than Drucksilben, and with this it is not only the loss of ALL syllable-closing elements (including the second part of diphthongs) which is in agreement, but likewise the treatment of the syllabic nucleus - the vowels themselves. Their treatment was strictly in accordance with their inherent sonority; the closest, /i/ and /u/, became further reduced to /ь/ and /ъ/, being able to keep their timbre only under length, which favored greater sonority; /e/ and /o/ kept their place, though not without some vicissitudes, being of the middle degree of sonority as well as of length, while their long degrees /ē/ and /ō/ increased their aperture to /ǟ/ and /a/ respectively, the latter in agreement with the original /ā/. Short /ă/ and /ŏ/ fell together, i.e. sonority and length went hand in hand. It is as part of the same principle of unimpeded sonority that all I.-E. diphthongs were homogenized, thereby entailing further changes in the phonological system including the consonants, which therefore can be seen to be ultimately due to syntagmatic and not paradigmatic features. The syllable is the natural syntagma of the phoneme. Within it, all major sound changes of Common Slavic that give it such a different appearance from closely related Baltic were contained, while the one or two exceeding the limits of the syllable, like the Third (Baudouin de Courtenay's) Palatalization effected the breach precisely at the point of least resistance involving the least sonorous /i/ and /u/. It was also at these weak points that the syllabic structure of Common Slavic eventually broke down.

The maximal assimilation which prevailed in Common Slavic in the sequence CV (tautosyllabic)(8) is the reverse of the principle of the open syllable; hence the recurring palatalizations of the velars with their typically shifting locus (hub); the combinations of consonant + yod establishing new phonemes, the velarization of the /l/ etc. From the very opposition of the sequences CV and VC there evolved in Slavic their most consistent consequences in a truly dialectic harmony. The Common Slavic syllable was maximally homogeneous within, maximally delimited without, and only against this background do the individual changes make any sense.

References

1) N.S. Trubetzkoy, Grundzüge der Phonologie, Travaux du cercle linguistique de Prague VII (1939), p. 14.

2) *mātēr, cf. O. Szemerényi, Einführung in die vergleichende Sprachwissenschaft [2], Darmstadt 1980, p. 109; T.V. Gamkrelidze - V.V. Ivanov, Indoevropejskij jazyk i indoevropejcy I, Tbilisi 1984, p. 184.

3) André Martinet, Économie des changements phonétiques, Berne 1955, p. 59; on the double pressure on each phoneme from context and system cf.p.25.

4) F. de Saussure, Cours de linguistique générale, Paris 1922, P.79 ff.

5) R. H. Stetson, Motor Phonetics [2], Amsterdam 1951, passim.

6) Three Areas of Experimental Phonetics, Oxford University Press 1975, p. 23.

7) op. cit., p. 345; for the following cf. pp. 129, 252.

8) cf. L.V. Bondarko, Struktura sloga i xarakteristike fonem, Voprosy Jazykoznanija 1967, 1, p. 33-46.

RELATION BETWEEN SEGMENTAL PHONEMES AND TONES IN DIACHRONY

VYACHESLAV VS. IVANOV

Institute of Slavonic and Balkan Studies
Academy of Sciences of the USSR, 125040

ABSTRACT
The interconnection between the events usually described as supersegmental (such as phonemic tones) and segmental units is studied mainly on the example of phonological systems with laryngealization and pharyngealization used as supersegmental features in a syllable or a word. Synchronic and diachronic typology of prosodic systems with laryngealized and pharyngealized tonemes are discussed in connection with those phonetical data that give experimental support for the reconstructed historical evolution.

The problem of the relation between segmental phonemes and tones in diachrony is of utmost importance for the correct theoretic distinction between those aspects of speech sounds that are grasped by means of phonetic equipment and the purely functional use of the same sounds.

One should stress the importance of the problem for the general phonology since many other facts too point to the interconnection between the events usually described as supersegmental (particularly phonemic tones) and those phonetic manifestations that are considered mostly as segmental ones, cf. /I/, /2/.

In the recent studies on the tone phonology two results have been deduced that may be connected with each other. First in a lot of languages the segment inventory of the syllable phonemes can influence in unforseen degree the supersegmental characteristics of the tone. From the point of view of natural phonology stressing (as the academician Shcherba's conception has done) the importance of the phonetic substance one might speak about the phonologization in separate languages of those regular relations that on the phonetic level may be discovered between the feature voicedness-unvoicedness of stops and the feature high-low in the adjoining vowel /3/. From many diachronic consequences of this universal one might give only one example: Verner's law may be interpreted as the continuation of the old differences between the high and the low tones (and between the unvoiced and voiced cosonants in Proto-Germanic)in the given forms /4/ cf. the typologically similar opposition between /pólù/ 'back' and /bólù/ 'his back' in Kpelle (the group Mande, Liberia), where according to Welmer in the second form the low tone is reconstructed that could cause voicing. In the synchrony the causal link between supersegmental units and the complex character of prosodemes are found: in them the features are phonologized that can be described as phonologically segmental.

Secondly it is found that phonological oppositions of supersegmental units - tonemes (or prosodemes) in a number of languages are formed by differences not only in pitch and melody but also by some other features: quantity (for example in some Chinese dialects such as Shang-Khai and Amui, in Burmese, modern Yenissey languages), intensity, laryngealization and pharyngealization etc. These features are intertwined with those of pitch and melody. That makes the notion of tone in such a classical tone language as Chinese much more complicated than it had been supposed earlier.

From the point of view of general phonetics most understandible is the functional similarity between pitch and melody differences and the glottal stop since the glottalization is localized in the same part of the vocal mechanism where the different types of phonation are produced. For the concrete understanding of these phenomena important are the results of the investigation of the relation between the low tone and the voicedness of the consonant which is connected with the non-tenseness of the vocal chords. The suggestion according to which the pitch is becoming higher due to the growth of the tenseness of the vocal chords (and of the subglottal pressure) can be verified experimentally. In this connection one should study the possible link between the glottal stop and the rising tone that had been studied first in "Indo-Pacific" language Kate. In the glottal stop takes part in tone oppositions. The causal link between rising melody and the glottal stop was proved long ago by studies in experimental phonetics. The same results are achieved in the synchronic phonological description of languages where this relation is given phonological status to. In Kachari (the Bodo subgroup of the Tibeto-Burmese, Assam) the glottal stop and the rising tone are in complementary distribution. In Mixteo dialect of Santo Thomas Oxtepec the tone is higher in the verbs with the second syllable beginning in a glottal stop; in the Ayutle dialect of the same language a similar rising of the tone takes place in a syllable after which the glottal stop follows. Particularly interesting are the data of the Northern Thai dialect of Tang-pa. In it the tones of the high series (the first one and the fifth one) and the glottal stop are united and are tonemically close to the initial preglottalized consonants ('b, 'd, 'j). One can think that the glottal stop and the voiced preglottalized consonants are always similar as to their influence on the tone: usually they do not lower it, but they can rise it. In Ngizim (Chad subgroup of Afroasiatic, Nigeria) the implosive (injective) b /6/that is pronounced with ingressive in-taking of the air with the closed vocal chords, does not cause the lowering of the tone in distinction to other voiced cosonants. This can be related to the data of experimental phonetics according to which implosive (preglottalized) consonants might rise the tone more definitely than all other types of consonants. In this respect they are different from the other consonants which may usually lower the tones. From that point of view the facts of such a Modern Indo-Aryan language as Gujarati are important in which the implosive character of the articulation of the voiced consonants may be linked to the relation rising of the pitch during the closure and the growth of the number of vibrations. It can be supposed that just such articulation may lead to the rising of tone found in some modern Indo-Aryan languages (particularly in the Eastern Bengalian dialect of Dacca) in connection with the implosive character of consonants (in Sindhi) or the development of tones.

The universal that leads from the existence of the glottal stop to the rising of the tone may have some diachronic implications. Among them most interesting is the development of the high rising tone in Lahu (the Lolo division of the Lolo-Burmese subgroup of Tibeto-Burmese). The Proto-Lolo *-p, *-t, *-k of Proto-Lolo-Burmese had merged in a glottal stop due to the law that the end stops are worn out and lose the articulation in the mouth that can be given parallels to from many languages, Tibeto-Chinese aming them. Since in Lolo (as also in Proto-Lolo-Burmese) a number of words have initial glottal stop /ʔ/ in Lahu the dissimilation of the old initial glottal stop and of the new final one developed by merging of neutralized final *-p, *-t, *-k took place. This dissimilation might be to a principle according to which two glottalized stops do not occur in the same morphs as it can be seen in Kartwelian and some Amerindian languages /5/. The final -t had developed into the glottal stop in some Tibeto-Chinese languages in which later laryngealized (glottalized) tones such as Chinese Zhu-Sheng and Lolo stopped tones appeared. Later in Lahu in morphemes of this type the dissimilative disappearence of the glottal stop might be seen as means uniting segmental consonants structures and supersegmental tone systems. The Vietnamese 5-th tone (săc, rising) corresponds to the final /ʔ/ in other Austro-Asiatic languages of the group Palaung-Wa. One may think that here the phonetic events that were caused by the glottal stop have become phonologized: after the disappearence of the glottal stop the rising melody has become phonologically independent. According to the Vietnamese diachronical model similar process might be reconstructed also for the Old Chinese. It is supposed that in it the rising tone might be traced up to the glottal stop. According to the theory about the Indo-European laryngeals developing into Balto-Slavic syllabic intonations the rising intonation has arisen from a lost laryngeal (a glottal stop consonant after some interpretations). In a similar way the disappearance of a glottal stop might lead to a rising tone in Triqué according to its comparison to other Mixteo dialects.

The inverted relation between the glottal stop and the rising melody developing into a glottal stop may be supposed in later periods of the history of the same Baltic languages: Latvian, where the interrupted intonation (lauzta) has developed from the ancient rising (acute) one in mobile acute paradigms with the movement of the accent toward the old syllable before the accentuated one. A similar process of developement of the interrupted intonation from an acute rising one is found in North-Western-Lithuanian (Zhemaite). In Danish the stød (phonetically a glottal stop) is traced back to an old rising tone. But from the point of view of the Indo-European phonology it can have developed from an old glottalized cosonants not only in Danish but in English dialects too: */foʔt/ ← */phot'/ 'foot' etc.

With this type of synchronic and diachronic events telling about the functional line between the glottal stop and the rising intonation one may also connect the possibility of the use of the glottalization (as a part of a more complex articulation) as means of syntactical intonation in those sentences particularly interrogative that are usually marked by a rising melody in many languages6 Thus in Australian language Nyangumata glottalization is used for the most part in the end of the interrogative sentences with the rising intonation.

The connection between the glottalization and the rising melody as well as with the quantity (in Thai, Burmese, Chinese dialects, Yenissey an languages) might be counted among the universals that might find the general phonetic explanation in mechanism of the production of the source of the speech message. At the same time this universal is manifested in a number of languages either in synchronic events or in diachronic developement at different levels starting from the segmental phonemic up to the supersegmental (tonemic and intonational) one.

The use of laryngealization (or pharyngealization) as a supersegmental feature in a syllable or a word is functinally different from its application as syntactic intonational device which (expecially in the final position in a sentence) is found in several languages as in Mikasuki; but in both cases the parallelism with the use of pitch and melody contours is striking. One might distinguish the symmetrical prosodic systems with the equal or multiple relation of the laryngealized tonemes and the non-laryngealized ones and the asymmetric systems. In the latter one can find empirically that laryngealization is connected with the differential feature of the brevity of the vowel (Shan-Khai dialect of Chinese, Burmese, the Yenisseyan Ket dialects). Diachronically the asymmetrical systems might develop from those in which due to the neutralization the number of the tonemes in the syllable ending in a voiceless stop (developing later into the feature of laryngealization) is less than that of the tonemes in the other syllables (as in Middle Chinese and in some Tibeto-Burmese languages such as Atsi). For Twi, Livonian, sone Yenisseyan languages and for the earliest periodes of the history of Tibeto-Burmese and Chinese it appears possible to trace the line between the development of pharyngealized or laryngealized tonemes and the decrease in the differential potentialities of the phonemic inventory: just as in the history of nasalization (opposed to laryngealization in some languages such as Haimu in Melanesia) the lessening of the differential possibilities of phonemic components of words is compensated for by the expansion of supersegmental features. In different periodes of history of Tibetan, Burmese and some cognate languages laryngealized tonemes cyclically reappeared. Due to that the laryngealization as such persisted, but later it manifested itself in different morphs if compared with the older epochs. One might add that in many languages the glottal stop should be considered as phonetic realization of supersegmental (prosodic)feature of glottalization or laryngealization stretching on the whole syllable, the whole morph or the whole word. In the languages in which tonal differences exist glottalization and laryngealization (as well as also pharyngealization) are usually connected with the pitch-melody differences that build one system with them. Many examples of such prosodic systems have been analyzed /6, 7, 8, 9/.

REFERENCES

/1/ Maddieson I. Tone effects on consonants. - Journal of phonetics, 1978, vol.6, p.327-344.

/2/ Maddieson I. The effects of F_0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis.- Journal of phonetics, 1984, vol.12, N 1, p.9-15.

/3/W.Dempwolff Grammatik der Jabem-Sprache auf Neuguinea, Hamburg, 1939.

/4/ Ivanov V.V. Die historisch-vergleichende Indoeuropäische Sprachwissenschaft und die Typologie. - Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 1981, Bd.34, Heft 4, ss.415-417.

/5/ (Gamkrelidze T.V., Ivanov V.V. Indo-European and Indo-European vol.I, II, Tbilisi; Mecniereba, 1984.

/6/ Ivanov V.V. On the interrupted intonation (in Russian). - In: Rakstu krajums. J. Endzelins-Festschrift. Riga; Latvian Academy of sciences, 1959.

/7/ Ivanov V.V. On the origin of laryngealization-pharyngealization in Yenisseyan languages (in Russian). - In: Phonetics. Phonology. Grammar. A. Reformatsky - Festschrift. Moscow; Nauka editors, 1971.

/8/ Ivanov V.V. On synchronic and diachronic typology of prosodic systems with laryngealized or pharyngealized tonemes (in Russian). In: Essays in phonology of oriental languages. Moskow, Nauka, 1975.

/9/ Ivanov V.V. On the functions of the glottal stop (In Russian) - In: Sound pattern of language. Acad. Shcherba-Gedenkschrift. Moscow: Nauka, 1979.

PROSODIE ALS LINGUISTISCHE GESTALT

HERBERT PILCH

Englisches Seminar
Albert-Ludwigs-Universität
Freiburg i.Br.

1. PROSODISCHE MERKMALE

In ihrer grundlegenden Monographie zur Intonation führt Ceplitis /2/ eine Liste von acht "Intonationselementen" des Lettischen an. Sie sind ausgewählt einerseits nach ihrer "Semantizität" (d.h. als voneinander unabhängige, bedeutungstragende Elemente), andererseits nach den vier kennzeichnenden ("dominirujuščije") akustischen Eigenschaften:

1. Intensität kennzeichnet die Elemente Pause, Lautstärkeregelung (z.B. slur /2, p.85/) und logische Betonung (d.h. Hervorhebung eines Wortes im Syntagma /2, p.89/).

2. Frequenz kennzeichnet Melodie und Stimmlage.

3. Zeit kennzeichnet Sprechtempo und emphatische Länge.

4. Spektrum kennzeichnet Stimmqualität.

Vom allgemeinen (damaligen und heutigen) Diskussionsstand aus gesehen, geht Ceplitis einen weiten Schritt voran, insofern als sie nicht einfach akustische Eigenschaften mißt, sondern zuerst nach den linguistischen ("linguo-akustischen") Elementen fragt, die sie messen will. Mit Recht weist sie die verbreitete Vorstellung zurück, zunächst sei uns eine Reihe von segmentalen (durch ihre spektralen Merkmale definierten) "Lauten" gegeben, denen dann suprasegmentale "akustische Modifikationen" von Grundfrequenz, Intensität und Dauer beigegeben würden: " ... takogo rjada v real'noj reči net" /2, p.65/. Ihren Schritt voran geht Ceplitis einerseits gegenüber den messenden Experimentalphonetikern und den testenden Psychoakustikern. Bei diesen beiden Gruppen herrscht die Zuversicht, wenn man nur recht raffiniert messe bzw. recht breite Populationen teste, werde man am Ende alles Wissenswerte herausfinden.

Genauso läßt Ceplitis andererseits die konventionelle linguistische Typologie hinter sich, die Ton, Akzent und Intonation als die drei suprasegmentalen Kategorien postuliert und letztere auf die drei "prosodischen Merkmale" Tonhöhe, Lautstärke und Dauer reduzieren möchte /4, p.22; 5, pp.77-90; 6, p.55/. Insbesondere analysiert sie für das Lettische fünf "melodische Motive" /2, pp.113-115/ und stellt sie den fünf (inzwischen sieben) "Intonationskonstruktionen" des Russischen (nach Bryzgunova /1, p.86/ gegenüber. Damit tut sie den entscheidenden Schritt von der Intonation als Menge fallender und steigender Kurvenstücke zur Intonation als phonologischem (d.h. einzelsprachlichem) System.

2. PROSODISCHE GESTALTEN

Wir gehen in der gleichen Richtung noch einen Schritt weiter. Erstens lösen wir die gesamte linguistische Kategorie Prosodie aus ihrer traditionellen Bindung an bestimmte akustische Merkmale und begreifen Prosodien als Ganzheiten ("Gestalten"). Ihnen können wohl akustische Merkmale zugeordnet werden, sie lassen sich aber nicht auf bestimmte akustische Merkmale reduzieren. Auszugehen brauchen wir deshalb nicht mehr von jener schier unübersehbaren Menge akustischer (und sonstiger) Eigenschaften, wie sie immer wieder in der Literatur aufgeführt werden, als da sind Grundfrequenz, Intensität, Tempo, Rhythmus, Pausensetzung, Stimmlage, Stimmqualität, Gesamtintensität, Steigungswinkel usw. /2, p.62/. Im Gegenteil - Prosodie als einzelsprachliches System definiert sich nicht akustisch, sondern durch die Unterscheidbarkeit der zum System gehörigen Einheiten ("Prosodien"): "Dans la langue il n'y a que des différences" /18, p.166/.

Zweitens präzisieren wir die allgemeine Kategorie Semantizität einerseits als zeichenunterscheidende Denotation (Distinktivität), andererseits als Textzusammenhang stiftende Diskursivität (kommunikative Funktion /14/). Es ist eben nicht dasselbe, ob ich durch bestimmte Prosodien zwei schwedische Wörter wie anden[1] 'Geist' und anden[2] 'Ente' unterscheide. Das ist ein Wortakzent mit distinktiver Funktion. Oder ob ich durch andere Prosodien eine Antithese andeute (z.B. bei der Aussprache beider Wörter als minimales Paar). Das sind Intonationen mit diskursiver Funktion. Letztere berührt die Identität der beiden Zeichen nicht; denn eben mit ihrer diskursiven Funktion grenzt sich die Intonation kategoriell ab vom distinktiven Ton /15/.

Kleine prosodische Einheiten, die wir als Ganzheiten begreifen, nennen wir Figuren, und zwar im Anschluß an Mulder /6, p.43/ und Hjelmslev /3, p.43/. Im Gegensatz zum Zeichen mangelt es der Figur an denotatio, nicht aber an Diskursivität. Im Gegensatz zum Phonem (bzw. distinktiven Merkmal) braucht die Figur keine distinktive Funktion auszuüben (sie tut dies nur als Ton bzw. als distinktiver Akzent). Es genügt die diskursive Funktion. Im Gegensatz zum Intonem hat die Figur somit keine feste Bedeutung ("intonacionnyje jedenicy, každaja iz kotorych obladajet opredeljonnym značenijem" /2, p.160/).

Genau wie andere phonologische Einheiten bilden auch die Figuren Hierarchien. In einzelsprachlich sehr unterschiedlicher Weise verbinden sie sich zu größeren Einheiten wie Konturen /10/ bzw. Konstruktionen /1/. Umgekehrt lassen sie sich in kleinere Teilstücke segmentieren, z.B. die aus der griechischen Grammatik bekannten Teilstücke Proklise, Tonsegment, Enklise. Die Prosodien lassen sich ordnen zu phonologischen Korrelationen, aus denen sich ihrerseits distinktive Merkmale abstrahieren lassen, z.B. die drei gebrochenen gegenüber den drei kontinuierlichen Figuren des Kymrischen /13/. Außerdem lassen sie sich (wiederum in einzelsprachlich sehr unterschiedlicher Weise) den segmentalen Einheiten wie Silbe, phonologisches Wort, Diskurseinheit ("sentence" im Sinne Mulders /6, p.56/) zuordnen. Die dem Wort zugeordnete Prosodie nennt man allgemein Akzent, die der Diskurseinheit zugeordnete Intonation. Das ist häufig die gleiche Prosodie (z.B. das Tonsegment des Russischen, vgl. unten Abschnitt 4) je nachdem, welche Zuordnung wir im Einzelfall betrachten. Die Diskurseinheit braucht kein Syntagma zu sein, sondern es ist ein Textsegment wie These : Antithese in der bekannten rhetorischen Figur. Nur bei der Leseintonation fallen Diskurseinheit und Syntagma zusammen (daran erkennen wir alle den vorgelesenen Text).

Unsere prosodische Analyse bestimmter Sprachen (besonders des Englischen, Deutschen, Kymrischen und Bretonischen) haben wir an anderer Stelle vorgelegt /7//10//13//17/, ebenso unsere Typologie prosodischer Systeme /15/. Wir möchten hier von zwei anderen Erfahrungen berichten, einmal didaktisch von der Erlernbarkeit prosodischer Systeme, sodann von der Zuordnung russischer Prosodien zur Kategorie Akzent.

3. ERLERNBARKEIT

Aus der Auffassung der Prosodien als Ganzheiten folgt, daß sie nur als ganzes erlernbar sind. Das erweist sich sowohl an Kindern, die sie auf natürlichem Wege mit ihrer Muttersprache(n) lernen, als auch im Sprachunterricht. Niemand lernt z.B. getrennt erst Tonhöhenbewegung, dann Lautstärkeregelung, dann Rhythmus, Sprechtempo u.dgl. Kinder lernen das alles im ganzen. Sie spielen mit prosodischen Gestalten, ordnen ihnen zunächst segmentales Kauderwelsch zu, erst allmählich Wörter und Diskurseinheiten. Die Wörter und Syntagmen, die sie im Laufe der Zeit erlernen, ordnen sie den bereits erlernten Prosodien zu. Deshalb klingen sie stets "muttersprachlich".

Der Schulunterricht geht dagegen umgekehrt vor, beginnt mit Wörtern, Sätzen und (segmentalen) Phonemen und läßt die Prosodien bis zum Schluß. Im allgemeinen werden sie dann nicht mehr gelernt, weil die Prosodien der Muttersprache vom Schüler bereits fest in die Fremdsprache übernommen sind. Meinen (deutschsprachigen) Studenten, die Englisch lernen wollen, rede ich deshalb zu, zunächst den amerikanischen Touristen zu parodieren, wie er Deutsch "mit englischem Akzent" spricht. Das lernen sie leichter als richtiges Englisch, weil es Spaß macht. Als nächstes parodieren sie den amerikanischen Touristen mit dem gleichen "Akzent", aber mit englischen (statt deutschen) Wörtern. Damit lösen sich ihre Ausspracheprobleme sowohl im segmentalen als auch im suprasegmentalen Bereich, und wir können auf phonetischen Kleinkram verzichten /9/.

Wer eine Tonsprache lernen will (z.B. Chinesisch), bekommt im allgemeinen gute Ratschläge, wie er auf das Steigen bzw. Fallen der Melodie achten müsse. Erfahrungsgemäß nützen solche Ratschläge nichts. Warum nicht? Weil es um ganze Figuren geht, nicht um steigende oder fallende Melodiestücke. Tatsächlich bewegen sich die Figuren nie stetig aufwärts oder abwärts (bzw. eben), sondern immer sehr unregelmäßig. Sonst klänge die Figur nämlich nicht wie gesprochen sondern wie ein glissando. Wenn wir von den "steigenden" oder "fallenden" Tönen einer Sprache sprechen, so meinen wir damit die tatsächlich hörbare (und meßbare) Bewegung nur im hoch stilisierten Sinne des distinktiven Merkmals, und dieses hören wir nicht von Natur aus, sondern müssen es erst "hören" lernen.

Das ist der Übergang vom angeborenen, audiometrischen Hören zum erlernten, phonematischen Hören /8/. Deshalb müssen wir auch die steigenden und fallenden Figuren für jede Sprache eigens neu lernen. Wenn ich z.B. den fallenden Ton des Mandarin kann, kann ich deshalb noch nicht den fallenden

Ton des Thai, geschweige denn das fallende Tonsegment der ersten Konstruktion des Russischen. Das distinktive Merkmal "fallend" stilisiert für jede der drei Sprachen eine prosodische Gestalt _sui generis_ - typologisch vergleichbar, aber nicht akustisch gleich.

4. RUSSISCHER WORTAKZENT

Hören wir flüssiger russischer Rede zu in der Absicht festzustellen, welche Silben betont sind! Wenn wir uns dabei von der Lehre vom "Intensitätsakzent" des Russischen leiten lassen, so erkennen wir die Akzente nicht (es sei denn, wir wissen schon im voraus, wo sie liegen). Dagegen höre ich den russischen Akzent leicht, wenn ich auf die Vokalqualität achte, auf den vollen, betonten Vokal. Auf manchen (aber nicht auf allen) betonten Vokalen höre ich darüber hinaus das Tonsegment einer Intonationskonstruktion (d.h. den "udarnaja čast" der Konstruktion /1, pp.17, 23 passim/). Nur wenn einzelne (phonologische) Wörter vorgesprochen werden, so trägt der betonte Vokal stets ein Tonsegment, der "nebentonige" Vokal der unmittelbar vorangehenden Wortsilbe dagegen nur bei besonders "deutlicher" Aussprache (es geht hier um das erlernte, phonemətische Hören!). Auf Grund der unterschiedlichen Wortgrenzen kann ich deshalb (für die etwas minder deutliche Aussprache) minimale Paare erfinden wie (bei der ersten Intonationskonstruktion): _da^1_, _bu^{1}det_ ≠ _dobu^{1}det_. In gleicher Weise hörbar sind die "Nebenakzente" in Komposita kirchenslavischen Typs wie _bògomáter'_, _mnògouvazájemyj_. Auch für sie kann ich minimale Paare erfinden wie _bo^{1}ga ma^{1}ter_ ≠ _bogoma^{1}ter'_. Ich brauche dazu nur meinem "Hauptakzent" ein Tonsegment zuzuordnen, meinem "Nebenakzent" nicht - auch wenn die genannten Wörter keineswegs immer so gesprochen werden.

Der russische Akzent ist also weder ein "Druckakzent" noch ein "melodischer Akzent" (im Sinne der klassischen Lehre), sondern "betont" ist per definitionem jeder volle Vokal (und kein anderer Vokal). Der betonte Vokal (aber kein anderer Vokal) kann außerdem das Tonsegment einer Intonationskonstruktion tragen. Wir halten das Russische für die Akzentsprache _par excellence_, weil jedes russische Wort genau eine Tonstelle hat (wir sprechen lieber von Ton_stelle_ als von Ton_silbe_, weil die unmittelbar vorangehende, "nebentonige" Silbe mit zur Tonstelle gehört). Anders als im Englischen und Deutschen gibt es (mit Ausnahme des genannten Kompositionstyps) keine Nebenakzente im (isolierten) Wort. Anders als im Schwedischen und Norwegischen kommt es nicht auf eine bestimmte (distinktive) Figur an. Anders als im Französischen und Kymrischen (und teilweise im Englischen) liegt die Akzentstelle lexikalisch fest, d.h. die jeweilige (phonologische) Wortform (flektierte Form, präpositionale Gruppe) hat jedes Mal, wenn sie in der Rede auftritt, den Akzent an der gleichen Stelle. Das kommt uns vielleicht selbstverständlich vor, ist es aber in anderen Sprachen nicht (wenn wir nur flüssiger Rede zuhören statt isolierten Wörtern /12//16/).

Gewiß ordnen wir auch dem russischen Akzent Meßwerte für Intensität, Dauer, Grundfrequenz u.dgl. zu - genau wie allen segmentalen Einheiten auch /2, p.65/. Die volle Vokalqualität bringt sogar im allgemeinen höhere Meßwerte mit sich als die reduzierte. Nur macht das solche Meßwerte nicht zum _definiens_ des Akzents. Im Gegenteil: "Odnako, naskol'ko nam izvestno, metodiki, pozvoljajuščej po odnim liš' fizičeskim parametram opredeljat', javljajetsja li dannyj segment udarnym ili bezudarnym, poka ješčo net" /2, p.90/. Eine solche Methode wird es auch nie geben. Der Grund dafür ist einfach: Prosodien definieren sich als phonologische Ganzheiten, nicht als reduzierbar auf bestimmte akustische Eigenschaften.

Literaturverzeichnis

/1/ Bryzgunova, E.A., Zvuki i intonacija russkoj reči, Moskau 1969.

/2/ Ceplitis, L.K., Analiz rečevoj intonaciji, Riga 1974.

/3/ Hjelmslev, L., Omkring sprogteoriens grundlæggelse, Kopenhagen 1943.

/4/ Jakobson, R. und Halle, M., Fundamentals of Language, Janua linguarum, series minor 1, den Haag 1956.

/5/ Martinet, A., Eléments de linguistique générale, Paris 1960.

/6/ Mulder, J. und Hervey, S., The Strategy of Linguistics, Edinburgh 1980.

/7/ Pilch, H., Baseldeutsche Phonologie auf Grundlage der Intonation, Phonetica 34 (1977), 165-190.

/8/ ---, Auditory Phonetics, Word 29 (1978), 148-160.

/9/ ---, Intonation als Grundlage der Ausspracheschulung, Festschrift Wächtler, ed. P. Kunsmann und O. Kuhn, Berlin 1980, 221-232.

/10/ ---, English Intonation as Phonological Structure, Word 31 (1980), 55-66.

/11/ Pilch, H., Der phonologische Bau des russichen Wortes, Gedenkschrift Hildegard Schroeder, ed. H. Riggenbach und F. Keller, Bern 1981, 497-511.

/12/ ---, Les mots anglais à accent mobile, Festschrift George Faure, ed. P. Léon und M. Rossi, Brüssel 1981, 3-8.

/13/ ---, The Structure of Welsh Tonality, Studia celtica 18/19 (1983-4), 234-252.

/14/ ---, Intonation als kommunikative Funktion, Festschrift F. Link, ed. U. Halfmann et al., Berlin 1984, 195-210.

/15/ ---, La tonalité linguistique, La linguistique 20 (1984), 29-48.

/16/ ---, Tonal Cues in Welsh Discourse, Studia celtica (erscheint).

/17/ ---, L'accentuation de la langue bretonne, La Bretagne linguistique 2 (Brest 1987), 1-26.

/18/ de Saussure, F., Cours de linguistique générale, Paris 1949.

SEGMENTAL AND PROSODIC VARIABILITIES IN CONNECTED SPEECH. AN APPLIED DATA-BANK STUDY

GUNNAR FANT, LENNART NORD, ANITA KRUCKENBERG

Dept. of Speech Communication and Music Acoustics
Royal Institute of Technology (KTH), Box 70014
S-100 44 Stockholm, Sweden

ABSTRACT

As a subset of the KTH data bank, we have recorded several subjects reading the same passages from a selection of various texts. We have studied variations in the realization of segmental and prosodic characteristics and to a less extent reading style. Data is reported on the degree of closure of voiced consonants, ambiguities in segmentation and vowel durations. Vowel-consonant contrasts may be highly reduced even in non-weak stress forms. The multi-cued realization of syntactic boundaries are discussed in relation to subjective assessments and to rhythmical structures. In addition to physical pauses, final lenghtening, formant-pattern changes and intonation contours, there are also local voice source features other than F0 to consider, e.g., creaky voice junctures.

INTRODUCTION

Advanced work on text-to-speech synthesis and speech recognition demands a continuous updating, extension and renewal of knowledge from speech analysis. We have to adopt a rule-oriented search to efficiently encode phonetic features, speaker typology and behavior. The data-bank storage and processing system of Carlson and Granström /1/ have provided a format and practical tool. A more complete account of the work is given in a report by Fant et al. /2/ which contains observations of speaker behavior under various conditions, not included here.

Within this limited frame, we have gained a fresh insight in several fundamental acoustic-phonetic problems and a view of what kind of problems we will encounter as the analysis proceeds. We have studied segmentation problems and underlying variabilities in articulatory gestures and, furthermore, the realization of syntactic boundaries, and how subjective juncture assessments correlate with acoustic factors. Our overall impressions is that of a richness of variability on all levels as well as potentials of structuring variabilities. One prevailing impression is that segmentals and prosodics share a common basis of acoustic correlates. Therefore; they should be treated together as seen from an underlying model of speech production. Our study has also provided some limited data on vowel durations and prosodic realizations which can be extended to support the up-dating of our synthesis rules.

When instructing subjects, we laid an emphasis on attaining a neutral but semantically distinct reading. In addition, we have also recordings of more engaged readings, occasional mannerisms and deliberately dramatized versions. However, even in the more normal readings, we observed a rather large span of intonation and overall prosodic patterns. Deviations from average and preferred patterns attain a subjective personality marking which attracts our attention without affecting the overall quality of the reading. It would be of interest to certify which prosodic factors remain intact and which are allowed to vary.

In the present pilot study we have concentrated on 14 subjects' readings of two sentences. Spectrograms and associated oscillograms, intensity and F0 plots were produced by means of our laboratory computer processing routines.

SEGMENTAL STUDIES

We have studied various coarticulation and reduction phenomena that affect the segmental composition of phonemes and complicate the task of boundary assignments.

Boundaries are more clearly realized by changes in "manner" cues than in "place" cues. Thus, it is easy to find the boundary between a fricative and a vowel but we have no clear rules for finding boundaries between vowels or between voiced consonants like /v/, /j/ and /r/ and their combinations with vowels. A voiced intervocalic stop is not always associated with a stop gap, and phonemically unvoiced stops in unstressed positions may attain voicing. Lack of oral closure may affect nasals as well as stop sounds or any consonant, and an incomplete abduction of the glottis in an /h/ causes a continuation of voicing.

In order to understand these ambiguities, we should consider a basic parameter of speech production related to the extent to which vocal-tract constriction targets are reached in connected speech. This parameter which has a strong descriptive power could be labeled "articulatory contrast" or more generally, "dynamic contrast". It affects not only the supraglottal articulators but also the glottal articulation. Thus, a sufficient adduction/abduction contrast is needed for preserving a voiced/voiceless boundary. Also, the boundary between a voiced /h/ and a following vowel becomes obscured by insufficient glottal contrast.

Articulatory contrast implies acoustic contrast in terms of envelope intensity modulation as well as an extended range of formant pattern dynamics. Decreased contrast, thereby, also affects the rate of change of formant patterns at segment boundaries.

Although these phenomena are by no means new in phonetic theory, we had not anticipated the full extent of their realizations. Thus, most speakers did not produce a full closure of the voiced stop /g/ in "legat". For one speaker, LN, the intensity modulation was marginal only and the formant pattern that of a connecting glide, see Fig. 1.

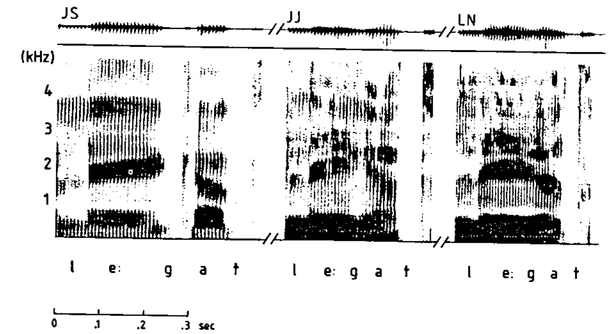


Fig. 1. Three degrees of articulatory contrast. The same word "legat" from three subjects' readings.

This is typical of voiced stops in fluent rapid speech and probably dependent on both the place of articulation and the vocalic context.

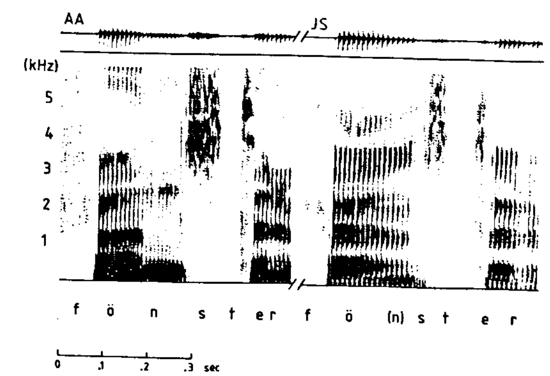


Fig. 2. Two subjects contrasting in oral closure of consonant [n].

Another example of incomplete closure is in the nasal consonant /n/ in the word "fönster" which often is realized by nasalization of the vowel only, see Fig. 2. The appearance of an orally closed segment for the /n/ of "i en" was even less frequent which is to be expected since "en" is a function word. On other occasions we have noticed this effect to be particularly strong in phoneme strings consisting of vowel-nasal-fricative. For American English, Malécot described this phenomenon in word pairs with nasal-homorganic unvoiced stop , like "camp" versus "cap", differing in nasalization only /3/.

The two-word string "han hade" in the initial part of the sentence "Han hade legat och skrivit det i en stor sal vars fönster vette mot Klarälven" is produced out of focus and with higher tempo and reduced articulatory contrast. The /n/ is realized by nasalization only, and the second /h/ is hard to detect being glottally coarticulated with the following vowel /a/. The second /h/, when present, can thus be said to be realized by aspiration of the following vowel. A further complication is that nasalization and aspiration share cues, e.g., the reduced F1 intensity. Only few speakers produced a sequence of clearly identifiable segments.

The Swedish /r/-sound appears in a variety of acoustic realizations ranging from a pronounced trill to a slight /r/ coloring of a neighboring vowel. /r/ also occurs frequently in consonant clusters with subsequent forms of coarticulation and reduction as a result /4/.

In "fönster vette" and "skrivit", the /r/-sounds are often reduced and segmentation becomes a problem. The acoustic cues become diffuse, a brief constriction phase is often found, but not always, and the same is true of the F1, F3 and F4 lowering cues. When present, the constriction phase of /r/ in "skrivit" may mark the right boundary of an inserted vowel after /k/. It may also reside in the unvoiced k-release. Segmentation rules for /r/-sounds are still undefined. Shall we concentrate on the stop gap if it is present or should we choose a larger domain of perceptual importance including a possible inserted vowel or a short segment of the same nature?

With the latter choice, the segmentation principle will deviate from that of handling stop sounds where, by convention, the voiced part of a following transition goes with the next segment. When the acoustic cues become weak, the auditive impression of the /r/ prevails though weakened.

More examples of variabilities of segmental realizations and segmentation ambiguities will be discussed in connection with the study of syntactic boundary regions in the following section.

SYNTACTIC BOUNDARIES AND PROSODICS

Our standard sentence with each word assigned a lexical stress pattern according to SAOB* attains the following structure:

```
 4    3  2  3 2   4        3  2    4   4  4     4    4
HAN  HADE  LEGAT  OCH  SKRIVIT  DET  I  EN  STOR  SAL
 4     4    0   3  2   4     3 2  0
VARS  FÖNSTER  VETTE  MOT  KLARÄLVEN
```

This transcription of each word read in isolation is irrelevant to connected speech. Following established notations we transform it into a more realistic form omitting the stress of function words except the pronoun "det" which generally attains the prominence of its substitute.

*Swedish normative word dictionary.

HAN HADE LEGAT OCH SKRIVIT DET | I EN STOR SAL |

VARS FÖNSTER VETTE MOT KLARÄLVEN

"i en stor sal" is a preposition phrase. Crosses denote grave accent. A vertical short bar denotes acute accent, if above the line, and the secondary syllable of grave accent, if below the line.

One object of the study was to study the realization of the syntactic boundaries before and after the preposition phrase. We found a considerable variation in both acoustic cues and subjective impressions. In a listening test, ten subjects assessed the degree of perceived boundaries on a scale from 0 to 5. The first boundary got an average rating of 2.2 with a standard deviation of 0.8 whilst the second boundary was rated 3.7 with a standard deviation of 0.75 within the jury. The standard deviation between speakers was 1.1 and 2.2, respectively.

The most prominent acoustic cue appeared to be segmental durations. Since the three words "det i en" in several cases merged to a single voiced gross segment [eI] with no clear boundaries, especially not in the formant juncture between /e/ and /I/, we selected an interval from the onset of voicing in the /e/ of "det", to the onset of the /s/ of "stor", thus potentially including final and initial lengthening effects. It should be noted that three of the 14 speakers omitted the /t/ of "skrivit" and produced a voiced stop gap of 40-70 ms duration for the /d/ of "det". The remaining 11 speakers' spectrograms showed an unvoiced stop gap of 70-140 ms duration appropriate for the /t/ plus /d/ with an uncertainty of whether the /d/ was realized at all and, if so, with no obvious boundary towards /t/. According to the sandhi rules of Gårding, /t+d/ are transformed to unvoiced /d/ /5/.

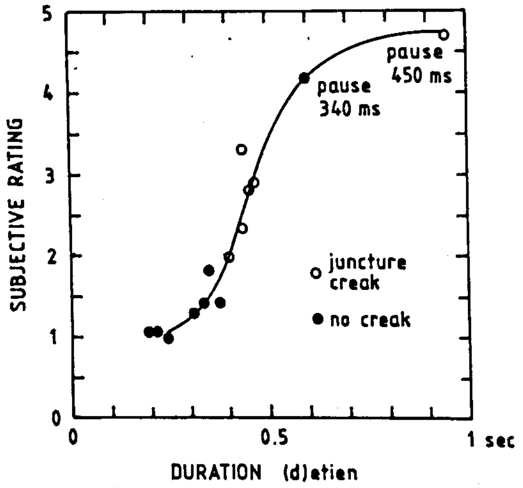


Fig. 3. Subjective rating versus durational measure of the first phrase boundary.

Fig. 3 shows a fair correlation between boundary region duration and the subjective boundary assessment. A tendency may be observed of a doubling of the subjective rating per 200 ms increase of the juncture duration. Deviations from this trend are within the standard deviation of listener judgements. A most apparent trend associated with more marked boundaries is the appearance of creaky voice, i.e., glottalization at the end of the /e/ which causes a local drop of F0 and/or an alternation between strong and weak glottal excitations which is especially apparent in the second and higher formants. These boundary cues have earlier been noted by Gårding /5/, Lehiste /6/ and Kreiman /7/. These alternations may cause an ambiguity in the definition of the local F0. The two speakers of the highest boundary rating produced a proper pause at the phrase boundary. A general phonological rule is to omit the /t/ of "det". Only one of the speakers, EJ, had a proper combination of /t/ unvoiced stop gap + release at the following vowel. Eleven speakers did not have a /t/ stop gap, and the two with proper pause after "det" omitted the /t/.

In absence of glottalization, most speakers produced a level or slightly rising F0 contour at the juncture. An exception was subject BB who had a falling F0 into the beginning of the second phrase. His reading style was in general more personal and affected than others.

As a durational measure, for the second phrase boundary "-----sal vars-----" we selected the /1/+occasional pause+/v/. There were seven subjective ratings between 3.0 and 3.6 with a duration of about 120 ms and six with a rating between 4.2 and 4.4 with associated durations from 200-540 ms. Of the later, four of these included a proper pause and two displayed a brief /1/-release. The F0-contour was mostly a fall+rise at the boundary with some correlation between magnitude of the movement and subjective rating. Here again, subject BB deviated from the rest by an F0 rise+fall. None of the subjects displayed a glottalization. Two subjects, however, had shown such tendencies in earlier informal recordings.

From a second sentence containing a sequence of enumerations, we found similar correlations between subjective boundary impressions and durational measures. Rather constant subjective ratings independent of durational measures were found when a boundary was terminated by strongly stressed syllables on either side.

Our data on vowel durations are summarized in Table I. They are compared with data from Carlson and Granström (ref. /1/) and from the text-to-speech (Rulsys) generated version of our sentence (in May 1986). A correction for overall tempo has been made, the Rulsys sentence being 20% longer. This comparison confirms our awareness of the insufficient contrast between present Rulsys-generated unstressed and stressed short and long vowels, see ref. /2/ for further details. This restricted study can only provide a tendency and more representative data will eventually be gathered.

Table I. Vowel durations in milliseconds

	Short unstressed	Short stressed	Long stressed
Present study	42	105	155
Rulsys	78	93	134
C & G (ref. /1/)	60	90	125

The average reading speed was five syllables or 14 phonemes per second. The standard deviation was rather low, 7% only.

We have looked into the rhythmical structure of the sentence. A rhythmical unit, "stress interval", has been defined as a subpart of the utterance located between the onsets of two successive vowels carrying main stress. Since function words are down graded, the main stresses are confined to content words. We find an overall tendency of two main stresses per second. Similar findings have been made by Goude and Malmström /8/ and Dauer /9/. Distances between vowel onsets in stressed syllables are, thus, of the order of 500 ms but vary with the number of phonemes typically from 350 ms for three phonemes to 600 ms for nine phonemes. The weak tendency of isochrony in reading, see also the study of Strangert, is probably more a matter of constraints in number of phonemes per stress interval than an intention of the reader /10/. As an average for the nine sentences of the central passage, we find two words or eight phonemes per stress interval. It remains to quantify the actual performance of the reading of this and more extensive tests. Even though Fig. 4 refers to a single sentence, it exemplifies typical trends such as the relative weight of stressed vowels and that stress intervals which cut across phrase boundaries are longer whereas the sentence initial group, leading up the first stress, is shorter than within phrase stress intervals /11/. Further studies along these lines might give some insight in reading behavior.

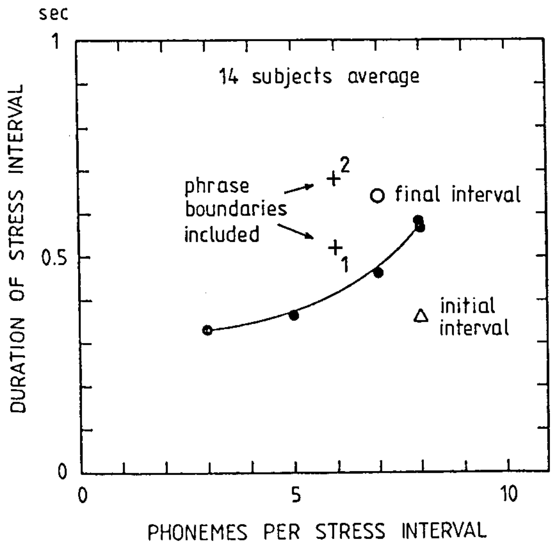


Fig. 4. Duration of stress intervals versus number of phonemes contained. Stress intervals that cut across a phrase boundary are lengthened.

REFERENCES

/1/ R. Carlson, B. Granström, "A search for durational rules in a real-speech data base", Phonetica 43, 140-154, 1986.
/2/ G. Fant, L. Nord, A. Kruckenberg, "Individual variations in text reading. A data-bank pilot study", STL-QPSR 4/1986 (KTH, Stockholm), 1-17.
/3/ A. Malécot, "Vowel nasality as a distinctive feature in American English", Language 36, 222-229, 1960.
/4/ L. Nord, "An acoustic and perceptual study of /r/ varieties in Swedish", forthcoming.
/5/ E. Gårding, "Internal juncture in Swedish, Gleerup, Lund, 1967.
/6/ I. Lehiste, "Perception of sentence and paragraph boundaries", in B. Lindblom, S. Öhman eds., Frontiers of Speech Communication Research, Academic Press, 191-201, 1979.
/7/ J. Kreiman, "Perception of sentence and paragraph boundaries in natural conversation", J. of Phonetics 10, 163-175, 1982.
/8/ G. Goude, S. Malmström, "Ett exempel på experimentalpsykologiskt studium av rytmupplevelse av poesi", University of Stockholm, 1970.
/9/ R.M. Dauer, "Stress-timing and syllable timing reanalyzed", J. of Phonetics 11, 51-62, 1983.
/10/ E. Strangert, "Swedish speech rhythm in a cross-language study", Almqvist & Wiksell Int., Stockholm, 1985.
/11/ I. Lehiste, "Isochrony reconsidered", J. of Phonetics 5, 253-263, 1977.

THE TYPOLOGY OF SENTENCE INTONATION SYSTEMS

TATJANA NIKOLAYEVA

Institute of Slavistics and Balkanistics;
Academy of Sciences of the U.S.S.R.
Moscow, USSR,125040

The study of the intonation system is efficient if the distinctions are made between universal,typological and specific phenomena.The report sets out some data on the prosodic units from this point of view and the ways of reconstruction intonation system of different languages.Some phonetic laws are reconsidered.The "law of language evolution" is put to discussion:increase of information per speech unit leads to the development of the suprasegmental sentense features(presuppositions and intonation contours including)since the human perception capacity is limited.

1.1. We consider the intonation system as a complex of three functional prosodic parameters: F_0 frequency, duration and intensity.

1.2. At the contemporary stage of intonological investigation it is reasonable to distinguish two complementary plans:
1/ synchronic intonational typology ;
2/ diachronic intonational typology.

2.1. In both cases it is necessary to differentiate such three strata: universal, typological and language specific one /1/. This distinction is not an easy one, and we have to elaborate a method for it. Let us turn our attention to one of prosodic phenomena: intensity curves in a word and in a sentence. We know that in a word an intensity curve is usually declined, so in Slavic languages the stressed syllable occupying the last place (or the penultimate syllable) can be a little lower than the unstressed initial one. For example, the Polish word zakładzie has such intensity curve (in mm): 15-13/12/8; the Ukrainian word tikáti has 9/4/4; the Russian word ubežál has 9-8/7-3/2-3 and so on.The data of other languages confirm this regularity /2/. And there seem to exist some premises to consider the intensity declination as a prosodic universal. And it is easy to find an explanation for it in communication and articulatory programmes. Yet in Turkish languages the situation is quite different /3/. The intensity has rather a rising than a falling contour in the word. Moreover, it turns out, that in Turkish environment the Slavic speech is influenced by this Turkish tendency (we have data of the Bulgarian language and some Russian Volga dialects). The intensity curve in them may be risen either to the stressed syllable or to the end of the word end (Bulg. papagál - 30/45/50; Russian Volga dialectal: tópota - 6/2/8 ; general Russian:tópota - 6/4/4 ;/4/).Thus the word intensity direction is not universal but a typological phenomenon.And now let's analyse the sentence intensity curve. All languages known (Turkish including) demonstrate a clear tendency to declinate intensity / 5/. In any case we don't know any other data. So, these two intensity curves - in a word and in a sentence- are factors of different character. The former is typological, the latter is universal.

2.2. The rise/fall opposition (in questions vs in declarative sentences) has been interpretated many times /6/. But in reality the solution is not quite so trivial. Namely: 1/ the falling melody in declaratives is universal; 2/ the question melody movement is of special interest. First of all, generally speaking, there is no question as such: we have some types of questions each with its own function and specific melody. And thus we have a certain correlation between two sets: the intonation construction set and the number of questions types. In the former set we should in the first place separate the wh-question which has a quasi-universal intonation type with a falling end/7/. It is universal too.

2.3. The second issue is an ontological one. What is namely the rising melody? For example, there are two dominant types of questions melody in Slavic languages: they split the Slavic group into two subgroups, the Western group and the South-Eastern one.

The first type is characterized by the rising tone in the final unstressed syllables (boundary tones) including the last unstressed syllable; the final stressed syllable (under sentence stress) can take a lower position.

The second type is characterised by a high steep rise on the stressed syllable, post-stressed syllables have a falling tendency, thus the whole contour has a low end. West Slavic languages use the first type in general (yes-no) questions and the second type - in alternative and repeated questions. South Slavic languages and, specially, East Slavic languages, on the contrary, use the second type in general questions and the first type in repeated questions /8/. Now if we are to accept O'hala and Bolinger's viewpoint literally, then we may have two answers: 1/ the first answer is: no, the second figure is not rising one, if rise means to be rising to the very end;2/ the second answer is: yes, it is, if we treat question- declarative melody opposition as the high-low opposition because the frequency interval in East-Slavic (Russian) yes-no question is very large and the stressed syllable is intoned very high.

And now we propose to replace the opposition:rising-falling (as universal question-declarative opposition) to the opposition:high-low (with rising-falling as part of it).

Thus the intonation types are heterofunctional and may be represented in the following way:
Melody curve in declaratives is Universal;
Melody curve in Wh-questions is Universal;
Intensity curve in declaratives is Universal;
Intensity curve in words is Typological;
Melody curve in not-Wh-questions is Typological.
(This list may be continued in fact).

The above described analysis proves that the description of language reality is not simple,it may be contradictory and not at all economic.

2.3. In our description the term "typological" means "non-universal". It shows that the intonation system of a language is to be correlated with a certain set of features which it can share with another language.This set is not always conditioned by genetic or areal reasons. For example,it is known that in African languages with lexical functions of tones the differentiation of questions and declaratives is realized by pitch registers, and not by contours /9/. R.F.Paufoschima has found traces of lexical tones in North Russian dialects /10/. Later in the same dialects she has discovered the register type opposition between questions and declaratives /11/.Here we ought to see only typological correlations.

I.Lehiste and P.Ivić recently have introduced a new hypothesis about the existence of a "new Balcanism" - yes-no question intonation type /12/. Here we have a case of areal convergency.

2.4.Some years ago I discovered a striking coincidence between the second Slavic yes-no question contour (with falling poststressed syllables) and the yes-no question melody in Finno-Ugrian languages. At the VI International Finnougrian congress in Syktyvkar in 1985 (Komi Republic) I recorded on magnetic tapes the melodies of Finnish,Estonian,Hungarian,Sami,Udmurt,Komi,Mordva and Mari questions. We investigated the melody of the following question types: general,wh-questions,alternative,the so-called precising questions and repeated questions. It turns out that the main types of question melody in Russian and Finnougrian coincide. (I would not like to discuss the reasons of this coincidence now). And yet in order to have the right to speak about this coincidence being systematic, we ought to discover the really rising question in Finnougrian languages /13/,otherwise we won't be able to formulate the distribution type similarity. We have found this rising-to-the-end-melody in repeated questions, echo-questions, questions with the word "and" (And mama?) at the initial position.

3. We have so far discussed the form-function correlation. Yet we have found one very important criterium comparationis: the force (or weakness) of influence of sentence prosody on the word prosody. In other words, it is significant to know how the word prosody is modified in different sentence positions.We have experimental data that there is a regularity of the changes in the word prosody due to the position of the word in the sentence (the beginning- the middle- the end;the sentence type and so on).There may be strong sentence positions and weak sentence positions / 14/.But this prosodic change is maximal in some languages and minimal in others. In the second case we may speak about a word-per-word (by-word) prosodic realizations, in the first case - about strong grammaticalization of the intonation type,.thus the word is no longer a basic prosodic unit,but only a part of a larger unit,i.e. syntagm,or a sentence /15/.

4. The criterium discussed above deals with an idea of diachronical intonation typology.This is a new problem and has not been considered so far. In other words we deal with sentence intonation reconstruction.Linguists do not yet possess any direct evidences for it. But intonologists are capable to draw conclusions from indirect proofs of archaic intonation states.We have to the point of adopting some hypotheses now.In my opinion it is high time to part with an idea of sentence as

a certain wire upon which words stressed and unstressed are moving. It is not the case. Sentence intonation has the right to have its autonomous structure and its own autonomous history as segmental phonetics has.And we have to reveal it and to describe it.

What indirect evidences do we have to discuss?

1/ The folk music and folk songs. The intonologist can draw from it information on word prosody, stress and pause means.

2/ Church music and church reading. For example,in Russian church reading there is a manner called "po-glasica" /16/ - a structural unit for reading which corresponds to a sense group.In Vladyshevskaya examples we can see a narrow pitch range (F_0) and very important role of T-structure.There exist special recordings with such data.

3/ Archaic texts such as fairy tales, legends, etc.These texts have specific prosodic form even to-day.

4/ Signs on manuscripts.In this case the intonologist can help the graphic interpretation.

5/ Re-interpretation of well known "phonetic laws" from the prosodic point of view.For instance, to my mind, the famous Wackernagel's law about the clitics and unstressed words in the second position can be re-formulated. I assume that it was not the words themselves but the position itself was so weakened,unstressed /17/that it was hardly audible.

6/ Certain concepts of comparative syntactic accentology.For instance,the data that the verb in the main clause was unstressed and in the subordinate clause- it was not. It can be explained in the following way: in Indo-European the subordinate clause usually preceded the main clause.So when the verb was in the end in both cases it had a rising melody in the end of the first clause and the falling one in the end of the main clause.

7/Inversion data (the movement by which I mean the location of the semantically loaded words to the initial position). A.Meillet underlined that in Old Greek the important word can occupy the initial position: μία δὴ λείπεται τῷ ἡμετέρῳ νόμῳ ὁδός. The same is true of Vedic /18/.It means that the beginning of the sentence was distinct, prosodically loaded.

8/The experimental investigation of poetry data.Experimental observations of Russian verses /19/demonstrate acoustic prosodic model quite similar to the Ukrainian language intonation,which is,so to speak,more "central" between the Slavic languages for its similarity feature index.It is necessary to note that Ukrainian has non-Slavic contacts to the last degree.

9/Diachronical proofs of word segmental phonetics such as compensatory lengthening in some languages,prosodic oriented manipulation in word paradigms in Greek, Latin,Sanscrit,etc.(γείτων - γείτονος ἀνήρ - ἀνέρος).They proved the prosodic autonomy of words during those periods.

10/ Phonetic evolution of different parts of speech conditioned by their different syntactic position.For example, initial Indoeuropean particles changed very little (pa-,ku-) and usually had a fortis consonant in word initial position.Thus the sentence beginning was tense and sharp.

11/ The spoken language phenomena.We know that spoken syntax of many languages has similar structures.The syntactic relations of spoken speech demonstrate the proximity to the "pragmatic code" speech model, as it was described by T.Givon /20/; the syntactic features can help to reconstruct the intonation.

The above mentioned considerations lead us to the hypothesis about different stages in the prosodic evolution.It is necessary to understand that we should distinguish between the unidirection of prosodic evolution on the one hand and the absolute chronology of its manifestation, on the other.For example, the North Russian dialects (see above) have a word-by-word pronunciation,relicts of lexical tones and register oppositions in intonation. This dialect co-exists with literary Russian which has a syntagm speech unit contrary to word-by-word pronunciation,has no proved tones and has large sets of grammaticalized intonation configurations.

5. I want to discuss the following very important hypothesis.In my opinion, the main diachronic universal of intonation typology is a movement in three stages: 1/before-word stage; 2/ word stage;3/postword stage.At the first stage the unit is utterance ,which is perhaps syllable divised and has a certain metric model.It is probable that there were sentences of gnomic or gimnic character, rithmically convenient for memorizing /21/. The word period (for many languages even to day) is connected with understanding the wordform as an autonomous and well-formed unit(i.e. with morphological development). The post-word period embraces languages with ancient traditions the sentence prosody of which is characterised by well-elaborated melodic contours semantically loaded, and by large prosodic units and meaningful juxtaposition of these large prosodic units.

We dont want to impose the obligatory uniformity for all languages and their prosodic evolution.Languages can stop their development at a certain stage, or they can develop some compensatory means.This naturally calls for future investigations. In fact all recent interesting works on tonogenesis cannot explain why some languages have conserved tones, and others have not.

6.D.Bolinger recently has named intona-

tion the linguistic "Cinderella"/22/.I invite to review this name and propose for discussion the law of language evolution based mostly at the prosodic data.In other words.Language changes transmitting more and more information in one expression unit.Because of the double articulation language has two possibilities of information augmenting: to increase the semantic load and to increase the speech rate,i.e. sense compressing and phonation compressing ways. All the fusion phenomena (such as flexions from pronouns, prepositions from nouns, compounds arising, etc.) are sense compressing.Yet perception capacity of human beings lag behind sense compressing: that is we cannot speak faster than 50-60 msec per sound.Consequently language should develop all suprasegmental phenomena in a broad sense: presupposition of all kinds, semantics of prominence, grammaticalized melodic contours, etc. All these factors help to create supplementary sense lines,i.e. supersegmentals of content plan.

References

/1/ See:T.M.Nikolayeva."Frasovaja intonacija slavjanskich jazykov",Moskva, 1977,Chapter 2.
/2/ I.Lehiste:"Suprasegmentals",Cambr.-Mass.-L.,1970.
/3/ T.M.Nikolayeva,Op.cit.,pp.64-65; D.A.Pavlov,T.S.Esenova,"Foneticheskaja characteristika i fonologicheskij status kalmyckogo i mongol'skogo jazykov,"Fonetika jazykov Sibiri i sopredel'nych regionov",Novosibirsk,1986.
/4/ T.M.Nikolayeva.Op.cit,p.64.
/5/ A.N.Nurmachanova,"Ob intonacionnych osobennost'ach tipov predloženij v t'urkskich jazykach",Učenuje zapiski MGPIIJA,1971,t.60.
/6/ See,specially: D.Bolinger,"Intonation across languages",Human languages, Standford,1978; J.J.Ohala,"Cross-language use of pitch:ethological view",Phonetica,1983, N 1.
/7/See,specially:Svetozarova N.D:"Intonacionnaja sistema russkogo jazyka", L.,1982.
/8/ T.M.Nikolayeva ,Op.cit.,Chapter IV.
/9/ V.I.Petr'ankina,"Funkcional'nuj aspect intonacii i tipologija jazykov", Prosodija sloga- slova - frazy,M., 1981.
/10/R.Paufoschima ," Sledy muzikal'nogo udarenija v sovremennom vologodskom govore",Dialektografija russkogo jazyka,M.,1985.
/11/ R.F.Paufoschima," Ob ispol'zovanii registrovych različij v russkoj frazovoj intonacii",Slav'anskoje i balkanskoje jazykoznanije.Prosodija /in print/.
/12/ I.Lehiste,P.Ivič,"The intonation of yes-or-no questions - a new Balkanism?",Balkanistica,VI,1980.
/13/See the opposite opinion: A.Iivonen, "Is there interrogative intonation in Finnish?",Nordic prosody,Lund,1978.
/14/T.M.Nikolayeva,Op.cit.
/15/ We use now L.Shcherba's term without explaining it.
/16/ T.Vladyshevskaja,"K voprosu o roli vizantijskich i nacional'nych russkich elementov v processe vozniknovenija drevnerusskogo cerkovnogo penija,M.,1983.
/17/T.M.Nikolayeva,"Funkcii chastic v vyskazyvanii",M.,1985,pp.116-27.
/18/ A.Meie/Meillet/,"Vvedenije v sravnitel'noje izuchenije indoevropeiskich jazykov",M.-L.,p.371.
/19/L.V.Zlatoustova,"Izuchenije zvuchashchego sticha i chudozestvennoj prozy instrumental'nymi metodami", Kontekst,1976,M.,1977.
/20/ Talmy Givon," On understanding grammar", N.-Y,etc,1979.
/21/ T.M.Nikolayeva,"Stichotvornaja i prozaicheskaja stroki:pervichnoje i modificirovannoje",Balcanica, M., 1979.
/22/ D.Bolinger,"Intonation and its parts. Melody in spoken English",Standford, California,1986,p.3.

LINGUISTIC FACTORS IN SENTENCE STRESS (EVIDENCE FROM RUSSIAN)

N.D.SVETOZAROVA

Department of Phonetics
Leningrad University
Leningrad, USSR 199034

ABSTRACT

The paper deals with various linguistic factors of sentence stress assignment and detection in spoken and written Russian texts.

The aim of the study is to test such different factors as "parts of speech", "syntactic relations", "word order", "semantic structure of the word", "rhythmic patterns", "newness/givenness of information" and "actual sentence division" for their reliability in real text conditions and thus to establish a hierarchy of these factors both in production and perception of spoken and written material.

INTRODUCTION

This paper deals with sentence stress which is understood as a system of accents varying in relative strength (prominence) and quality (melodic shape). The system of sentence accents is treated as part of the intonation system of language.

The experimental basis for this study consists of 4 types of speech material: a) spontaneously spoken texts (monologues and dialogues), b) written texts on different subjects which were read aloud, c) summarized tonetic transcriptions of text fragments, d) listeners' reactions to different types of accentual patterns.

The special aim of the study is to summarize various factors influencing the distribution of sentence accents within texts in order to test their contribution to two practical tasks: teaching of sentence stress assingment and detection to foreign language students and working out rules for automatic sentence stress assignment and detection in the systems of speech analysis and synthesis.

BACKGROUND

A great deal of experimental research has been done, and is still being done, on intonation. However, within the intonation system it is specifically sentence stress which lacks experimental data both in speech production and perception. Most of the work on this subject is done not by phoneticians, but by syntacticians, especially those who study word order and Functional Sentence Perspective. Yet one may recognize that the importance of accentuation has often been underestimated in these studies.

It is impossible to cite all of the most valuable works pertaining to our subject-matter. From Western authors I will mention just a few whose ideas and data are especially relevant to my understanding of sentence stress and who summarize previous studies. These are: D.L.Bolinger /1,2/, D.R.Ladd /3/, A.Fuchs /4,5/, C.E. Keijsper /6/, C.Gussenhoven /7/, I.Fougeron /8/, A.Lötscher /9/.

In the Soviet Union research on sentence stress is not a new phenomenon. L.V.Shcherba based his linguistic analysis of poetic works on the detailed transcription of various degrees of prominence /10/. The notion of the accent in a sentence plays an important role in the works of I.I.Kovtunova /11/, V.E.Shev'akova /12/, G.A.Zolotova /13/, D.N.Shmelev /14/ and E.V.Paducheva /15/ on communicative aspects of Russian syntax.

One of the most inspiring books on the semantics of sentence stress is that of T.M.Nikolaeva /16/.

Recently some experimental works on sentence stress in Russian have appeared, and especially those by O.F.Krivnova /17/, T.P.Skorikova /18/, T.M.Nadeina /19/, A.V. Pavlova /20/ pertain to this paper.

FACTORS INFLUENCING SENTENCE STRESS

In the search for factors which determine different kinds of sentence accentuation within a text, the investigator may come to the conclusion that these factors are various and occasionally contradictory.

The well-known controversy initiated in "Language" in 1971-1972 as to whether accent placement is determined by syntax or by semantics has not been satisfactorily resolved yet: compare the recent polemics between D.Bolinger and C.Gussenhoven in "Journal of Linguistics" /2,7/.

In trying to isolate one factor which governs rules of sentence accentuation we risk oversimplification. In admitting a set of equally important factors we cannot explain cases in which different factors are in opposition or even in conflict. Therefore we have established a hierarchy of factors to account for the experimental data. I will now turn to the linguistic factors which will figure in the paper and present each of them separately beginning with those which are more easily formalized.

1. Parts of speech

In Russian, as in other languages, nouns, verbs, adverbs and other so called "full" words usually carry stress whereas most of the "function" words are unstressed, unless they bear contrastive or emphatic stress. This rule held for the texts studied in ca. 90% of cases, but it does not account for words carrying different degrees of stress. Both "full" and "function" words in Russian carry various degrees of accentual prominence which cannot be explained by such a simple rule. Thus it is not only the part of speech that determines the degrees of prominence, but also its function in a given word-group.

2. The syntactic factor

In Russian, as in many other languages, there exist rules which assign accent patterns to different kinds of sentence constituents. Thus in Subject-Predicate groups, nouns are usually more weakly stressed than verbs, whereas, conversely, in Predicate-Object groups nouns are stressed more than verbs:

Ivàn chitáyet - chitàyet knígu
/John is reading -(he)is reading a book/.

In our texts the number of such "normal" cases was only ca. 75% of all word-groups and the percentage seems to be very sensitive to the kind of the text - spoken or written, literary or scientific. The frequent deviations from the standing rules indicate either that the rules are incorrect (or too rough), or that other factors influence sentence stress, or both.

3. Word order

In Russian there exists a strong tendency for words and word-groups to receive stronger stress in certain sentence positions. Thus sentence-final and initial positions are most often connected with stronger stress, whereas sentence-medial position carries weaker stress.

It is obvious, however, that word order in Russian - flexible as it is - is governed by various influential factors and only in some cases it is the position itself that determines sentence accentuation (for example in enumerations).

It is not the task of this study to account for the complicated system of word order in Russian. There is a great body of work not only on word order itself but also on the complex relations between word order and sentence stress (which is in my opinion the only explanatory way to discuss word order). Here it is important to underline the fundamental difference between word order rules for written and spoken texts in Russian. When writing, it is obligotary to follow word order rules if one wants his intended emphasis to be understood correctly, for the place of the word in a sentence is one of the primary means to detect the degree of its prominence. In oral speech, however, there is a possibility of expressing one's intentions directly by prosodic means. In our texts, neutral (non-emphatic) sentence stress occured in sentence or clause final position in ca.90% for scientific texts and in ca.70% for dialogues.

4. The rhythmic factor

Strongly connected with all the factors already discussed is what one may call the rhythmic factor, i.e. a tendency to alternate accents of different degrees. In Russian, this tendency is most obvious in spontaneous speech /21/. In the texts analyzed we found that non-final strong and extra-strong accents were preceded by relatively weaker accents or followed by them in ca.95% of cases. Our data indicate that the structure of accent pattern is asymmetrical in Russian, that is, that accents following the main stress are usually weaker than those preceding it. Compare:

On chitàyet knígu /He is reading a book/
On chitáyet knìgu/He is reading the book/

It is clear that the rhythmic factor can explain the weakening of the accent on one word in the presence of a stronger accent on another, because due to the rhythm all words within a Russian sentence cannot be equally accented. This factor also explains how deaccenting of one word in the sentence results in the accentual prominence of some other word, because again for rhythmic reasons, all words cannot be equally unstressed. But which word will be stressed depends obviously on other factors to which I will now turn.

5. The semantic factor

The fact that some words in Russian are more likely to be accented because of their inherent meanings has not until recently been properly considered by re -

searchers. Recent investigations /16,18, 20/ have shown that in many cases it is the peculiarity of the meaning of a given word which forces the speaker to stress it more than other words in the sentence. The role of the semantic structure of the word is especially evident for those words which do not bear the main accent due to the factors discussed above. Such words are in Russian, for example, adjectives before nouns and verbs with an objective complement. The former have been studied in connection with sentence accent by T.P.Skorikova /18/, the latter - by A.V.Pavlova /20/.

Adjectives and verbs in Russian which normally do not bear the main accent in such word-groups may get this accent if they contain the meaning of evaluation:

novaya kniga /a new book/
prekrasnaya kniga /a wonderful book/.

Factive verbs, verbs with the meaning of reaction, retrospection, contrast, negation (both grammatical and inherent) usually bear a strong accent /20/:

Ivan vspomnil o knige
/John remembered the book/
Ivan zabýl o knige
/John forgot about the book/

The semantic complexity of such verbs is often combined in Russian with definiteness or even "emptiness"of the nouns:

Eto raznye véshchi
/These are different objects/
Eto ráznye veshchi
/It's quite different/

In an experiment made by A.V.Pavlova, subjects read sentences of this type with different accent patterns.

In the texts analyzed, cases of verb accentuation due to semantic factor were quite numerous. The lack of accents on such verbs must be accounted for by the operation of other factors, which, in my opinion, can also be formalized. (For more detail see /20/.

6. Given/new information

In Russian, as in other languages, words expressing new information are usually accented and words expressing given information are either accented and pronounced with rising tone or left unaccented. There are not enough experimental data on how sentence stress reacts to different kinds of new and given information (for English see /22/).

I will not discuss this apparantly universal tendency any further. My concern is to test the possibility of formalization of the given/new distinction in Russian texts. That is why I put aside such hardly formalizable concepts as presupposition and the so called "fund of common assumptions".

In the experimental material only ca.75% of "new" words had the main stress and in ca.15% of cases the main accent was on "given" words despite the formal indication of their givenness in the context. Thus the degree of accent is not fully determined by the context.

Many of the deviations from the noted tendency can be explained by the factors discussed above. For example, if a new concept is expressed not by one word but by a word-group, then only one word of the group will get a stronger accent according to syntactic and semantic rules. This indicates that the word-group is understood as a whole. Otherwise the unity of the concept is destroyed.

Other deviations can only be explained if we consider new factors. Both deaccenting the "new" and accenting the "given" can be explained by the concept of "focus", which is closely connected with the theory of Functional Sentence Perspective (in Russian linguistics commonly "Actual Sentence Division").

7. Actual sentence division

This universal and high level factor influences both the placement of accents and their type (thematic vs. rhematic accents) in Russian. Sentence stress patterns are essentially different in "undivided" (actually non-segmented) and "divided" utterances. In both kinds of structures a further subdivision into different "degrees of importance"/23/ or "information centres" /24/ is often possible. In oral speech the main means of expressing the ASD in Russian is the system of sentence accents which is far more common than either the syntactic or lexical means of distinguishing the theme from the rheme.

In the texts studied, the coincidence of different accent patterns and the corresponding ASD types was very high. As for the formalization of this factor we must admit the free choice of the author in selecting the type of ASD (divided/undivided) and the type of focus (broad or narrow). But this freedom is not unlimited. Certain communicative conditions show a high degree of affinity to certain ASD types. For example, when the speaker wants to express the meaning not only of contrast but also of result, reaction, estimation or negation, he tends to use a narrow focus, thus singling out a part of the whole by means of a "special" stress. Lack of emphasis, on the opposite, leads to the use of neutral stress, that is, to accenting the word-group as a whole.

CONCLUSION

This pilot study of the distribution of sentence accents in Russian has shown that the place and the type of sentence stress in Russian texts is to a certain extent predictable and therefore formalizable, but only with due regard for the highly complicated interplay of various factors which expose different hierarchies under different communicative conditions. In reading - both by man and machine - we proceed from word order, constituent structure, lexical meaning and other formalizable factors to the underlying actual sentence division type and from that to accent pattern. In listening we move from the perceived accent pattern to the intended actual sentence division and from that to communicative meaning. In speaking - from communicative meaning through actual sentence division to the realization of an accent pattern.

ACKNOWLEDGEMENTS

The author is very much indebted to all colleagues who have participated in the experiments and in the discussion of the data, especially to A.V.Pavlova and I.Panas'uk. I am also very grateful to Mr.John Bailyn for help in reading the English version of the paper.

REFERENCES

/1/ D.L. Bolinger Accent is predictable (if you're a mind reader).-Language, 48, 1972,633-644
/2/ D.L.Bolinger Two views of accent. - Journal of Linguistics, 21,1985,79-123
/3/ D.R.Ladd The structure of intonational meaning. Bloomington a.London,1980
/4/ A.Fuchs "Normaler" und "kontrastiver" Akzent.-Lingua,38,1976,293-312
/5/ A.Fuchs "Deaccenting" and "Default Accent".// Intonation, Accent and Rhythm. Studies in Discourse Phonology. Ed.by D. Gibbon and H.Richter. Berlin, New York, 1984,134-164
/6/ C.E.Keijsper Information structure. With examples from Russian, English and Dutch. Amsterdam,1985
/7/ C.Gussenhoven Focus, mode and the nucleus. - Journal of Linguistics, 19, 1983, 377-417
/8/ I.Fougeron Les moyens prosodiques dans l'organisation du message. // IIIe Colloque de linguistique russe. Paris, 1983, 111-121
/9/ A.Lötscher Satzakzent und Funktionale Satzperspektive im Deutschen. Tübingen, 1983
/10/ L.V.Shcherba / Selected writings on Russian language/.Moskva,1957,26-44, 97-109 /in Russian/
/11/ I.I.Kovtunova /Word order and actual sentence division in modern Russian/. Moskva,1976 /in Russian/
/12/ V.E.Shev'akova /On logical stress/.- Voprosy jazykoznanija,1977,N 6,107-118 /in Russian/
/13/ G.A.Zolotova /Communicative aspects of Russian syntax/. Moskva, 1982 /in Russian/
/14/ D.N.Shmelev /Syntactic division of the sentence in modern Russian/. Moskva, 1976 /in Russian/
/15/ E.V.Paducheva /The utterance and its reference to reality/. Moskva, 1985 /in Russian/
/16/ T.M.Nikolaeva /Semantics of accentual prominence/. Moskva, 1982 /in Russian/
/17/ O.F.Krivnova /On the accentual function of melody/.- In: Intonacija. Kiev, 1978,119-143 /in Russian/
/18/ T.P.Skorikova /Functional capacity of the intonation pattern of word-groups in speech-flow/. Cand.of Sci.Diss. Moskva, 1982 /in Russian/
/19/ T.M.Nadeina /Accentual structure of the utterance in Russian/. Cand.of Sci. Diss. Moskva, 1985 /in Russian/
/20/ A.V.Pavlova, N.D.Svetozarova /Factors determining the degree of accentual prominence of a word in an utterance/. - In: Slukh i rech v norme i patologii. Leningrad, 1987 /in Russian/
/21/N.N.Rozanova /On dynamic instability of words in spoken Russian/.- In: Sintaksis teksta. Moskva, 1979 /in Russian/
/22/ G.Brown Prosodic structure and the given/new distinction. // Prosody: Models and Measurements. Ed. by A.Cutler and D. R.Ladd. Springer Verlag, 1983, 67-77
/23/ I.G.Torsueva /Intonation and the meaning of an utterance/. Moskva, 1979 /in Russian/
/24/ O.A.Lapteva /Syntax of colloquial Russian/. Moskva, 1976 /in Russian/

WORD STRESS IN PROSODIC CONTEXT

A.M. SLOOTWEG

Institute of Phonetics, Nijmegen University
Erasmusplein 1
6500 HD Nijmegen, The Netherlands

SUMMARY*

This contribution deals with the acoustic realisation of 'primary' versus 'secondary' stress in monomorphemic and compound words in Dutch, and with the effect of sentence accent on the acoustic realisation of syllables carrying lexical stress in simplex words. Measurements show a linear phonetic factor interacting with phonological variables to determine the acoustic realisation of lexical stress. Also, the important acoustic correlates of stress turn out to react differently to phonological prominence on higher prosodic levels. The unit under investigation is the syllable with 'main' stress; we believe the syllable to be the domain in which stress is manifested.

INTRODUCTION

Most phonetic work on stress has involved the search for the acoustic manifestations of lexical stress, and the question of the perceptual 'cue value' of each of a number of acoustic parameters which have come to be associated with stress. Stressed syllables differ from unstressed ones in that they show longer duration, specific pitch movements and more intensity. The difference between stressed and unstressed syllables within one word has been studied extensively (see Niemi (1984) for an overview). However, we know of no systematic investigation as to the influence of prosodic levels higher than the word on the realisation of lexical stress.
In phonological theory, the prominence relations on these higher levels have been specified, albeit in non-acoustic terms. Pertinent to the present study are such relations on the word-formation level, the level of the phonological phrase, and sentence level.
Whenever two words are joined to form a compound, one turns out to be stronger, and the other less prominent. The recently developed 'metrical' framework in phonology captures this generalization by always grouping constituents into binary pairs, within which the strong-weak relation is formally defined. Thus, it is impossible to end up with two 'primary' stresses.
Under the shorthand name of 'status', we studied the acoustic difference between identical syllables in the weak parts of compounds or as the stressed syllable of simplex words.
This material was embedded in the higher-level prosodic structure known as the phonological phrase. Within this type of word group, there also exists a strong-weak relation between constituents. In accord with Liberman & Prince (1977)'s Lexical Category Prominence Rule, the second part of Dutch phrases is more prominent than the first. The rule operates without regard to the internal structure of the constituents, so that the weak element in the compound retains its status of being weaker than the first element. In the experiment reported on here, the phonological phrase is the context in which compounds and simplex words appear and within which phonetic conditions were rigidly controlled.
The highest level where prosodic prominence relations occur is that of the sentence. Sentence accent distinguishes strong and weak (more prominent or less so) intonational phrases, on pragmatic and contextual bases. The effect of sentence accent on syllables of either status is the second issue that was studied. Traditionally, sentence accent is primarily associated with specific prominence lending pitch movements, while acoustic features like intensity and syllable duration are affected to a lesser degree. In the context of the present experiment, there were two questions to be answered:
1) What is the difference, in acoustic terms, between a syllable assigned phonological main stress and the same syllable bearing 'secondary' stress in the weaker phonological environment formed by the second part of a compound?
2) How does the presence or absence of sentence accent affect these characteristics?
To answer these questions, acoustic measurements were performed on some 500 syllables, pronounced in controlled phonological and phonetic contexts. Along with providing insights into the acoustic effects of both status and sentence accent, this setting disclosed the influence of a linear phonetic parameter, namely the position of the syllable within the word.

THE EXPERIMENT

Material

Speech material was devised to study the influence of status and accent, and to introduce a third independent variable, the position of the lexical stress within the word. The material was based on fifteen tri-syllabic target words with lexical stress on the first, middle or last syllable, five times each. The phonetic variable of position of the target syllable in the word shall be called 'type n', the numeral indicating first, second or third position. To minimize acoustic differences caused by syllable make-up, we strove for uniformity in the consonant-vowel structures. Accordingly, (almost) all syllables consisted of one consonant and one long vowel. Five vowels occurred, balanced across word position types. The words are listed at the end of this paper.
To form two kinds of word groups in which the target word would be either strong or weak, we embedded that word in an adjective-noun phrase, where the target would be strong, and in a nominal compound, where it was the second element and, thus, weak. So, the syllable could end up in a monomorphemic word ('main stress') or in the weaker part of a compound ('secondary stress'). For example, (the target syllable is the middle syllable of 'kimono'):
(main stress): 'die rode kimono'(that red kimono)
(secondary stress): 'die mode-kimono'(that fashion-kimono)
As the examples show, the phonetic context within the phrases was kept as similar as possible. The number of syllables preceding the stressed one was kept constant, through the introduction of the adjective. The CV-structure of all syllables within the phrase was the same, with identical vocalic segments.
In order to study the effect of the presence or absence of sentence accent, we used two different sentences of which the phonological phrase could be a part. One sentence prompted a reading with sentence accent on the phrase in question. In the other sentence the phrase was relatively unimportant; it was easy to pronounce it without sentence accent. The sentences were:
(accent): 'Vergeet niet die .... in je toespraak te noemen.'
(Don't forget to mention that .... in your speech.)
(no accent): 'Ik geef NIKS om die .... die ze daar aanprijzen.'
(I don't care AT ALL for that .... they're selling there.)
Both sentences were to be pronounced as one intonational block.

Speakers, instructions and the recording session

Speakers were 8 males, who pronounced the set of sentences presented to them in random order. Each speaker produced 60 utterances: the product of 15 words x 2 status possibilities x 2 sentence accent conditions. After some training, they were all able to produce the accent patterns desired. This does not mean they delivered exact replicas of an example: performance was checked on the basis of perceptual equivalence. Recordings were made in a soundproofed studio with semi-professional equipment, the sessions led by the experimenter.

Measurements and computed values

Acoustic measurements were performed on the syllables with lexical stress with and without accent, both in compounds and phrases. 1) For each syllable, a number of values to capture pitch, duration and intensity features were obtained. From these, three new variables were computed, taking into account speakers' speech rate, and melodic and dynamic speech characteristics.
-The new duration variable expressed syllable duration as a proportion of the sum of the four unreduced syllabes in the compound or comparable phrase.
-Syllable intensity was expressed in dB above the intensity baseline. This was defined as the lowest mean value observed on unstressed syllables, for each speaker and condition separately.
-To arrive at a new pitch variable, the aim was to express pitch changes rather than absolute values. These changes were then to be related to speakers' melodic ranges. Pitch range was defined as the difference between highest pitch observed on stressed accented syllables and lowest pitch as found on unstressed syllables in a condition without accent, both mean values for each speaker and condition separately. The pitch change was chosen as the largest pitch movement occurring on a given syllable, and expressed as a percentage of the pitch range. 2) While using the variable names of 'pitch', 'duration' and 'intensity', reference is made to the variables defined as above.
Mean values were computed for these new variables, to gain an overview of the acoustic differences in each linguistic condition. Analyses of variance served to determine which of the independent linguistic variables (status, accent and the linear phonetic factor described above) most influenced the dependent acoustic variables.

RESULTS

The results for the three variables are presented separately. An overview of the results of the analyses of variance is in the Summary Table of Effects, given in the final section.

Intensity

As Figure 1 shows, there is an effect of word type on the values of the intensity variable. They range from means of 7.7 to 7.5 to 4.7 dB above the baseline for types 1 to 3: $F(2,384)=3.08, p<.001$. It is clear that the final position in the word leads to low intensities. This tallies with suggestions made in Pierrehumbert (1979), where the phenomenon of amplitude downdrift in sentences is described.

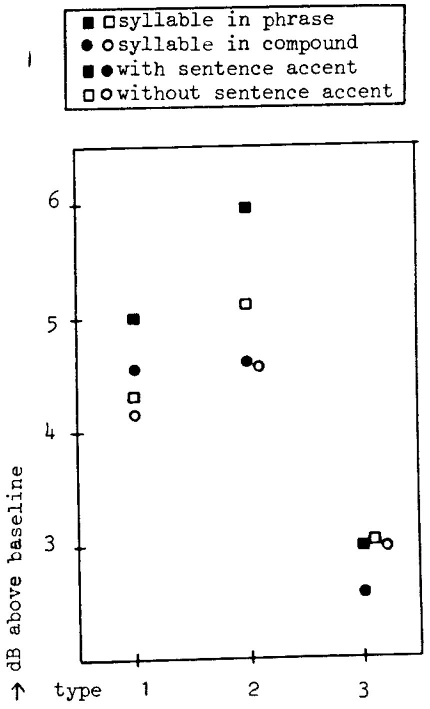


Figure 1: Mean intensity, expressed as distance from speaker's intensity baseline, of stressed syllables in various positions, in compounds and phrases, with and without sentence accent.

The status of the stressed syllable also produces a clear difference in intensities: means of 7.1 dB for syllables with primary stress, against 6.2 dB for syllables with 'secondary' stress in compounds. This difference, too, was statistically significant: $F(1,384)=17.77, p<.001$.
The presence of sentence accent resulted in higher intensities. The mean values are 7.5 dB for accented and 5.8 dB for unaccented items. Again, this difference was significant: $F(1,384)=72.50, p<.001$.
There were no interactions at the significance level employed in this study.

Proportional syllable duration

The results for the duration parameter are presented in Figure 2.
The most striking effect on the durational parameter can be seen on the horizontal axis. As the stressed syllable is situated further back in the target word, its proportional duration increases from 25% via 29% to 30%. This fully agrees with the observations in Nooteboom (1972), where vowel duration is a function of the number of syllables to follow in the word. The results show that the same holds for proportional duration of non-identical syllables: $F(2,384)=63.73, p<.001$.
The factor status also causes a significant effect: $F(1,384)=24.91, p<.001$. If the stressed syllable forms part of a simplex strong element in the phrase, its duration is longer than if it is in the weak part of a compound: 29% and 27%, respectively.
Remarkably, the presence or absence of sentence accent had no influence on the proportional duration. In the analysis of variance the effect of accent was not significant.
There were no significant interactions between the independent variables. This means that the actual increase of the proportional duration across word types is independent of the status of the word. The two factors of type and status operate separately.

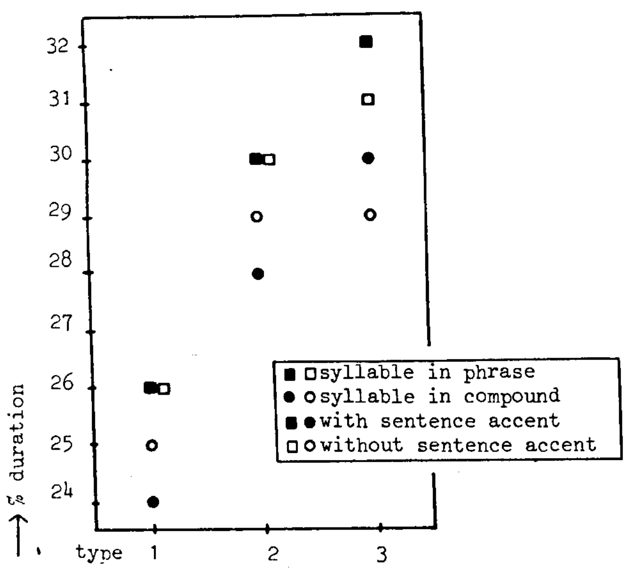


Figure 2: Mean proportional syllable duration of stressed syllables in various positions, in compounds and phrases, with and without sentence accents.

Pitch movements

The mean values for the proportional pitch movements are presented below. Figure 3 shows that word type

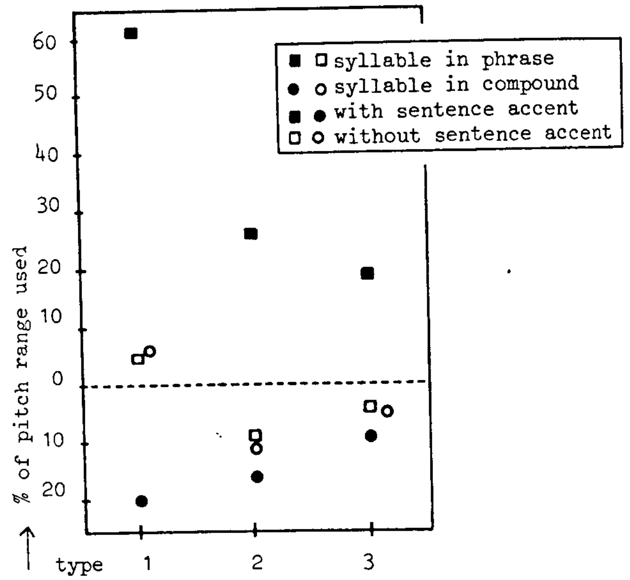


Figure 3: Mean pitch movement, expressed as a proportion of the speaker's pitch range, for stressed syllables in various positions, in compounds and phrases, with and without sentence accent.

has an effect on the size of the pitch movements. Disregarding the direction of the movements, (and the important effects of status and accent) they take up 23%, 15% and 8% of the speaker's pitch range for types 1, 2, 3, respectively. This effect was found to be significant $F(2,384)=12.78, p<.001$. 4)
A strong effect was caused by the factor status: if the word was itself the head of the phrase, the mean pitch movement on the stressed syllable was 16% of the pitch range, while it was only 8% on the syllables in the second elements of compounds. Also, in the latter case, the direction of the change was falling rather than rising, as in the former. The analysis of variance yielded a significant effect: $F(1,384)=90.05, p<.001$.
Accentuation within the sentence causes clear pitch movements, both rising and falling, with an overall mean of 11% of the pitch range. Syllables in unaccented words showed falling pitch through 3% (mean) of the pitch range. The effect of accent was, of course, significant: $F(1,384)=28.04, p<.001$.
First order interactions play a part in the variance of this variable. The values for status depend on the word type, as figure 3 shows: from type 1 to 3, the difference between 'secondary' and 'primary' stressed elements diminishes. This is a significant effect: $F(2,384)=10.19, p<.001$.
Also, status and accent cooperate to produce significant differences: $F(3.384)=84.52, p<.0001$. The effect of accent on pitch changes is much more marked on syllables in simplex words, than on the 'secondary' stresses.
The second order interaction of type and status and accent yielded significant results as well: $F(8,384)=11.61, p<.001$.)

DISCUSSION

Table 1 below gives the results of the analyses of variance, with a significance level of $p<.001$. 4)
Position of the target syllable within the word had a significant effect on all three variables studied. As the figures showed, this effect was not the same for the three variables; duration increased towards the end of the word, pitch movements got smaller, while intensity was negatively affected by the final position in the word. The status of the syllable had an effect as well. If a syllable was embedded in a compound, as the least prominent element, its intensity was lower, its duration is (generally) shorter and its pitch movements were smaller than in a simplex word. As to sentence accent, notice how this increased syllable pitch and intensity, but not duration. Higher order interactions occured with the pitch variable only.
Of the three variables studied, duration is the most

effect	p	d	i
type	*	*	*
status	*	*	*
accent	*	ns	*
type x status	*	ns	ns
statusxaccent	*	ns	ns
typexaccent	ns	ns	ns
typexstatusxaccent	*	ns	ns

Table 1: Summary of the analyses of variance performed on the P(itch), D(uration) and I(ntensity) data.

'straightforward'. Only type and status play a role in the values obtained. Intensity is also affected by sentence accent. The pitch movements constitute the most complex variable; not only are status, accent and the linear phonetic factor important as such, interactions of these variables further determine the behavior of the pitch variable.

The results of the present study show unequivocally that what is loosely referred to as (relative) prominence in prosodic literature, can be specified in a precise and insightful fashion when dealing with production aspects of prosodic structures. The cover-all term prominence can be split up into a number of separate acoustic correlates of stress and accent, each of which reacts differently to another phonetic parameter, namely that of position in the word.
The strategy to tackle this wealth of subtleties produced in speech would be to find out which of the observed regularities are also perceptually relevant, and to what extent. Currently, we are investigating the interaction of syllable position and size of preferred pitch movements, in a pitch adjustment experiment with naive listeners. It is only on the basis of results in the perceptual realm that we can realistically assess what is essential in higher level prosodics.

NOTES

*This research was supported by the Foundation for Linguistic Research, which is funded by the Netherlands Organisation for the Advancement of Pure Research, ZWO. Thanks are due to Jip Wester, for criticism and patience, and to Bert Cranen, for help with the measuring procedure.
1) Measurements were performed by means of a speech editing system and the ILS speech analysis system.
2) No claim is made as to the perceptual relevance of this variable's definition. It could be argued that conferring Hz to semitone values would be a way to arrive at a direct coupling of measured acoustic differences and perceived stress differences. In this paper, we leave the matter undecided.
3) We have compared the behavior of the pitch movement variable to that of the absolute pitch reached in he syllable. Effects of type, status and accent were present, with this difference: from types 1 to 3, the pitch peak was somewhat lowered, while the pitch movement diminished considerably. Examination of the measurements revealed no inconsistency: in final positions, the pitch peak was reached via earlier steps on preceding syllables, so that the final step was indeed the smallest.
4)In spite of our efforts to rule out speaker-related variation in the variables computed, the factor of 'speaker' still caused significant effects on duration and intensity. Also, two-factor interactions of 'speaker' and other independent variables occurred in the values of both pitch variables.

REFERENCES

Liberman, M. & Prince, A., On stress and linguistic rhythm. Linguistic Inquiry (8) pp. 249-336, 1977.
Niemi, J., Word level stress and prominence in Finnish and English. University of Joensuu publications in the Humanities, Joensuu, 1984.
Nooteboom, S.G. Production and perception of vowel duration. Unpublished doctoral dissertation, University of Utrecht, 1972.
Pierrehumbert, J., The perception of fundamental frequency declination. Journal of the Acoustic Society of America, 66, pp. 363-369, 1979.

LIST OF WORDS USED

Stress on first syllable: bariton, risico, dominee, genius, junior.
Stress on second syllable: dynamo, familie, kimono, komedie, illusie.
Stress on third syllable: chocola, melodie, mirabeau, defilé, residu.

ENGLISH IN NATIVE AND FOREIGN MOUTH — A LARYNGOGRAPHIC STUDY OF POLISH-ENGLISH CONTRASTS

KATARZYNA DZIUBALSKA-KOŁACZYK

Department of English
Adam Mickiewicz University
Poznań, Poland

ABSTRACT

Lx waveforms and, consequently, fundamental frequency histograms manifest characteristic properties similar for all normal healthy speakers. It has been found out, however, that there exist differences in preferred Fx modes depending on a language. Also, learners of foreign languages may deviate from their Fx patterning when speaking the language learned.

The paper aims at discovering some of those differences and deviations by means of analyzing the data resulting from an experiment conducted with 15 Polish and English speakers, in which an electrical impedance method of observing vocal fold vibration was used.

0. The healthy larynx of any speaker pronouncing e.g. a clear long a sound produces a vibration which, when recorded as a waveform, has a characteristic shape with recognizable and repeatable features (the so-called Lx waveform). Consequently, fundamental frequency (Fx) histograms obtained from Lx waveforms also manifest characteristic properties similar for all normal healthy speakers. This, however, might not be the case when speakers of different languages are involved. Differences have been found out in preferred fundamental frequency modes in particular languages (cf. 10: 126). This, among others, could be the reason for learners of foreign languages to deviate from their characteristic pattern while speaking the language learned.

The paper aims at either discovering or disconfirming the existence of some of those differences and deviations with reference to Polish and English. No matter the nature of the results, the study might prove useful in building up a theory of second language acquisition and, consequently, it may be insightful for theoretical, contrastive and foreign language teaching purposes.

In what follows, first, an electrical impedance method of observing vocal fold vibration will be described; second, the subjects, conditions, apparatus and procedure applied in the experiment will be presented; third, both a statistical and visual analysis of the results will be conducted; finally, a discussion of the results against the background of some other experiments administered by the author will end the paper.

1. An electrical impedance method of observing vocal folds activity possesses several advantages: it is non-invasive and thus relatively easily applicable (although still some speakers are hard to persuade to place the electrodes correctly); a resulting Lx waveform is unaffected by any acoustic noise; and, Lx can be recorded on unsophisticated equipment.

A device used to monitor the varying impedance of vibrating vocal folds is an electrolaryngograph consisting of two electrodes applied superficially to the neck on both sides of the thyroid cartilage. The resulting output waveform is recorded on one track of a tape while the other track is occupied by a speech waveform (Sp) from a microphone.

The Lx waveform manifests characteristic features indicating the voice quality used by the speaker (e.g. normal, breathy, creaky or falsetto) as well as certain pathological deviations from normal speech and individual idiosyncrasies. Lx also provides a basis for the analysis of fundamental frequency patterning (Fx) for particular speakers: periods of vibration of the vocal folds are easily convertible into Fx values so that an Fx histogram is obtained. This frequency distribution of vocal folds vibrations also manifests a characteristic shape with speaker-specific ranges and preferred modes. Fx histograms are obtained from single, double or triple period analysis, increasingly emphasizing the modal values against frequency irregularities indicated by low probability figures. Fx distribution can also be presented in the form of a scattergram showing a correlation between subsequent larynx frequencies (for a better visual presentation a scattergram might be converted into a 3D plot).

2. Fifteen Polish and English speakers (7 Polish, 7 English and 1 bilingual) were asked to read an IPA demonstration passage ("The north wind and the sun...") to the microphone while having laryngograph electrodes on. Polish subjects were asked to read both an English and Polish version of the passage, each of them twice; English subjects read their native text twice. A double reading was elicited in order to: a) obtain a sufficiently long speech sample, b) allow for a degree of text customization through the second reading which was thus performed in a more relaxed manner.

English subjects were all speakers of a broadly defined RP i.e. with, at the most, slight residues of a different accent. Among the Polish subjects, three were formal setting learners (they learned English in Poland), and four — natural setting learners acquiring English in England.

All recordings were done on a professional Marantz cassette recorder with the use of a dynamic microphone and a laryngograph set designed in the Department of Phonetics at University College London. Recorded waveforms were analyzed by means of a waveforms and Lx distribution programme on a BBC Master Series microcomputer with the input coming from an Uher CR 240 filtered through telequipment S61 due to which a visual representation of Lx and Fx could be seen. The analysis was conducted at the UCL phonetics laboratory.

The output of the analysis for each subject consisted of:

- an exemplary speech waveform (Sp) of a selected clear vowel sound
- an Lx waveform
- 1st, 2nd and 3rd order Fx histograms
- a statistical table including calculated mode, mean, variance, standard deviation, median and sample size for each order of distribution
- an Fx scattergram
- an Fx 3D plot

3. The shape of Lx for a healthy larynx is relatively stable. Therefore, its fundamental frequency distribution also possesses characteristic fixed features: modal peaks and sharp edges. The mode values and frequency range are speaker-specific. Irregularities in the overall shape, however, result from some abnormal voice condition like laryngitis or speech pathologies. Can they as well occur as a consequence of difficulties a learner encounters when speaking a foreign language? Still further, is there a possibility of "mode-switching" for the same speaker dependent on a language he is using? What kind of a relationship holds between physiological limitations on the laryngeal apparatus and linguistic structures? The above might be guiding questions to the type of experiment discussed in this paper with reference to Polish and English tongue, and Polish learners of English.

a) In order to verify a null hypothesis about Polish and English demonstrating similar tendencies for preferred fundamental frequency values, the t-test for small independent samples was used. It tests the significance of differences between two means.

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\left(\frac{\Sigma x_1^2-(\Sigma x_1)^2/N_1+\Sigma x_2^2-(\Sigma x_2)^2/N_2}{N_1+N_2-2}\right)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$$

where

$\bar{x}_1$ and $\bar{x}_2$ are sample means

x_1 and x_2 are variables

N_1 and N_2 are sizes of sample 1 and 2

t was calculated three times using different variables: means, modes and medians from Fx histograms for Polish and for English of the Poles and of the English respectively. In all cases it proved non-significant. Thus, the above null hypothesis cannot be rejected: linguistic structures of Polish and English are not distinct enough to introduce significant difference between Lx's (and, consequently, Fx's) of the respective native-speakers of these languages.

b) t-test was also used to compare English of the English subjects with English of the Poles. The null hypothesis this time was that Poles in general do not alter the output of their vocal folds activity when speaking English. Again, there was no significant basis for accepting any alternative hypothesis.

c) A study of correlation between Polish and English of the Polish subjects, however, did show a certain tendency for different Fx patterning for some speakers depending on a language spoken. The assessment of correlation was conducted for three cases: correlation between Polish and English for all Polish subjects together; correlation for Polish formal setting learners, and correlation for natural setting learners. The Pearson product-moment correlation coefficient was calculated with Fx 3rd order mean as a variable.

$$r=\frac{N\Sigma xy-\Sigma x\Sigma y}{\sqrt{\{N\Sigma x^2-(\Sigma x)^2\}\{N\Sigma y^2-(\Sigma y)^2\}}}$$

where

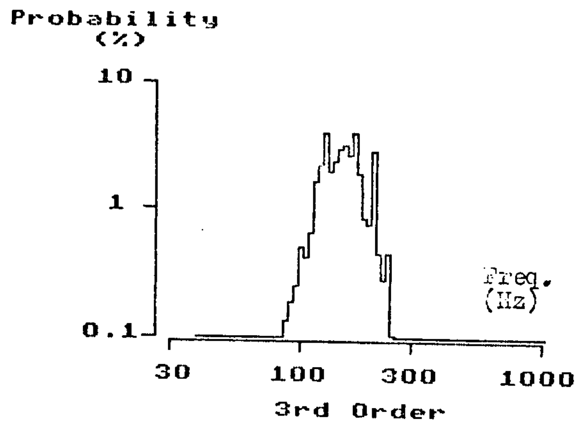
x and y are variables
N is a number of pairs of observations

The results are the following:

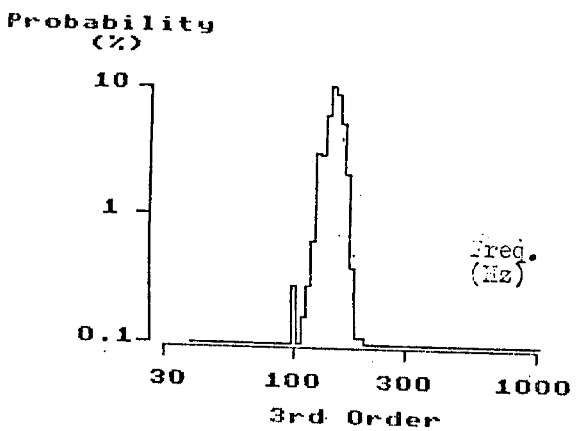
1) r = 0.84 for the whole group of Poles, i.e. a correlation is significant at 0.05 level

2) r = 0.999 for formal setting learners, i.e. shows a strong positive correlation
3) r = 0.879 for natural setting learners - significance at 0.20 level only, i.e. the probability for correlation is almost 20% lower than in 2).

Visually, English Fx histograms for Poles show some minor divergencies from their English counterparts, namely: irregularities in lower frequencies, or higher probabilities for lower frequencies, or a smaller frequency range.
Compare an example:

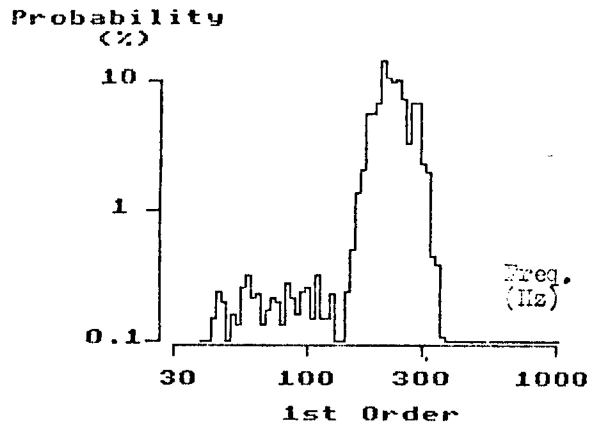


An Fx histogram of a natural setting learner based on the Polish text.

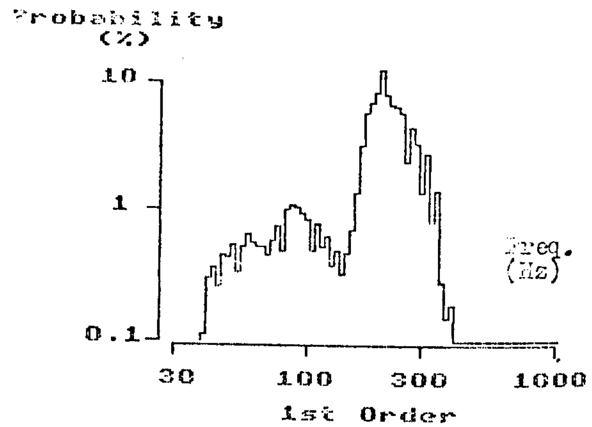


An Fx histogram of the same learner based on the English text.

A bilingual speaker demonstrates a similar tendency. Compare below:



An Fx histogram of a bilingual speaker (born in England) based on the English text.



An Fx histogram of the same speaker based on the Polish text.

4. On the whole, English and Polish turn out to be similar enough to manifest no significant difference between Fx distributions for the native-speakers of the two languages. Particular speakers, however, vary with respect to the amount of irregularity present in their foreign language production. The irregularity itself might be due to natural difficulties faced by the learner in the process of SL acquisition: being limited by his phonatory mechanism he attempts to produce an auditorily acceptable foreign output. There is, therefore, a degree of a conscious control involved in his performance as a recompense for being a non-native speaker. This, together with foreign linguistic structures which themselves require minor (for a Polish-English pair, at least) modification of the native Fx, leads to some deformation of the latter in foreign speech.

It appears that natural setting learners tend to deviate a great deal more from a characteristic to them Fx distribution than formal setting learners do. This suggests that a natural setting learner forces his phonatory mechanism to function appropriately in a foreign tongue without former preparation, which in effect gives a phonetically unsatisfactory output (in terms of segments, segment sequencies and suprasegments) whose only aim is communicativeness. Formal setting learners usually exercise their vocal folds, together with the whole articulatory structure, to let them gradually accommodate to the new, foreign language requirements. Consequently, their Lx patterning remains relatively stable.

The above distinction between formal vs. natural setting learners is confirmed by other experiments conducted by the author which all point out to: firstly, different mechanisms employed by those learners in the process of the acquisition of SL phonology; secondly, formal setting learners being, in general, nearer to success in producing a phonetically acceptable foreign speech.

Naturally, an observation about the lack of statistically significant difference between Polish and English fundamental frequency distributions which would be attributable to language does not presuppose the same state of affairs for any given pair of languages. For instance, one could expect differences in Lx patterning between languages which are typologically distinct, like tone vs. stress languages, especially when the speakers are also anthropologically differentiated. This, however, remains to be investigated.

*I would like to express my gratitude to all speakers who kindly agreed to serve as subjects in the experiment and to the staff of the Department of Phonetics at UCL for supplying me with a necessary equipment.

REFERENCES

1. Borden, G.J., Harris, K.S. 1984. Speech Science Primer. London: Williams and Wilkins.
2. Butler, C. 1985. Statistics in Linguistics. Oxford: Basil Blackwell.
3. Abberton, E., Fourcin, A. 1971. "First applications of a new laryngograph." Medical and Biological Illustration 21, 3. 172-182.
4. Abberton, E., Fourcin, A. 1972. "Laryngographic analysis and intonation." The British Journal of Disorders of Communication 7, 1. 24-29.
5. Abberton, E., Fourcin, A. 1976. "The laryngograph and the voiscope in speech therapy." XVIth International Congress of Logopedics and Phoniatrics, Interlaken 1974. 116-122.
6. Abberton, E., Fourcin, A. 1984. "Electrolaryngography." In: Code, Ch., Ball, M. eds. Experimental Clinical Phonetics. London: Croom Helm. 62-78.
7. Fasold, R. 1985. The Sociolinguistics of Society. Oxford: Basil Blackwell.
8. Fourcin, A. 1974. "Regularity of vocal fold vibration." 8th Internat. Congress on Acoustics. London 1974.
9. Fourcin, A. 1974. "Laryngographic examination of vocal fold vibration." In: Wyke, B. ed. 1974. Ventilatory and Phonatory Control. OUP. 315-333.
10. Fourcin, A. 1981. "Laryngographic assessment of phonatory function." Proceedings of the conference on the assessment of vocal pathology. Maryland 1979. ASHA Reports 11. 116-127.

ON THE PERCEPTUAL EQUIVALENCE BETWEEN JAPANESE AND SPANISH SOUNDS

SATORU AIBA

Dept. of Behavioral Science
Hokkaido University
Sapporo, Japan

MIGUELINA GUIRAO

Laboratorio de
Investigaciones Sensoriales
CONICET
Universidad de Buenos Aires
Buenos Aires, Argentina

JUN-ICHI ABE

Dept. of Behavioral Science
Hokkaido University
Sapporo, Japan

ABSTRACT

Identification tests were performed with Japanese listeners using Spanish sounds consisting of V, CV, CVC, CCV syllables and CVCV words. All five vowels were correctly identified. Nine of the seventeen consonantal phonemes /m,n,ɲ,b,g,s,f,ʒ,ʧ/ reached over 80% accuracy. One /x/ scored 75% (taken for /f/ 23%). Liquids /l/ and /r/ were mutually confused with /l/, being twice more recognized as /r/ than vice versa. Unvoiced stops (average correct 40%) were changed for their voiced counterparts, and voiced /d/ (correct 50%) was displaced toward /r/ and /θ/. In 20% of the responses syllables CCV were transcribed as CVCV. When unvoiced stop-vowel syllables were in the second position in CVCV words, all responses rated nearly 100%. For another group of listeners /d/ resulted in 100% accuracy and unvoiced stops rated between 80 and 90%. Both Japanese and Spanish listeners seem equally good in identifying sounds of the other language, while misidentified phonemes are different for the two languages.

INTRODUCTION

Working with Spanish speaking listeners, we have previously presented some evidence on the perceptual similarity between Japanese and Spanish sounds (Guirao. M., 1978; Guirao, M. and Luis, C. R., 1982). In the experiments described here the Spanish speech material was presented to Japanese listeners.

PROCEDURE

Two speakers, both native of Argentina, recorded isolated syllables and words. Speech sounds consisted of the following syllabic types : a) five vowels /i,e,a,o,u/, b) eighty five CV combining seventeen consonants with each one of the five vowels, c) seven CVC, starting with /p,k,m,s,f/ and ending with /m,n,s,l,r/, d) twelve CCV pairing the six stops and /f/ with /l,r/ as in /pla/, /pra/,/fla/ and /fra/. Eleven words, formed by stop V - stop V combinations, as in /dote/, /dike/, /codo/, were also included.

The material was presented individually to ten students of Hokkaido University, Sapporo, Japan, who were instructed to listen and to write down the sounds in kana characters and in romanized letters.

RESULT

Vowels /i,e,a,o/ were 100% identified. Vowel /u/ resulted less familiar, being 60% correctly reproduced and written ou by the rest of the cases. For the CV syllabic types, nasals /m,n,ɲ/, voiced stops /b,g/ and fricatives /s,f,ʒ,ʧ/ were identified over 80% of the cases. Sound /x/ rated 75%, was perceived as /f/ 23%. Unvoiced stops /p,t,k/ scored 40% being shifted toward their voiced counterparts /b,d,g/. In turn, voiced /d/ was misidentified for /r/ 17%, /b/ 9% and /l/ 6%.

Liquid /l/ and /r/ rated 42% and 62% respectively. Sound /l/ was taken for /r/ 42% and /r/ for /l/ 29% of the cases. We tested these two sounds in CVC and CCV syllables. At the end of CVC the identification of /l/ improved to about 80% but /r/ remained close to 50%. When in the second position of syllables CCV, /l/ was confused with /r/ twice as much as /r/ for /l/.

Unvoiced stops were also tested at the onset of CV and CCV syllables and were misidentified for voiced stops. Moreover some of the subjects transcribed CCV type as CVCV, e.g gara instead of gra.

Unvoiced stops and /d/ were presented again in CVCV words. In this case /t/ and /d/ reached 100%. Recognition of /p/ and /k/ improved to 83 and, 90% respectively in the initial word position and to 93 and 100% when located in the second syllable.

An extra experiment was run presenting unvoiced stops and /d/ in CV and CVCV combinations to listeners trained in phonetics. This time /d/ scored 100% and the other stops gave about 80% for /p/, 86% for /t/ and 90% for /k/.

FINAL REMARKS

It is observed that in general Japanese listeners gave equally good performance as the Spanish listeners in recognizing each others speech sounds. It is also noted that when sounds are confused, tendencies are different for the two language groups.

As it was anticipated (Guirao, M. and Luis, C.R., 1982) vowel /u/ does not seem to have as close correspondence in both vocalic systems as the other four.

Among the periodic non vocalic sounds, the Spanish speaking group converted most of the Japanese sound /r/ into /l/ and Japanese listeners showed the opposite tendency making more bias in Spanish /l/ toward /r/.

With respect to fricative sounds (bands of noise) consonant /f/ was somewhat changed for /x/ by Spanish listeners and /x/ for /f/ by Japanese listeners. While the Spanish participants found it more difficult to label some fricative sounds such as /ʒ/ /f/ /z/ /ts/ and /ʃ/, the last three non-existent in Spanish, Japanese speakers could not easily recognize the unvoiced sounds (bursts) /p,t,k/. In both listeners confusions were between sounds of the same acoustic group.

REFERENCES

(1) Guirao, M. "Similarity between Japanese and Spanish sounds", The Study of Sounds <u>18</u> (Proceedings of the 3rd World Congress of Phoneticians, Tokyo, Japan, 1976), 211-216, 1978.

(2) Guirao, M. and Luis, C.R. "Identification of Japanese syllables by Spanish speaking listeners", Journal of Acoustic Society of Japan. (E), 3, No.1, 21-26, 1982.

A CONTRASTIVE PHONETIC STUDY OF JAPANESE AND DUTCH

TJEERD DE GRAAF

Institute of Phonetic Sciences
Groningen State University
The Netherlands

ABSTRACT

As part of a research project on the pronunciation of Dutch by foreigners (representing more than 20 languages), we have paid special attention to the contrast between Japanese and Dutch. This contrast forms a factor to explain the pronunciation of Dutch by Japanese speakers. For a period of more than 200 years, Dutch was the only European language known in Japan and during that time this language had an important influence on the Japanese culture. Many loanwords of Dutch origin were introduced, for which the pronunciation and writing in Japanese graphemes were adapted to the Japanese sound system. In our study the phonetic properties of these loanwords are related to the results of the present contrastive phonetic investigation.

1.INTRODUCTION

During the last decades, large groups of foreigners have settled in the Netherlands. Many of them use Dutch with a very strong accent, which makes communication difficult and can be felt as a strong social handicap. Dutch language instruction to foreigners is often provided for heterogeneous groups, where most of the attention is paid to grammatical and lexical aspects of the Dutch language. In many cases, the pronunciation errors (which may vary considerably within the group) are not taken into account sufficiently.

In order to analyse these problems for various groups of non-native speakers, we followed the example of Gårding and Bannert (1980) in setting up foreign accent archives : recordings of Dutch spoken by representatives of 23 language groups. These recordings consist of a text read aloud, isolated words and spontaneous speech (for more than 80 subjects). For separate languages, whose phonology differs from Dutch in an interesting way, a further analysis is made and typical errors are collected and classified. In this classification we distinguish vowels, consonants prosodic properties and further aspects.

2.DUTCH SPOKEN BY JAPANESE SUBJECTS

One of the groups of foreigners we paid special attention to in our project consisted of Japanese speakers of Dutch. This choice was motivated by the large differences between both languages, the historical importance of the Dutch language in Japan and the existing collaboration we have with Japanese institutes. In Figure 1 we present the phoneme systems of Japanese and Dutch, illustrating the differences between both languages.

Figure 1

Japanese vowels and consonants:

i		u	p b	t d			k g
e		o		s z	ʃ ʒ		
	a		m	n			
				r			
			w			j	h

Dutch vowels and consonants:

i	y	u	p b	t d			k g
I e:	ø: ə	o:	f v	s z			x ɣ
ɛ	œ	ɔ	m	n			ŋ
a:		ɑ		r l			
			w			j	h

The most important differences are:

for vowels:

(2.1) the Japanese vowel system, with five vowels only, is much less complicated than the Dutch system

(2.2) in Japanese, the phoneme /u/ is realised as the unrounded high back vowel [ɯ]

(2.3) there are no rounded front vowels in Japanese

for consonants:

(2.4) The fricatives /f/,/v/,/x/ and /ɣ/ do not exist in Japanese

(2.5) /h/ is realized as an unvoiced palatal fricative before /i/ and as a bilabial fricative before /u/

(2.6) /r/ is realized as an alveolar flap; there is no separate phoneme /l/

(2.7) /t/ becomes [tʃ] before /i/ and [ts] before /u/

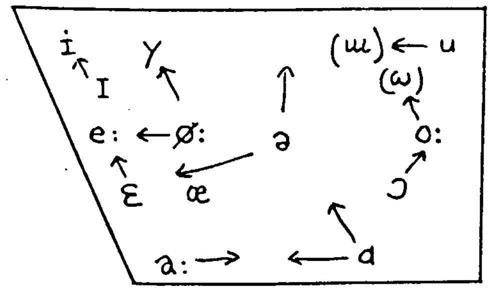
(2.8) /n/ undergoes regressive assimilation of place, e.g. to [m] before /p/,/b/ and /m/. It is the only consonant that can be followed directly by other consonants and is considered as a separate syllable

(2.9) there are no clusters of consonants within a syllable: these have the structure V (only vowel) or CV (consonant-vowel-combination)

3.COMPILATION OF PRONUNCIATION ERRORS

We analysed the sound recordings of Japanese speakers who read the Dutch text, the isolated words and produced spontaneous conversation. First, a phonetic transcription was made of the material and with the aid of this we gave an inventory of the main pronunciation errors for each speaker. Afterwards, their common errors were summarized in a compilation that can be considered as a typical pattern representative of the speakers of Japanese. For the vowels, we illustrate this in a vowel diagram with the Dutch vowels, where the typical deviations of the vowel pronunciation are indicated by arrows.

Figure 2. The pronunciation of Dutch vowels by Japanese speakers



The following differences can be observed:

for the vowels:

(3.1) the pronunciation is in general higher and more fronted

(3.2) short vowels are too long in general

(3.3) the vowel /I/ is pronounced like [i]

(3.4) lip rounding is not realized in the pronunciation of /u/

(3.5) the Dutch diphthongs are produced as monophthongs, e.g. /œy/ becomes [ø:] or [ɛ:]

for the consonants:

(3.6) aspiration of initial voiceless plosives takes place

(3.7) /h/ is realized as a palatal or velar fricative

(3.8) /f/ or /v/ are produced as a bilabial or as a glottal fricative

(3.9) /r/ and /l/ are realized in the same way: namely as an alveolar flap; they can be ommitted in final position

(3.10) /x/ or /ɣ/ can be pronounced as a velar plosive [k]

(3.11) the velar nasal becomes [ŋg]

for clusters of consonants:

(3.12) vowel insertion, like in [sətura:t] for 'straat' (Engl.'street') takes place

(3.13) deletion of consonants, like [sətɛ:k] for 'sterk' (Engl.'strong') takes place

(3.14) at the end of words a vowel is added, like in [tɛkisuto] for 'tekst'

The above changes can be explained by the properties (2.1)....(2.9) of the Japanese phoneme system as compared to the Dutch system (Section 2)

4. DUTCH LOANWORDS IN JAPANESE

For a long period in the history of Japan, the Dutch were the only Europeans who could visit the country. Due to the policy of seclusion from the outside world by the Tokugawa government, this isolation started in 1639 and it lasted until the second half of the 19th century (Vos,1963 and 1978). The Dutch representatives imported the European culture from their settlement on Deshima and the Dutch language was used by Japanese interpreters. A special field of Dutch studies (Rangaku) developed and Japanese publications like 'Rangaku kaitei' ('Guide to Dutch Learning') by Otsuki Gentaku (1783) appeared, providing a Dutch vocabulary and a concise Dutch grammar.

Figure 3. Fragment from the 'Rangaku kaitei'

In Figure 3, we reproduce a section from this book, where the Dutch sentences are translated into the kanji (the logographic writing with Chinese characters) and into the katakana (one of the syllabic writing systems where the sounds of the syllables are reproduced). The katakana represent the possible pronunciation of the Dutch words by the Japanese speakers, like Dutch [Ik] (Engl.'I') as [iki]. In part of this text, we have compared the katakana writing to the original Dutch words and we also collected a number of loanwords from Dutch in the Japanese language.These occur in particular in the terminology of science (medicine) and ship building, but also quite general new words have been introduced into Japanese through their Dutch equivalent. According to Vos (1978), more than 160 words of Dutch origin are still in daily use in Japanese.

5. COMPARISON OF PHONETIC DIFFERENCES

In order to demonstrate the phonetic changes, we quote the following examples of Dutch loanwords with their Dutch original and English translation in brackets.

1.biiru,'bier'(beer).The English word 'beer' is found in biiya-hooru(beer-hall).

2.bisuketto,'beschuit'(rusk).

3.garasu,'glas'(plate glass).

4.gasu,'gas'

5.inki,'inkt'(ink).

6.karan,'kraan'(tap).

7.karuku,'kalk'(lime).

8.koohi,'koffie'(coffee).

9.kokku,'kok'(cook).

10.madorosu,'matroos'(sailor).

11.masuto,'mast'(mast).

12.mesu, mesu(knife).

13.orugooru,'orgel'(organ).

14.pisutoru,'pistool'(pistol).

15.ponpu,'pomp'(pump).

16.porudaa,'polder'(polder).

17.ranpu,'lamp'(lamp).

18.ransetto,'lancet'(lancet).

19.renzu,'lens'(lens).

20.sukoppu,'schop'(shovel).

In these words, we can observe the following changes when comparing the Dutch originals and relating their deviations to the rules (3.1)...(3.14) following from our contrastive study:

(5.1) final consonants are followed by a vowel, cf. examples 3,5,11; see rule (3.14)

(5.2) vowels are inserted in consonant clusters, cf. examples 10,13,14; see rule (3.12)

(5.3) final -ər and -əl become -V (vowel), cf. example 16; see rule (3.9)

(5.4) the alveolar consonant l becomes r, cf. examples 17,18,19; see rule (3.9)

(5.5) diphthongs are replaced by monophthongs, cf. example 2; see rule (3.5)

(5.6) fricatives are realized differently, [f] as [h], [x] as [k], cf. examples 2,20; see rules (3.8) and (3.10)

6. CONCLUSIONS

The structure of loanwords in Japanese deviates considerably from the Dutch originals; this also holds good for the much more numerous and more recent English loanwords (cf. Pierce, 1971, Vos 1963). We could compare the deviations with the present day pronunciation of Dutch by Japanese subjects and find similar properties in our contrastive study. Both can be explained by taking into account the differences between the phoneme systems of Japanese and Dutch.

LITERATURE

E.Gårding och R.Bannert: Praktisk Lingvistik; Institutionen för Lingvistik, Lund (1980) 347

Y.Igarashi: Basic Phonological Problems confronting the Japanese Speaker of English; in: The Teaching of English in Japan; Eichosha Publ.Comp.Ltd.; Tokyo (1978) 346

J.E.Pierce: Culture, Diffusion and Japlish; Linguistics 76(1971)45

G.Sampson: A mixed System: Japanese Writing; in: Writing Systems, Hutchinson, London (1985) 172

F.Vos: Dutch Influences on the Japanese Language; Lingua 12(1963) 341

F.Vos: Dutch Words in Kimono: Dutch Influences on the Japanese Language: in: A Contribution to the Historical Relation between Japan and the Netherlands; The Netherlands Association of Japanese Studies (1978) 41

ЯВЛЕНИЯ САНДХИ И РИТМИЧЕСКАЯ ОРГАНИЗАЦИЯ СИНТАГМЫ В РУССКОМ И БОЛГАРСКОМ ЯЗЫКАХ

ТОТКА СТОЕВА

Первая кафедра русского языка Софийского университета
София, Болгария

На основе аудиторского и экспериментально-фонетического анализа русского и болгарского публицистического текста, в дикторском его прочтении, рассматриваются явления сандхи на уровне фонетического слова и сандхи между отдельными фонетическими словами в составе синтагмы, роль этих явлений в ритмической организации синтагмы обоих языков.

Русский и болгарский относятся к типологически разным структурным типам в группе славянских языков, потому исследование особенностей стыка слов и их роли в ритмической организации синтагмы обоих языков представляет большой интерес.

Учитывая тот факт, что исходной и основной единицами звучащей речи являются фонетическое слово и синтагма, мы рассматриваем явления сандхи и ритмику речи на уровне этих единиц, а именно: сандхи между клитикой и знаменательным словом в рамках единой акцентно-ритмической структуры - фонетического слова - и сандхи между отдельными словами в составе синтагмы.

Наблюдения и анализ явлений осуществлялись на материале публицистического текста и его перевода на болгарский язык в дикторском прочтении. Тексты были начитаны на магнитную ленту тремя болгарскими дикторами, являющимися носителями нормативного литературного произношения современного болгарского языка, и тремя русскими - носителями московской произносительной нормы, в полном стиле произношения. Аудиторский анализ каждого текста выполнялся тремя аудиторами по напечатанному тексту с целевой установкой на указание границ фонетических слов, синтагм и фраз. Данные в процентах усреднены от трех дикторов. В целях выборочного экспериментально-фонетического анализа речи, магнитофонная запись воспроизводилась и регистрировалась на бумаге с помощью установки, включающей осциллограмму, выделитель частоты основного тона, выделитель интенсивности.

I. Перед тем как рассмотреть явления сандхи на стыке клитики и знаменательного слова, необходимо отметить, что в болгарском тексте 42.4% фонетических слов - это структуры, объединяющие от двух до четырех словоформ; в русском - этот процент почти в два раза меньше - 22.5%. Из них в болгарском 24% приходятся на структуры, объединяющие предлог с существительным или прилагательным, числительным, местоимением; в русском эта цифра - I0%. Большая разница - I4% в частотности структур, объединяющих предлог и именные формы, естественно, коренится в аналитическом выражении синтаксических отношений имени в болгарском языке, где расширяются функции служебных слов. В результате аналитизма болгарского языка обнаруживается закономерность и в употреблении частиц-актуализаторов - проклитик и энклитик - при глагольных формах, как и употребление глагола-связки СЪМ. Эти данные заставляют нас сделать важный вывод о том, что более частое присутствие клитик в стуктуре болгарского фонетического слова, в сравнении с русским, является существенным фактором для выявления специфики сандхи на стыке клитики и знаменательного слова и роли явлений сандхи в просодической организации потока речи обоих языков.

Для болгарского языка характерно активное участие клитик, оканчивающихся на гласный: предлог НА /самый частотный в качестве клитики и составляющий 30% всех клитик/, частицы ДА, СЕ, глагол-связка СА, Е и др. Широкие гласные [a], [e] в этой позиции способствуют увеличению фонетической самостоятельности клитик в составе единой акцентно-ритмической структуры; увеличивается частотность зияний на стыках, как например, в структурах типа: на‿основни, да‿укрепва, се‿явява, са‿убедително, показателно‿е и др. Очень часто отсутствие связывающего согласного в первом слове способствует отделению его от последующего небольшой паузой; в осциллограммах регистрируется некоторый спад интенсивности первого гласного и прекращение фонации на 20-40 мс перед началом реализации второго гласного. Это явление способствует ослаблению коартикуляции слогов на стыках и реализации более слабой степени редукции гласных в структурах.

Тенденция фонетической выделенности клитик в составе болгарского фонетического слова и ослабление коартикуляции на стыках наблюдается и при простых предлогах-клитиках, выраженных одним согласным звуком В, С. В этих случаях очень часто встречается удваивание и вокализация предлога, не всегда регламентированное нормами орфоэпии, как например, въ[ф]‿епоха, въ[ф]‿крак, със‿времето, със‿[ф]сички и др. По всей вероятности, удваивание и вокализация этих предлогов в норме произношения перед словами, начинающимися с того же самого согласного звука или его парного варианта по глухости/звонкости, мотивировано и вышеизложенными тенденциями явлений сандхи, например, със‿съюза, със‿загубата, въ[ф]‿Варшава, въ[ф]‿филиала и др. Отметим при этом, что звонкий согласный [в] в конце клитики оглушается по общему правилу оглушения звонких шумных согласных в абсолютном исходе слова и перед паузой. Подобная реализация клитик является дополнительным свидетельством фонетической самостоятельности клитик в потоке речи в лингвистическом сознании носителя болгарского языка.

В отличие от болгарского, особенностью русского языка является тот факт, что структуры с клитиками в потоке речи характеризуются слитностью элементов. В некоторых случаях проклитики и энклитики имеют характер подвижных префиксов и суффиксов; они как бы растворяются в структуре и слитная единица становится тождественной слову, например, в‿эпоху, в‿общем, со‿временем, к‿новым, под‿общим, рост‿[ы]х и др.

Необходимо также отметить, что в болгарском языке из общего процента клитик на проклитики приходятся 35.8% и на энклитики - 6.6%. Эта тенденция развития проклитичности в структуре болгарского фонетического слова влияет на увеличение предударных слогов структуры. В результате этого ослабляется начало слова, меняется интенсивная кривая в нем. И если сравнить данные по распределению места ударения в фонетическом слове, то приходится констатировать, что процент ударности на первом слоге в структурах болгарского языка во всех случаях меньше, чем в русском, а именно: двухсложные структуры с ударением на первом слоге в болгарском тексте - 9% против II.6% в русском, трехсложные - в болгарском 6.3% против 7% в русском, четырехсложные - 0.7% против I.5%.

Обобщая вышеизложенное, следует сделать вывод, что в потоке речи частое участие клитик и реализация различных явлений сандхи в структуре фонетического слова болгарского языка способствуют развитию более слабой коартикуляции составных элементов и более слабой степени редукции /полуредукции/ гласных; наоборот - сильная количественная и качественная редукция гласных и слитность произношения в русском развиты за счет меньшей частотности этих явлений в составе фонетического слова. В зависимости от этого определяются и специфические особенности ритмической организации этой единицы звучащей речи в обоих языках, т.е. специфическая для каждого языка реализация параметров, выражающих объединение ударным слогом неударных.

II. Различия в сандхи между отдельными фонетическими словами в составе синтагмы русского и болгарского языков обоснованы структурными закономерностями и функционированием именных флексий обоих языков.

В болгарском языке, в анализируемых нами текстах, 85% всех фонетических слов это структуры с открытыми слогами абсолютного исхода слова; в русском языке эта цифра - 62.5%, на 23.5% меньше. При этом 34% всех открытых слогов абсолютного исхода слова в болгарском языке приходятся на флексии членной формы мужского рода -а, -я, женского -та, среднего -то, множественного числа -те, -та. Членная форма, образуя со словом фонетическое единство, сильно влияет на акцентную структуру болгарского фонетического слова, как и на характер стыка слов в составе синтагмы.

Во-первых, вокалический исход слова дает возможность каждому из слов в синтагме стать фонетически более самостоятельным в потоке речи.

Во-вторых, широкие гласные флексий членной формы [a], [o], [e] являются более длительными и более интенсивными, чем русские в аналогичной позиции; в осциллограммах болгарского материала наблюдается меньшее их сокращение, чем в русском и очень часто регистрируется отделение слова в членной форме от последующего небольшой паузой.

Система русских именных /падежных/ флексий вносит большое разнообразие в их звукосочетаемость, а гласные в них, как правило, находятся в заударной - фонетически особенно уязвимой части слова в русском языке.

В русском тексте 37.5% всех фонетических слов это структуры с закрытым слогом в абсолютном исходе слова; в болгарском языке эта цифра - 15%, на 22.5% меньше. Кроме того, в рамках синтагмы консонантная форма связывания слов в сандхи в потоке речи русского языка часто реализуется так, будто конечная согласная находится внутри одного большого слова, как например, рост‿[ы]х‿мощи, в‿общем‿рынке и др. В болгарском языке при реализации стыкующихся согласных часто наблюдается полное их разделение, например, свят' без‿войни, капиталът' със‿[ф]сички се‿стреми и др.

Излагая выше наблюдения, касающиеся некоторых явлений сандхи на стыке клитики и знаменательного слова, как и сандхи между отдельными словами в составе синтагмы, мы попытались выделить из общей картины такие закономерности на стыке слов, которые окажутся инвариантными по отношению к ритмомелодической организации синтагмы и фразы в целом русского и болгарского языков, а именно:

- большая маркированность открытыми слогами клитик и конца фонетических слов в болгарском языке - важный параметр для появления внутрисловных и внутрисинтагменных пауз и нарушения слитности произношения составных элементов фонетических слов и синтагм; следствием этого в потоке речи является реализация меньшей степени редукции и большая фонетическая выделенность единиц, по сравнению с русским, что придает болгарской речи своеобразную четкость и "отрывистость"; в осциллографической записи наблюдается общая выровненность и "растянутость" формы контура интенсивной кривой и относительное сужение диапазона частоты основного тона;

- в результате действия противоположных тенденций - большей "свернутости" слова и слитности произношения синтагмы - в русской речи находят благодатную почву для реализации более разнообразные мелодические контуры, что придает русской речи характерную ей музыкальность и живость.

VOICE QUALITY JUDGEMENTS AND PHYSIOLOGICAL MEASUREMENTS IN ESOPHAGEAL SPEAKERS WITH AND WITHOUT A GRONINGEN BUTTON.

G.L.J. Nieboer

ENT-Clinic, Voice Res. Lab & Centre for Voice, Sp. and Lang. Dis., Univ. Hosp., and Inst. of Phon. Sc., State Univ, Groningen, The Netherlands.

H.K. Schutte, M.D., Ph.D.

ENT-Clinic, Voice Res. Lab & Centre for Voice, Sp. and Lang. Dis., Univ. Hospital, State University, Groningen The Netherlands.

T. de Graaf, Ph.D.

Institute of Phonetic Sciences, State University, Groningen, The Netherlands.

ABSTRACT.

Physiological measurements have been performed on 26 esophageal speakers, both with and without a Groningen Button. The measured variables are: intra-tracheal, sub- and supra- pseudoglottic pressure, trans-pseudoglottic flow and sound pressure level. Of the same set of speakers, tape recordings were made in view of a perceptual evaluation by a group of 85 judges. The evaluations were done on 13 bipolar semantic scales. The results of both parallel experiments are presented in this contribution, as well as the first results of correlation computations.

1) INTRODUCTION.

The measurement of physiological characteristics of esophageal voice has had a lot of attention during the last years. Part of this interest is due to the development of tracheo- esophageal valve prostheses [1]. Besides the advantages of these prostheses, a few disadvantages emerged too: the need to use one hand to close off the tracheostoma, the need for cleaning and exchanging the device, and the fact that relatively much effort is needed to phonate.

One more circomstance that leads to an interest from our side in this type of speech is the fact that in the Groningen ENT Clinic both injection- and button- esophageal voice are teached as a rule to the laryngectomees: this offers the opportunity to compare both types of speech on a physiological as well as on an evaluative level.

In the experiment reported on here this effort is assessed by measuring simultaneously the intra-tracheal, sub- and supra- pseudoglottic pressure, the trans- pseudoglottic flow, and the resulting sound pressure level. Furthermore, attention is payed to the pressure loss caused by the button, with the simultaneously measured air flow rate. The efficiency of voice production was measured, but it will not be reported on here. In fact, due to the relatively high intra- tracheal and sub- pseudoglottic pressures we encountered, these pressures will say as much about the effort of phonation as the efficiency.

The same speakers were asked to read a number of standard sentences. This speech material was judged by a group of 85 listeners, both naive judges and speech therapy students. The last group happened to consist of 96% female judges. Correlational computations have been made to relate these judgements to the physiological data of the same patients.

2) PHYSIOLOGICAL MEASUREMENTS.

In total, 1357 measurements were done in the phonations of 31 esophageal speakers. Not all variables were measured in every measuring point: during injection- esophageal phonations, we did not register the intra- tracheal pressure. The flow was measured in only 496 of the 1357 cases; this was done because the sound pressure level is influenced by the flow mask. The supra- pseudoglottic pressure data have not been processed so far.

The intra- tracheal pressure was measured with an open catheter, held by the patient himself in the trachea, under the thumb closing off the tracheostoma. The sub- and supra- pseudoglottic pressures were measured by means of micro pressure transducers, mounted on a catheter which has a diameter of 1.65 mm in the 6 cm of it between the two sensors. It was introduced through the nose into the esophagus, about 40 centimeters, and then gently pulled back again during phonation. By monitoring

the signal on a scope, evidence could be attained as to the position of the proximal sensor. When this sensor stops showing up pressure offset during phonation, it means that it is situated in the supra-pseudoglottic pharynx. Minor adjustments are sometimes needed in order to be sure that the distal sensor is situated in the air-filled sub- pseudoglottic room. The simultaneous registration of both sub- and supra- pseudoglottic pressure with high frequency sensors will enable us to investigate the acoustic phenomena occurring just below and above the pseudoglottis.

The spread in the data is quite high (see *Table 1*), especially in the sub- pseudoglottic pressure and the flow. The sound pressure level, on the other hand, has a rather small standard deviation, due to the generally small dynamic potential of these speakers.

When we consider the mean sub- pseudoglottic pressure, flow and SPL of our speakers it becomes clear which variables are able to differentiate between the two groups of injection- and button-esophageal speakers. The sub- pseudoglottic pressure seems to do that quite well. The means differ by 1.4 kPa. Four of the five patients where measurements were done during both types of phonation, showed a higher sub- pseudoglottic pressure (see *Table 2*). On the right hand side, you see the P-values from a comparison of the means with a t-test.

Table 1: Mean values and standard deviations of the physiological parameters; comparison of the button group and the injection group.

MEAN VALUES OF SUB-PSEUDOGLOTTIC PRESSURE, TRANS-PSEUDOGLOTTIC FLOW AND SOUND PRESSURE LEVEL.

physiological variable	all speakers	button group	injection group
Psub (kPa)	3.3	4.1	2.6
(S.D.)	(1.6)	(2.2)	(2.2)
Flow (ml/s)	108	131	82
(S.D.)	(93)	(112)	(62)
SPL (dB(A))	66.6	66.2	67.0
(S.D.)	(9.5)	(9.4)	(9.7)

Table 2: Mean sub- pseudoglottic pressure values of 5 speakers, with the p-level of a t-test on difference of the means (between brackets).

MEAN SUB-PSEUDOGLOTTIC PRESSURE VALUES + T-TEST; 5 SPEAKERS.

speaker	button		injection	p <
spr 4	2.8	<	3.5	.041
spr 9	3.9	>	3.1	.061
spr 11	5.3	>	2.2	.001
spr 24	5.4	>	3.1	.001
spr 34	2.6	>	2.4	.215

These differences ask for a physiological explanation. The question is: what will cause one and the same speaker to sustain two different pressures in order to vibrate one and the same sound source. In the first place the pressure build-up possibilities of the respiratory mechanism are responsible, allthough the pressure in the lungs and that in the sub- pseudoglottic room are not directly 1 to 1 related because the prosthesis is situated in between, and because the sub- pseudoglottic space is lying outside the thorax, hardly affected by the intra-thoracal pressure.

The flow values too differentiate between both types of speech. The mean registered value was 108 ml/s. Again, a significant difference was found between the two groups. Allthough we know from Schutte's data [2] that, for laryngeal voices, mean flow values are not very useful predictors of voice performance, the found differences might possibly be related to another voice variable: voice quality.

SPL did not discriminate between both voice types: both reached about 67 dB at 30 cm.

We measured a rather high in vivo trans-button pressure. These measurements were done without a selection based on the age of the devices. The mean age was about 11 weeks, so more than two and a half months; also, at the time we made our registrations (end 1985), no patients had anti-fungus medication [3]. The high trans-button pressures, with respect to the in vitro values we measured, must be attributed to the deterioration of the devices by fungal growth: it makes the material stiff, resulting in a higher flow resistance [3]. Consequently, research is going on at this moment to reduce the flow opposition of the prostheses.

3) PERCEPTUAL EVALUATION

Speech material of the same speakers was subjected to a perceptual evaluation by 85 listeners.. It was done by scoring one minute speech of each speaker on 13 semantic 7-points scales. The one

Table 2: The set of 7-points semantic scales as used in the perceptual evaluation experiment (with English translation). The scales are adapted from Fagel et al., 1982.

SEMANTIC SCALES FOR VOICE EVALUATION		
1) zwak	— krachtig	(weak — powerful)
2) onvast	— vast	(slack — firm)
3) niet hees	— hees	(not husky — husky)
4) eentonig	— melodieus	(monotonous — melodious)
5) schel	— diep	(shrill — deep)
6) traag	— vlot	(dragging — brisk)
7) hortend	— vloeiend	(jerking — smooth-flowing)
8) dof	— helder	(dull — clear)
9) uitdruk-kingsloos	— expressief	(expressionless — expressive)
10) slecht verstaanbaar	— goed ver-staanbaar	(not intelligible — intelligible)
11) langzaam	— snel	(slow — quick)
12) lelijk	— mooi	(ugly — beautiful)
13) laag	— hoog	(low — high)

minute speech was assembled out of the recorded material of read sentences. All pauses, coughs etc. were carefully cut out. We thought this useful to get right and reliable judgements, without the judges being distracted by all kinds of additional noises. In fact, it was not the noises that we wanted to evaluate, but the voices. The scales were 7-point scales. Of course we included in this experiment those 5 patients who were able to produce both types of esophageal voice.

We performed factor analysis on the scores, and three main factors turned out to be important. As could be expected, the scales 6, 7 and 11 formed one

Tables 3 a, b: Multiple correlations computed with the speaker mean values on the 13 scales and the speaker mean values of three (or five in the case of button- esophageal speakers) parameters from the physiological measurements. *Flow* is the mean flow, *SPL* the mean sound pressure level, *Psub* the mean sub- pseudoglottic pressure, *Ptra* the mean intra-tracheal pressure, and *dPbu* the mean pressure difference over the buttons. The figures in italics point at statistically significant correlations ($p < .10$) as found during the multiple correlation steps. As soon as one or more of the three physiological parameters are in the equation, the other figures in the row point to correlations of the rest-variance. In rows without italic figures, Pearson product moment correlations are printed. The figures *in italics* of the last column (between brackets) are the squared multiple correlation values (with $p < .05$). Multiplied by 100 they give the percentage of the variation on that scale that can be predicted from the physiological measurement values.

Table 3 a: All speakers.

MULTIPLE CORRELATIONS *(Italics: p < .10)* COMPUTED WITH MEAN VALUES; ALL SPEAKERS.

scales	Flow	SPL	Psub	Multiple r (and r^2) *(Italics: p < .05)*
TEMPO factor:				
scale 6	*.46*	*.32*	-.01	*.55 (.30)*
scale 7	*.54*	.08	.11	*.55 (.30)*
scale 11	*.46*	*.33*	-.03	*.54 (.30)*
VOICE QUALITY factor:				
scale 1	*.50*	.14	-.03	*.52 (.27)*
scale 2	*.58*	.28	-.08	*.63 (.40)*
scale 3	-.11	-.02	.15	.19 (.03)
scale 4	*.66*	*.35*	-.13	*.71 (.51)*
scale 8	*.56*	*.47*	-.30	*.71 (.51)*
scale 9	*.66*	*.36*	-.16	*.73 (.53)*
scale 10	*.59*	.29	-.24	*.67 (.45)*
scale 12	*.62*	*.32*	-.29	*.70 (.49)*
PITCH factor:				
scale 5	-.22	*.58*	-.09	*.61 (.38)*
scale 13	*.43*	*.50*	-.11	*.63 (.39)*

Table 3 b: Button-esophageal speakers.

MULTIPLE CORRELATIONS *(Italics: p < .10)* COMPUTED WITH MEAN VALUES; BUTTON-ESOPHAGEAL SPEAKERS.

scales	Flow	SPL	Psub	Ptra	dPbu	Multiple r (and r^2) *(Italics: p < .05)*
TEMPO factor:						
scale 6	.19	*.43*	*.59*	*-.46*	*.47*	*.83 (.69)*
scale 7	.10	.38	.33	*-.53*	*.62*	*.90 (.81)*
scale 11	.12	*.47*	*.45*	*-.49*	*.44*	*.79 (.62)*
VOICE QUALITY factor:						
scale 1	*.55*	*.49*	-.22	.05	.30	*.61 (.37)*
scale 2	.33	*.48*	.38	*.45*	*.61*	*.90 (.80)*
scale 3	-.17	-.37	-.35	-.03	-.13	.53 (.28)
scale 4	*.64*	.40	-.20	.00	.23	*.82 (.67)*
scale 8	*.44*	*.50*	*.54*	.17	.35	*.85 (.73)*
scale 9	*.65*	.37	-.25	-.00	.26	*.82 (.68)*
scale 10	*.57*	.41	-.47	.30	.37	*.86 (.74)*
scale 12	*.57*	*.52*	*-.49*	.25	.46	*.91 (.83)*
PITCH factor:						
scale 5	-.17	-.19	.09	.04	-.03	.32 (.11)
scale 13	.39	.25	-.32	-.06	.19	.56 (.31)

factor which one could call *tempo*. Another large factor can be considered to represent the *voice appreciation;* it comprises the scales 1, 2, 3, 4, 8, 9, 10 and 12. A third factor emerged as a *pitch factor:* scales 5 and 13.

This last factor showed a remarkable thing. As you can see, the scales have been arranged in such a way that the more negative side is on the left and the more positive side on the right. Now in the case of scale 5 it appeared that left and right should be changed. Normally, especially with male voices, the term "deep" would be more positive than "shrill". In the case of our set of esophageal speakers the reverse was the case. When the scale is turned around it correlates well with scale 13: "low - high", indicating that in fact "shrill" was considered a more positive attribute of the usually low pitched esophageal voices than "deep".

4) CORRELATION

Now the most interesting thing of such type of research is of course to correlate the physiological data on the one and the evaluative data on the other hand.

Considering all speakers *(Table 3 a)*, SPL and the "pitch" factor are relatively well correlated. The physiological relationship between these two variables in esophageal speech was established long ago. The mean flow values seem to be correlated with all factors, and especially with scales scoring on voice quality (e.g. scale 4 "melodious", scale 9 "expressive"). The same picture emerges from the correlations of the injection group. In the figures of *Table 3 b* the tendencies are somewhat different. The intra- tracheal and trans- button pressures correlate well with the "tempo" factor, and the "pitch" factor shows no important correlations at all.

One of the most striking things is that in all tables the mean flow shows relatively high correlations with the "quality" and "tempo" scales, and that the sub- pseudoglottic pressure hardly shows any correlation with the scales. We have no explanation as yet for this phenomenon, as one might expect this sub- pseudoglottic pressure to be an important determinant of voice quality; in any case, it is "closer to the voice" than e.g. the intra-tracheal pressure (see *Table 3 b*), but the latter has more to do with quality in the *Tables 3 a* and *b*.

REFERENCES

[1] H.F. Nijdam, A.A. Annyas, H.K. Schutte, H. Leever, «A new voice prosthesis for voice rehabilitation after laryngectomy», Clin. Otorhinolaryngol. 1982, 237: 27-33.

[2] H.K. Schutte, «The efficiency of voice production», Dissertation, Groningen, 1980.

[3] H.F. Mahieu, H.K.F. van Saene, H.J. Rosingh, H.K. Schutte, «Candida vegetations on silicone voice prostheses», Arch. Otolaryngol. Head Neck Surg. 1986, 112: 321-325.

PERCEPTUAL CONSEQUENCES OF EQUALIZING LOUDNESS DIFFERENCES OF VOWELS VARYING IN VOICE QUALITY

PERTTI HURME

Department of Communication, University of Jyväskylä
40100 Jyväskylä, Finland

ABSTRACT

Listening tests were organized in which vowels varying in sound pressure level (SPL) and voice quality were rated. The vowels were produced at 60, 70 and 80 dB; the SPL differences were equalized for the listening tests. The following voice qualities were simulated: normal, nasal, strained, breathy, rough. The results show that original SPL differences are reflected in the perceptual ratings in a voice quality dependent manner.

INTRODUCTION

The preliminary observations reported here are part of a research project on the perceptual, acoustic and clinical properties of normal and dysphonic voice [1]. One of our methods is to study the effect of controlled variations in the acoustic properties of voice on perceptual dimensions. Data from such studies may be useful in the construction of tools for examining normal and dysphonic voices.

The human voice shows much variation. Voices differ in fundamental frequency (related to pitch), sound pressure level (to loudness) and long-term spectral characteristics (to voice quality [2], resulting from long-term laryngeal and supralaryngeal settings) as well as other properties (e.g. temporal). The object of the present study is to evaluate voices varying in sound pressure level (SPL) and spectral characteristics on several perceptual dimensions.

More specifically, this paper investigates the perceptual impressions listeners extract from speech produced by simulating several voice qualities at three SPLs. Impressions of voices are certainly affected by voice quality. They are also influenced by SPL: for instance, voices with high SPL are generally judged as more "carrying" or powerful than those with low SPL. However, does this also hold when SPL differences originally present in the voices are technically eliminated by equalizing the SPLs to a common level? If impressions of voices manipulated this way are affected by the SPL, this must be due to reasons other than absolute SPL, probably the relative spectral properties of the voices. This is possible, as both momentary spectra of individual sounds and long-term spectra of speech are highly affected by sound pressure level [3, cf. also 4]. In general, the fundamental frequency dominates the spectra of sounds or speech with low SPL, whereas in sounds or speech with high SPL higher harmonics (especially in the first formant region) are much more prominent.

PROCEDURE

The vowel [ɑ] was produced by the present author in five simulated voice qualities: normal, nasal, strained (or tense, pressed), breathy (with a strong high-frequency noise component) and rough, cf. [5]. These qualities were clear and extreme. The five qualities were produced at three sound pressure levels: 60, 70 and 80 dB, by means of visual feedback from a decibel-meter when producing the 15 vowel tokens. Care was taken to produce the vowel in the same manner at all SPLs. The vowels were then digitally extracted from the master tape [6] and adjusted in SPL to the same level, i.e. the amplitude of vowels produced at 60 and 70 dB was raised to the level of those produced at 80 dB. Thus, the original loudness differences were leveled out. These vowels were then recorded in pseudorandom order on a test tape.

The vowels, presented at about 80 dB in the relatively small room where the listening tests were conducted, were rated by two groups of students of speech communication and logopedics (N=28) with some experience in assessing voices. The vowels were rated on the dimensions good/poor, "carrying" (powerful, stentorian, tragend)/weak, powerless and pleasant/unpleasant by all listeners. In addition, the vowels were rated by half of the group (N=14) on the dimensions nasal/normal, strained/normal, breathy/normal and rough/normal. Scales of seven points were used in the assessment in the first three dimensions, e.g. extremely pleasant (3), moderately pleasant (2), slightly pleasant (1), neither pleasant nor unpleasant (0), slightly unpleasant (-1), moderately unpleasant (-2), extremely unpleasant (-3) and of four points in the latter four dimensions, e.g. not at all nasal (0), slightly nasal (-1), moderately nasal (-2), extremely nasal (-3). The results of the listening tests were analysed statistically by means of Chi-square tests.

RESULTS AND DISCUSSION

Table 1 shows the results of the listening test for the five voice qualities on the seven perceptual dimensions (as median values, pooled for all subjects and the three SPLs). The median values are on the whole as expected: for instance, vowels produced with nasal voice quality are on the average moderately nasal and those produced with rough voice quality moderately rough. Vowels produced with normal voice quality are rated as moderately good and pleasant, neither breathy, rough nor strained, slightly nasal.

Table 2 shows the median values of the ratings separately in the three SPLs in cases where there was a statistically significant difference in the perceptual ratings. The perceptual ratings are affected by the original SPL level in a different manner in the five voice qualities here investigated. Normal vowels appear to be most susceptible to the effect of SPL. High SPL (in the original signal) of normal vowels is associated with less good, less pleasant, more rough and more strained impressions. Normal vowels with low SPL show no difference to those with medium SPL. Nasal vowels with low SPL are associated with less carrying (weak) impressions, whereas high SPL goes with more carrying (powerful) impressions. Strained vowels originally produced with low SPL were rated as more pleasant than those with higher SPL. Rough and breathy vowels generally give a negative impression on all dimensions; low SPL in rough vowels is associated with less strain.

The results imply that SPL information is important in perceptual studies of voice quality. SPL should either be standardized (for instance by means of visual feedback) when producing the sample to be

Table 1: Median values of the estimations of the stimulus vowels produced with five voice qualities (normal, nasal, strained, breathy, rough) on seven perceptual dimensions (good, carrying, pleasant, nasal, strained, breathy, rough) by 28 Ss; column and row medians are also given.

PERCEPTUAL DIMENSIONS:	STIMULUS VOWELS: NORMAL	NASAL	STRAINED	BREATHY	ROUGH	ALL
GOOD (3/-3)	2	0	-1	-3	-3	-1
CARRYING (3/-3)	1	1	1.5	-2	-2	0
PLEASANT (3/-3)	2	1	-1	-3	-3	-1
NASAL (0/-3)	-1	-2	-1	0	0	-1
STRAINED (0/-3)	0	0	-1	-2	-2	-1
BREATHY (0/-3)	0	0	0	-1	-3	-1
ROUGH (0/-3)	0	0	0	-2	-2	-1
ALL	2	1	1	-3	-3	

Table 2: Median values of the estimations on the perceptual dimensions with a statistically significant difference between stimulus vowels originally produced at 60, 70 and 80 dB.

STIMULUS VOWELS:	PERCEPTUAL DIMENSIONS:	p<	SPL 60	SPL 70	SPL 80
NORMAL	GOOD/POOR (3/-3)	0.01	2	2	1
NORMAL	PLEASANT/UNPLEASANT (3/-3)	0.01	2	2	1
NORMAL	ROUGH (0/-3)	0.01	0	0	-0.5
NORMAL	STRAINED (0/-3)	0.001	0	0	-1
NASAL	CARRYING/WEAK (3/-3)	0.01	-1	1	2
STRAINED	PLEASANT/UNPLEASANT (3/-3)	0.01	1	-1	-1
ROUGH	STRAINED (0/-3)	0.05	-1.5	-2.5	-2.5

evaluated or measured as an independent variable when recording speech for perceptual ratings. This is especially important when assessing dysphonic voices as they show much variation in SPL.

On the whole, the results indicate that extremes in the voice (here represented by SPLs of 60 and 80 dB) tend to be associated with less favourable ratings. However, the results presented here apply only to one speaker. Generalizations are perhaps to some extent valid for the Finnish speech culture as well. The cross-cultural comparison of voice quality (and presumably also the use of SPL) in different speech situations is a challenging but difficult task in view of the immense variation between languages and between cultures [7].

REFERENCES

[1] Hurme, P. & A. Sonninen 1986. Acoustic, perceptual and clinical studies of normal and dysphonic voice. Journal of Phonetics, 14, 489-492.

[2] Laver, J. 1980. The phonetic description of voice quality. Cambridge: University Press.

[3] Fant, G. 1973. Speech sounds and features. Cambridge, Mass.: MIT Press.

[4] Hurme, P. 1980. Auto-monitored speech level and average speech spectrum. Papers in Speech Research (Jyväskylä), 2, 119-127.

[5] Isshiki, N. & Y. Takeuchi 1970. Factor analysis of hoarseness. Studia Phonologica, 5, 37-44.

[6] Karjalainen, M. 1980. A computer-based system for speech and audio signal research. NAS-80 (Turku), 291-294.

[7] Ladefoged, P. & N. Antoñanzas-Barroso 1986. Computer measures of breathy voice quality. Working Papers in Phonetics (UCLA), 61, 79-86.

POST LARYNX RESECTION VOICE ACOUSTIC ANALYSIS

S.L. Taptapova, I.I. Yermakova

Serbsky Institute of General and Forensic Psychiatry
Labor. of Speech Pathology
Moscow, USSR 119839

ABSTRACT

The consonants of 26 patients, who had unergone partial resection of the larynx , were tested by the method of spectral analysis. The results of the test showed that the acoustic characteristics deviate from the norm: lack of high singing formant, the main vocal tone is lowered, the spectrum includes some additional noise components. The course of recuperative phonotherapy, including the exercises for modulation improvement and vocal range expansion allow for the rounding spectrum to coincide with the norm. However, the change of timbre does not significantly influences its endurance. The patients can continue their work with full vocal load.

At the present time the efficiency evaluation of recuperative treatment has to be substantiated by objective values. The most adequate of modern methods of analysis of the physical features of the voice is the method of spectral analysis. The consonant spectrum of a normal voice has been investigated by many authors /1, 2, 3, 4, 5,6,7,8,9,10,11/. Spectrographical analysis with vocal pathology was performed with cases of phonastenia /12,13/, of rhiniphonia /14,15,16,17/, of pseudovoice /18,19/, of functional larynx desease /20/.Any reference to investigations on acoustic vocal structure after larynx resection in the available literature has not been found.

The purpose of the present investigation was to determine objective criteria for evaluate of a restored voice after larynx resection of different modifications in case of tumors by the method of spectral analysis.

The investigation has been carried out according to standard procedure together with V.I.Konarev, the engineer. Spectral analysor F-4325 and tape-recorder "Mayak-205" were used. The subject of analysis were frequency characteristics of the fundamental vocal tone (FO), the first (FI) and the second (FII) formants, the most semantically important and constituting the base of a speech phoneme. There were 26 patients in the age groupe of 21 to 67 (21 men and 5 women). Most of the people were in their working age. 17 of them underwent front-side larynx resection and 9 - chordaectomy. All patients at the time could communicate in a whisper.

The course of phonopaedy consisted of the formation of vocal sound (due to activization of the remainders of vocal organs), introduction of this voice into spontaneous speech and a course of vocal exercises aimed at development of modulations and range expansion. The results of recuperative therapy before investigations were defined as "satisfactory" and "good" by the audio method. "Satisfactory" was understood as sonorous, but low modulated voice, easily growing weak and unable to provide satisfactory communication. "Good" was understood as a sonorous steady voice, that allowed the patients to continue their work without any restrictions of their voice.

Spectral analysis revealed the following.

The main vocal tone varied within the range of 100 to 14o Hz. In the majority of cases (73%) the main tone of male voices varied from 160-25o Hz, of female voices - from 250-400 Hz. The frequency of the main vocal tone coincided with the first formant of either the patients with low modulated, rated "satisfactory" voice, or the patients whose voice had been restored almost to the normal. In the first case the patients had to have an extensive larynx surgery, and the formants of almost all of their consonants coincided infrequency and were below normal. In second case, the formant characteristics of consonants coincided with the established standards which showed the compensating abilities of anatomic and physiological larynx structures.

The first formant of almost all patients was easily analysed. The great majority of its values (64.2%) corresponded to the norm. 19.5% of patients with low harsh voices were characterised by frequency decrease. In 16.3% of cases frequency of the first formant was increased, which was peculiar to all men and women, who before the operation had a high-timbre voice.

The second formant was determined in patients with "good" modulated sonorous voices. Its parameters were determined within the normal range for some exeptions (6.3).

The results of investigations also showed that consonant spectra of patients with harsh voices varied within a wide range of 100 to 1000 Hz. In the regions of high and low frequency there appeared additional noise spectra: single ones below 100 Hz and more often on the frequency close to 10 000 Hz.

After recuperative therapy, when the patients returned to normal life, their spectrum remained wide, and though it preserved its former noise components, there appeared the components within the vocal range, the main tone and formants were distinguishable. The rounding spectrum within the vocal range was approaching normal.

The vocal spectrum of patients with poor speech before vocal exercises could not even roughly match up with the normal vocal spectrum. Only some noise components at different spectrum sections were revealed. After the treatment course with the significant voice improvement the spectrum was narrowed, additional noise frequencies disappeared.

Thus, the acoustic characteristics of consonants in the colloqual speech of patients who experienced larynx resection of different modifications and underwent a course of recuperative treatment, can concide with acoustic characteristics of the norm, and the rounding spectrum can totally coincide with it.

Change of timbre of a voice restored after surgery can be accounted for the absence of a high singing formant in the spectrum, which makes the voice light and clear, as well as for the change of vibration frequency of the vocal source. Patients who had undergone larynx resection the range of 1oo-14o Hz which almost matched the frequency range of a normal voice of 200-500 Hz /6/. Lowering it in some cases to 100, 200 or 160 Hz was also peculiar to patients who had extensive larynx resections, with harsh, low voices.

Absence of correlations of resection modifications and quality of a restored voice confirm that the results of the recuperative treatment depend on the compensatoring abilities of a patient.

The coincidence of the rounding spectrum of the first and second formants which are semantically important and phonetically determining, with a variation of the norm allow to recommend the patients who underwent larynx resection and recuperative treatment to continue their work.

Thus, the spectral analysis of consonants in patients with larynx resection made it possible to provide an objective evaluation of the efficiency of recuperative therapy. Acoustic characteristics of the voice after larynx resection deviate from the norm: lack of high singing formant, the main vocal tone is lowered, the spectrum includes some additional noise components. The course of recuperative phonotherapy, including the exercises for modulation improvement and vocal range expansion allow for the rounding spectrum to coincide with the norm. However, the change of timbre does not significantly influences its endurance. The patients can continue their work with full vocal load.

/1/ R.Potter, G.Copp, H.Green and others, "Visible speech",N.T.Van Nostand,1947
/2/ L.I.Varshavsky, L.A.Chistovich,"Srednie spectri russkih glasnih fonem", Problemi fiziologicheskoj acustiki, v.4, 1959, p.9.
/3/ G.Fant, "Acusticheskaya teoriya recheobrasovaniya", M.,1969.
/4/ D.Flanagan, "Analys, syntes i wosprinyatiye rechi", M., Swyas, 1968.
/5/ L.B.Dmitriev, "Osnowi wokalniy metodiki", M., Prosweshenije, 1968.
/6/ V.P.Morozov, "Biophizicheskiye osnovi vokalnoy rechi", M., 1977.
/7/ B.Cuerin, L.J.Boc, "Etude de lénfluence du couplage acoustique sourse", Phonet., 198°, n°37, p.169.
/8/ G.S.Ramishvili, "Avtomaticheskoye opoznavaniye goworyashchego po golosu", M., j.Radio i swyas, 1981, p224
/9/ V.N.Sorokin, Acusticheskiy J., 1981, t.27, v.3, p. 434-440.
/10/ V.N.Sorokin, "Teoriya recheobrasovaniya", M., Radio i swyas, 1985.
/11/ V.I.Galunov, I.B.Tampel, Acusticheskiy J., 1981, t.27, v.3, p. 5-7.
/12/ T.E.Shamsheva, V.P.Morozov, J. Westnik otorinolaringologii, 1966, n°5.
/13/ L.A.Pokotilenko, J. Westnik otorhinolaringologii, 1986, n°4, p.32-34.
/14/ T.N.Woronzova, "Logopedicheskaya rabota s bolnimi otkritoy rhinolaliey posle uranoplastiki, M., 1966.
/15/ V.E.Ostanin, "K woprosu o funkzionalnoy ozenke lechenija wroshdjonnich rasshelin nyoba metodom spectralnogo analiza rechi", Kalinin, 1969.
/16/ E.Sedlackova, "Ein Beweis der Nasalität mit Hilfe der akustischen Analyse", Fol. Phoniatr., 1973, n°25.
/17/ L.I.Vansovskaya, "Spektralhiye harakteristiki glasnih zwukov w rechi bolnih s otkritiy phinolaliyey", w sb. Sowremenniye problemi phiziologii i patologii golosa i rechi, M., 1979, p. 60-66.
/18/ P.A.Derevshchikiv, S.L.Taptapova, J.Westnik otorinolaringologii, 1971, n°5.
/19/ A.V.Dmitriev,"Spektralnie harakteristiki rechi laringectomirowannih bolnih", w sb. Sowremennie problemi fiziologii i patologii golosa i rechi, M., 1979, p. 70-77.
/2o/ D.Doskov, B.Boyanov, "Obektiven analys govorniya signal na pazienti, stradayushchi ot funkzionalna disfoniya", Wtora nauchna jubileyna sessiya po ushni, nosny i gorleny bolesti, 1984.

DYSPHASIA (SPEECH DISTURBANCES), CAUSED BY HYPOACUSIS
(Phoniatric aspects)

Iraida Krushevskaya, Nina Nadeina

Dept. of Oto-Rhino-Laryngology
Research Institute of Capacity to Work
Minsk, Byelorussia, USSR, 220081

The clinic-social and pedagogic observations testify to the fact, that the absence of the auditory control in the formation of the communicative means is leading to the disturbance of the speech- and voiceformation. Due to the tedious work a hard hearing child may rehabilitate the speech communication and become social adequate member of our society.

People with the defect of hard hearing are regarded as people with damaged hearing in their speech communication, needing special measures for preventing these defects.

The communication abilities limitations, the decline of the working capacity and intellectual development of the people with defect of hard hearing are promoting the infringement of their psychosocial development. The aggravation of the defects of general and linguistic development is determined by the type, degree and time of the beginning of the hearing damage, the individual conditions of life and the social-cultural environment, as well as by the inadequate level of education process at special schools and it demands a particular form of influence, corresponding to the modern stage of society development. The speech-vocal disturbances of people with the defect of hard hearing are serving as risk factors for preserving the full working capacity, their social adequacy and integration into the federation of hearers.

With advance of the scientific-technological progress the need of the national economy in the labour reserves has increased, thus the social and vocational rehabilitation of people with the defect of hard hearing is of special significance.

The study of speech-formation process and its interdependence of the type and degree of hypoacusis is an urgent clinic-social and pedagogic problem.

The clinic-social and pedagogic observations testify to the fact, that the absence of the auditory control in the formation of the communicative means is leading to the disturbance of the speech- and voiceformation.

There are no objective data in iterature allowing to judge about the functional condition of the vocal apparatus of people with defect of hard hearing and its interdependence with the degree and time of the hearing damage in the age groups of the pupils of special schools. The usage of the modern acoustic-physiological methods of investigation and the original methodical approach helped to study the interdependence of the functional condition of the vocal apparatus and the acoustic voice and speech characteristics of the condition of the auditory function in age groups of the pupils of schools for the deaf and weak hearers.

One can notice that the data, obtained as a result of the carried investigations are promoting clarification of the phoniatric and phonetic aspects necessary for the correction of the existing surdopedagogic methods, concrete for the definite age group.

The study of the peculiarities of the speech of the earlier-and later-deaf pupils of special schools (who have earlier had normal speech function) has been carried out at the age of (7-I7). The determination of speech distinction has been carried out by the method of syllabic articulation with subsequent phonetic analysis of mistakes and it has helped to determine, that in the result of hypoacusis is occuring in the deterioration speech distinction percentage, directly depending on the degree and remoteness of the hearing disturbance. Predisposition to the replacement of the soft consonants by hard ones, the devoicing of the voiced consonants is typical. With the increasing degree of hypoacusis the deterioration of vowel distinction process (particularly ɘ,e,ɒ) is observed.

It has been noted, that if within the range of 500 - 3000 Hz the hearing loss constitutes more than 30 dB, speech delay is taking place and its melodics and tone formation are broken. If the hearing loss is more than 60 dB the acoustic ways of normal communication are broken.

When studying the changes of the volume of the resonant cavities in case of phonation of the above said groups of the pupils in comparison with the control group it has been noted, that if hypoacusis has been acquired in the early childhood, the volume of the resonant cavities is not changing, and as a result the articulation is indistinct. The difference in the volume of the resonant cavities during the phonation of high and low sounds is very insignificant, thus the speech is monotonous, deprived of melodiousness ans accents. The later deaf pupils, especially in the control group of children, during the pronunciation of the low sounds, the resonant cavities of the larynx are widening or are deeping, but in case the high sounds they are narrowing. The degree of volume changes of the resonant cavities is directly depending on the hearing acuity and the functional condition of the vocal apparatus.

The duration of the separate sounds and especially the vowels is shortened or lengthened, due to it the disturbance of phythmics is observed already at the age of 5-7. The melodics of speech is sharply changing since the ability to discern the pitch of the acoustic stimulus is being violated.

In case of the full absence of the hearing sensibility (the third and the fourth degree of hypoacusis), the rightness of stresses in the speech is violated. The disturbance of the hearing acuity with respect to the high and low tones are negatively influencing the function of the vocal folds right up to its complete cessation. The disturbances of the functional condition of the vocal apparatus, revealed during electronic laryngostroboscopy, expressed in the motor violations of the neuromuscular apparatus of the larynx of the functional character, the degree of which was not indirect dependence of the laryngoscopy data, but on the contrary it depends on the degree of hypoacusis and the time of its acquisition. So, in case of the first and the second degree of hypoacusis the vibrations of the vocal folds, uneven and asynchronous in amplitude have been revealed, when examining the phonation phases hypokinesia of the vocal folds in case of the phonation has determined their incomplete closure, the presence of displacement of the mucosal membrane on their internal edge. In case of hypoacusis of the third and the fourth degree closure of the vocal folds is strong, that determines the hard attack of the voice.

The pressure of the expiratory air is diminishing with the increase of degree and time of hypoacusis, closure force of the vocal folds is changing, the detonation of voice is observed, falset sounding is becoming evident. The above-stated pathology of the voice formation process is leading to the nodulation on the vocal folds in the vibration centre.

The investigations data of the speech quality at different degrees of hypoacusis testify to the dependence of the speech distinction on the degree of hypoacusis. Predisposition of the replacement of the soft consonants by the hard ones and the devoicing of the voiced consonants, apparent already at the second degree of hypoacusis is typical.

The time of maximum phonation is representing a motley picture and it is in close interdependence on an attack of the sound and correlation of inhalation time to exhalation. With the aggravation of hypoacusis and shortening expiration the time of maximum phonation has been decreasing.

Phonation coefficient of people with the defect of hard hearing is increasing with the hearing impairment.

The speech of people with defect of hard hearing is characterized by disturbances, concerning all the three types of stresses: rhythmical, dynamic and melodious.

The investigations data of the external respiration function are testifying to the fact, that voice disturbances of people with the defect of hard hearing is connected to a considerable extent with disturbance of the phonation breathing. Often during the phonation an inhalation is being used, instead of exhalation, that it is distorting the articulation and is making it impossible, inspiratory phase is being shortened in this position.Coordination between expiratory phase and phonation is violated.

During roentgenoscopy observations the paradoxical function and asymmetry between the right and left half of the diaphragm is being revealed from I2 to I4.

The results of the carried investigations are indicating to the fact, that the phoniatric treatment of hypoacusis should be expressed in the elaboration of the number of the conditioned reflexes: breathing, phonation and articulation.

The results of speech rehabilitation are better, if logophonopedic treatment begins earlier, since in the peripheric department of speech-vocal apparatus, functioning quite satisfactorily at the beginning (up to 4-5), in due course (at I4-I5 years and older) the mechanisms of speech formation are acquiring steady disturbances of phonation breathing, function coordination of the vocal folds, resonant cavities and articulation.

Due to the tedious and purposeful work of the specialists a hard hearing child may rehabilitate the speech communication and become a social adequate member of our society, that will help to expand the volume for the choice of professions during the vocation guidance.

ЭЛЕКТРОАКУСТИЧЕСКАЯ ДИАГНОСТИКА ПАТОФИЗИОЛОГИЧЕСКОЙ ФУНКЦИИ ГОЛОСОВЫХ СВЯЗОК

Иосиф Пешак, Алжбета Пойтелшмидова

Университет им. Палацкого
Оломоуц, ЧССР

Современная практика обследования речи не располагает ни одним из методов обследования, который бы использовал выгодную возможность электроакустической записи и обработки акустического сигнала для диагностики дефектов речи. Имеется в виду то, что акустический сигнал можно подробно анализировать с точки зрения физиологической акустики. Таким образом можно использовать информацию об изменениях инспираторного потока воздуха при разговоре и речи.

Каждый элемент речевого проявления, который является продукцией функций респираторного, фонетического и артикуляционного органов (рис. 1), можно изображать схематически и обратно преобразовывать в эти функции, иными словами, существует очень сложная координированная во времени деятельность органов, участвующих при формировании речевого проявления. Картину функционирования всех связанных во времени составных элементов представляет приведенная на рис. 2 схема замещения произнесения речи. Легкие – респирация – заменены функцией источника, обозначенного отрицательным знаком, так как речь реализуется экспираторным потоком выдыхаемого воздуха. Голосовые связки – фонация – представляют генератор, связанный с очень важной функцией выключателя, применяющегося в момент инициирования речевого акта. Совместное объединение полостей рта и носа прислуховыми замками образует схему артикуляции. Центром слухового органа является перекрытая правой мозговой полусферой большая невербальная область отрицательной обратной связи и перекрытая левой мозговой полусферой малая, вербальная область отрицательной обратной связи.

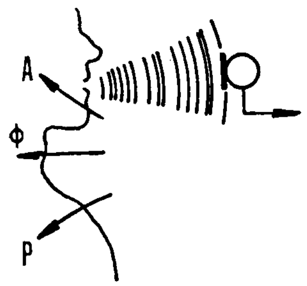


Рис. 1.

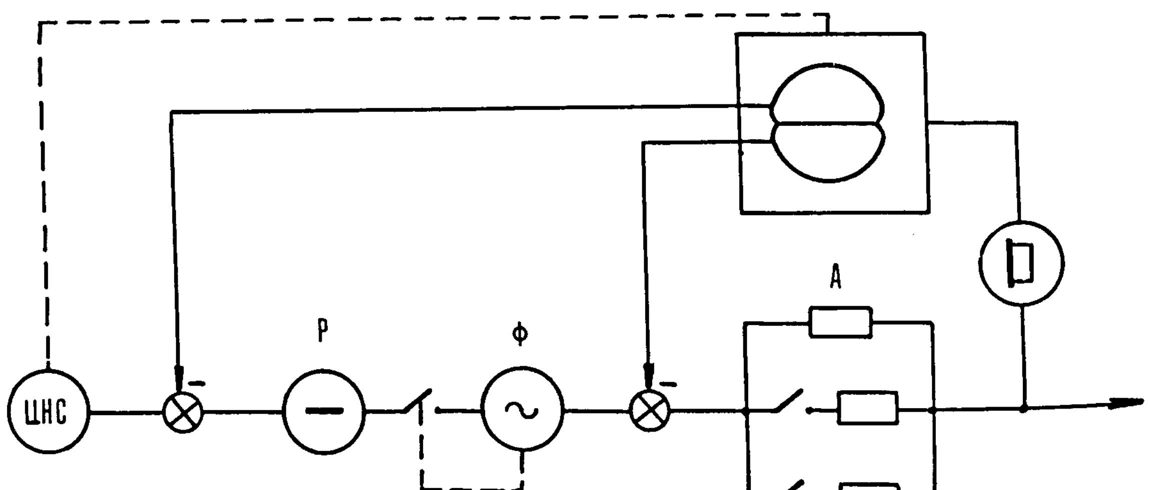


Рис. 2.

Визуализацию функции голосовых связок в момент инициирования фонации можно воспроизвести с помощью осциллоскопа, дополненного высокоскоростной кассетой подобно аппарату для высокоскоростной киносъемки. Таким образом, можно выполнить визуализацию очень быстрых процессов, в том числе – изображения прошлых и непосредственно предшествующих процессов.

Итак, можно обсуждать меру патофизиологической реакции голосовой щели при инициации фонации у балбутиков. Мы выяснили, что в спектре итерационных интервалов существует кроме области, которая у заикающихся детей определяет произношение заикания, и итерационная инфраобласть. Под итерационным интервалом мы понимаем временной интервал между повторяющимися процессами. При инициировании фонации у заикающихся детей зачастую появляются перед началом фонации очень малые колебания с очень малой продолжительностью. Эти итерационные интервалы на порядок меньше, чем интервалы произносимого заикания. Описанное явление, которое схематически показано на рис. 3, мы определили как "вибратио бревис". Частота схематически указанного явления приблизительно соответствует частоте движения губ. Инициирование фонации у индивидов без речевых затруднений зачастую связано с этими очень точными частотами.

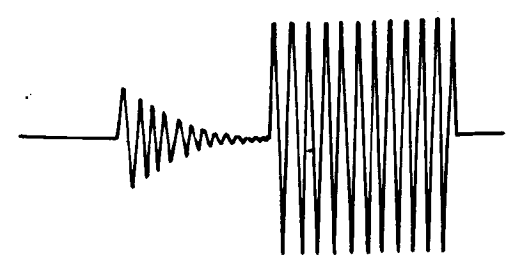


Рис. 3.

Аналогичную аппаратуру можно использовать и для записи звука кашля. Таким образом, можно еще более надежно регистрировать паратипию голосовой щели, чем при обычной фонации. Практически речь идет об обнаружении явления снижения плотности закрытия сомкнутой посередине щели непосредственно при ее пробивании. Физиологически сомкнутая голосовая щель, у которой можно предположить наличие равномерно-распределенного удельного давления при пассивном пробивании быстрым экспираторным потоком воздуха, открывается мгновенно, как видно из записи фонограммы, приведенной на рис. 4.

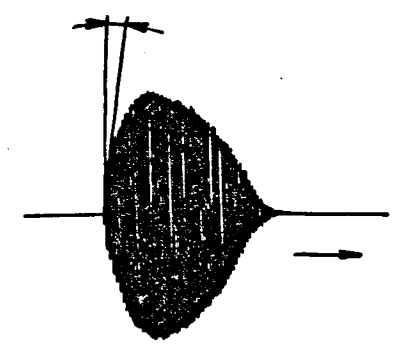


Рис. 4.

Угол отклонения от перпендикуляра в начале набега потока нарастающей амплитуды звука кашля мал. У голосовой щели заикающегося,

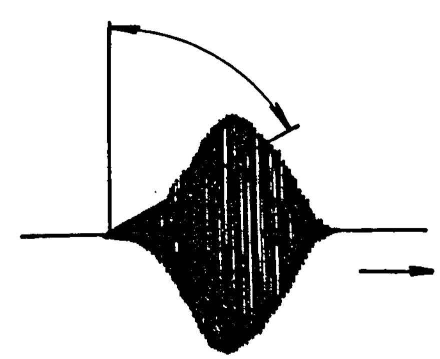


Рис. 5.

состояние которой связано со сниженным удельным давлением и которая посередине крепко сомкнута, имеет место ее постепенное воронкообразное открывание, как видно из рис. 5. Угол отклонения от перпендикуляра в начале набега больше. На рис. 6 обозначены допустимые пределы углов отклонения у заикающегося (1) и контрольного образца 60 детей одного возраста (2).

На основе установленных отличий в положении голосовой щели у заикающихся детей и у детей без речевых дефектов стало возможно высказать гипотезу, в соответствии с которой разговорные способности индивида определяются паратипическими изменениями голосовых связок.

На рис. 7 приведены данные последствий измененного паратипического профиля голосо-

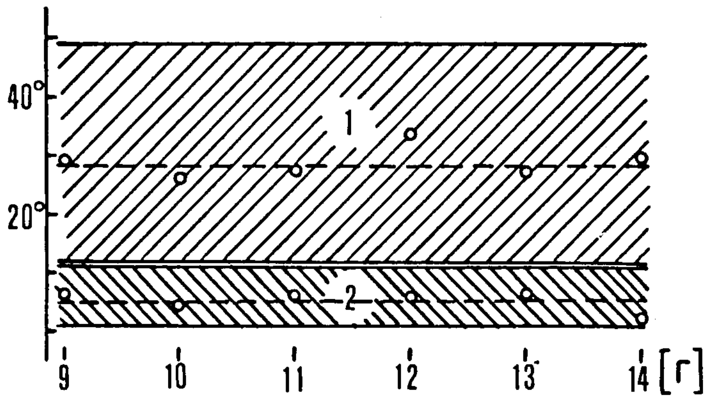


Рис. 6.

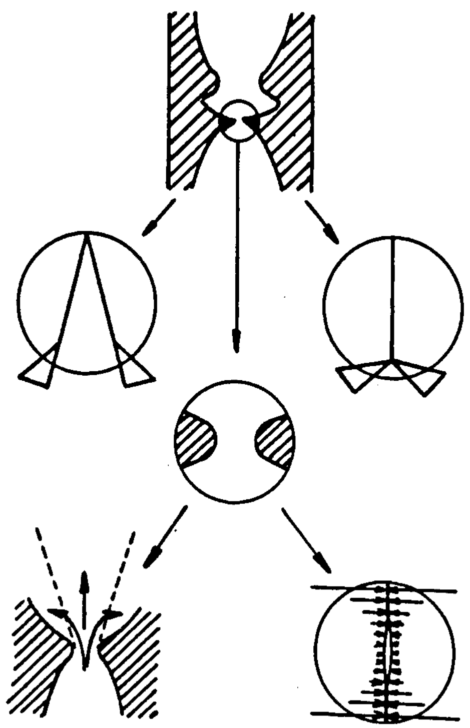


Рис. 7.

вой щели. Часть гортани с голосовыми связками изображена в верхней части рисунка в продольном сечении, перпендикулярном голосовой щели. Обоими стрелками обозначены основные функциональные положения голосовых связок. Влево - при легочной вентиляции, вправо - в момент перед самым пробиванием голосовой щели экспираторной струей воздуха при инициировании речевого акта. В центре рисунка приведена большая часть увеличенного радиуса закругления голосовых связок в месте их соприкосновения. Результаты соприкосновения голосовых связок с паратипичным профилем при инициировании фонации обозначены на рис. 7 в правой части, как мы уже отмечали. Следует обратить внимание на левую часть рис. 7, которая описывает процесс дыхания заикающегося.

Нами проведены измерения и детальный анализ параметров функционального рассмотрения устройства органов дыхания у различных групп заикающихся. Статистические результаты значений объемов легких, их взаимных соотношений и вентиляционные значения можно сравнить с результатами обследования совокупности здоровых детей и несовершеннолетних. Исследовали мы также объемы легких у заикающихся, вентиляцию легких и проходимость дыхательных путей. Мы установили, что как витальная емкость, так и инспираторная емкость существенно понижаются с возрастом при сравнении сопоставленных групп детей. К наиболее достоверным исследованиям следует отнести найденные нами изменения сопротивления потока и особенно последующее за этим изменение проводимости. Существенным является также понижение отношения инспираторной емкости к экспираторному резервному объему. По нашему мнению, очевидно, что наши предположения подтвердились и что найдено достаточно доказательств тому: что описанные изменения сигнализируют об аномальных функциях дыхания у заикающихся и у нормальных людей; что заикающийся не становится заикающимся только при разговоре, а что у него нарушена легочная вентиляция. Мы предполагаем, что приведенные результаты являются параллельным проявлением одной и той же периферийной причины, которая способствует нарушению речевого проявления в результате иного механизма, а ее проявления мы смогли зарегистрировать с помощью электроакустической диагностики патофизиологической функции голосовых связок.

ВЛИЯНИЕ ПАТОФИЗИОЛОГИЧЕСКОГО СОСТОЯНИЯ ГОЛОСОВЫХ СВЯЗОК НА РЕЧЕВУЮ СПОСОБНОСТЬ ИНДИВИДА

Алжбета Пойтелшмидова, Иосиф Пешак

Университет им. Палацкого
Оломоуц, ЧССР

Наше сообщение касается исследования взаимосвязей физиологического функционирования фонетического аппарата - гортани - и речевой способности индивида. Мы полагаем, что функциональная способность голосовых связок может быть оценена различной мерой в зависимости от их паратипического состояния.

Разговорную способность предполагаем необходимой, координированной во времени совокупностью всех органов, которые участвуют при звуковой реализации речи. Это означает координированное действие респираторного, фонетического и артикуляционного органов.

С точки зрения речевой способности главная роль принадлежит собственной реакции голосовой щели в момент инициирования речевого акта. Голосовая щель образована голосовыми сухожилиями, укрепленными на голосовых хрящах, как схематически показано на рис. 1.

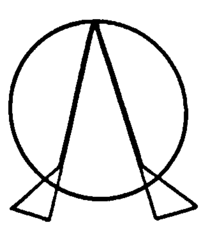


Рис. 1.

Незадолго до речи голосовые сухожилия занимают фонетическое положение, т.е. натягиваются и голосовая щель закрывается

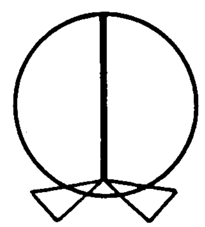


Рис. 2.

(рис. 2). Струя выдыхаемого воздуха, которая обозначена на рис. 3 стрелкой P, про-

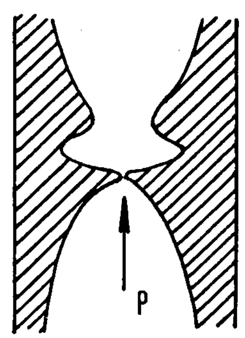


Рис. 3.

рывает голосовую цель в момент инициирования речевого проявления. Последующий периодический поток столба воздуха над голосовыми связками порождает ток, который усиливается в артикуляционных полостях, принимает окраску и приобретает характер человеческой речи.

Предположим, что важной предпосылкой хороших речевых способностей является достаточно тесное соприкосновение обоих сухожилий. Такое соприкосновение должно быть прочным и при этом одновременно обеспечивать готовность старта при фонации непосредственно после раскрывания голосовой щели выдыхаемым потоком воздуха.

Предположим, что отмеченная предпосылка прежде всего зависит от профиля голосовых связок в месте их соприкосновения. Вероятно, что чем меньше будет соприкасающаяся область утонченного профиля, тем проще достигается увеличение удельного давления между закрытыми голосовыми связками при достаточно тонком построении системы обоих голосовых связок (см. рис. 4).

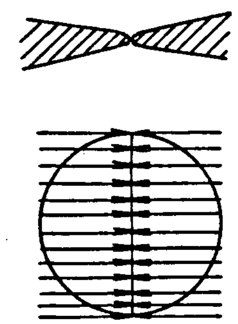


Рис. 4.

Если радиус закругления голосовых связок в месте их соприкосновения будет достаточно большим, то их сопротивляемость в области соприкосновения увеличится и в результате этого удельное давление будет недостаточным и, очевидно, достигнет наиболее неравномерного распределения (см. рис. 5).

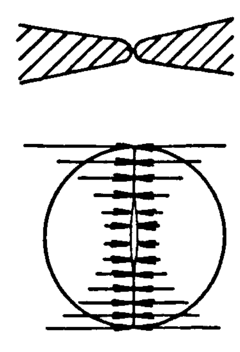


Рис. 5.

Каждый человек очевидно рождается с отмеченным профилем голосовых связок, имеющим большой радиус закругления. Само расположение гортани у новорожденного при имитированном приеме, пищи и дыхании значительно повышено. По мере того, как в процессе дальнейшего развития индивида гортань опускается по направлению к трахее, достигается, очевидно, и уменьшение радиуса закругления профиля голосовых связок, уменьшение их толщины и тем самым более плотное их смыкание, что является предпосылкой для хорошей будущей разговорной способности индивида. Врожденное и затруднительное переключение условных акустико-артикуляционных рефлексов в течение периода дальнейшей дислалии и итерации обусловлено физиологически недостаточным приспособлением речевых органов, вследствие описанных нами причин.

В течение дальнейшего развития радиус закругления профиля голосовых связок уменьшается так, что уменьшенная по толщине голосовая щель под влиянием достаточно большого удельного давления в контакте с тончайшими конвекционными соприкасающимися плоскостями становится достаточно прочной. Ход уменьшения радиуса закругления профиля голосовых связок в месте их соприкосновения под влиянием сопротивления потоку воздуха показан на рис. 6.

До тех пор, пока детская речь развивается спонтанно, требование определенного ограничения понятия разговорных способностей не является актуальным. Дело обстоит иначе в противоположном случае, когда в течение развития индивида радиус профиля голосовых связок не достигает достаточного уменьшения в месте их взаимного соприкосновения. Для такого паратипического закругленного профиля голосовых связок имеет место качественное изменение условий при прохождении экспираторного потока воздуха. В предположении изменения давления и изменения скорости выдыхаемого потока воздуха в области голосовой щели имеет место

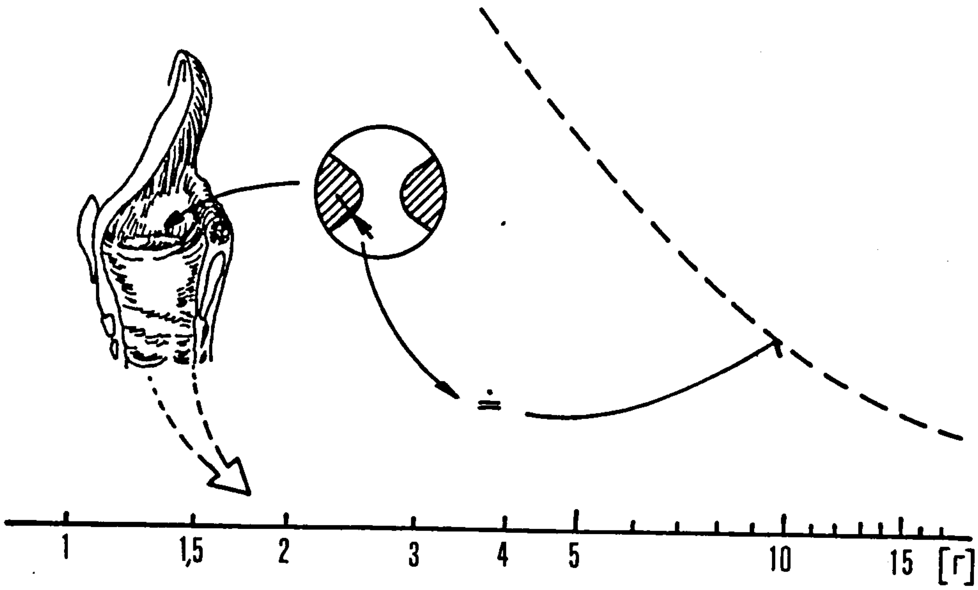


Рис. 6.

увеличение общего потокового сопротивления дыхательных путей.

Качественные изменения первоначального временного стационарного состояния имеют место при поступлении струи к аналогичной среде с постоянным потоковым сопротивлением, когда предельный слой отделяющегося экспираторного потока постоянен; при этом возникает выталкивание массы потока при нестационарном режиме при одновременной потере энергии рассеяния и, как следствие, возникает снижение относительной влажности в области над голосовыми связками и те высыхают. Очевидно при этом возникает нарушение биореологического режима при выходе экспираторного потока воздуха.

Голосовые связки с паратипическим профилем, с большим радиусом закругления вместе их взаимного соприкосновения не могут быть закрыты достаточно тесно при их взаимном сжатии. Пониженное удельное давление в наиболее критичном месте из-за этого не будет удерживать достаточно плотное закрытие, которое бы препятствовало начальному выдыхаемому потоку воздуха перед пробиванием голосовой щели в момент тесного смыкания перед инициированием речевого акта.

Такая неблагоприятная ситуация может быть частично решена непроизвольной коррекцией давления выдыхаемого легкими потока воздуха и дополнительной попыткой повышения упругости соответствующих мускульных групп, участвующих в закрывании голосовой щели. Таким образом, отмеченные попытки часто связаны с различными проявлениями итерационных признаков заикания.

Доказательства предполагаемого утверждения мы провели с помощью ряда несвязанных между собой методов, о которых докладываем в докладе "Электроакустическая диагностика патофизиологической функции голосовых связок". На основе этих доказательств обращаем внимание на существование паратипических изменений речевых органов, которые обуславливают речевую способность индивида и проявляются в своей экстремальной форме заикании.

Своим сообщением констатируем в первую очередь существование приоритета патофизиологии речевого аппарата на отмеченные дефекты речи, обусловленные невротическими симптомами. В настоящее время причины заикания часто противоположны и не дают исчерпывающего объяснения этиологии описанного мучительного дефекта речи - заикания.

ON A STRANGE FASHION OF RUMANIAN ACCENTUATION ABROAD. A CONTRASTIVE SKETCH WITH GERMAN AND ENGLISH

ANA TĂTARU

Heidelberg, FRG

ABSTRACT

Each language has its own, particular ways of accentuation, which must be observed in emphatic speech too. Thus, according to their prosodic specificity, an additional first-syllable accent is frequently allowed in German and English but not in Rumanian emphasized polysyllabic words.

Our paper presents a recent extreme development of an erroneous first-syllable accent in Rumanian words, affecting all prosodic features. It usually occurs in some western broadcasting programs, as a strange, deliberate speech fashion.

1. Introduction

1.1. A language is constantly undergoing change. But this change can hardly affect the specific physiognomy of the language, in general, thanks to its inner conservative tendencies and to the feeling for correct idiom of the native speakers. An excellent example offers us the Rumanian language, which could preserve its Romanic character for almost two thousand years, in spite of having been cut off from the main part of the Romanic world, developing among foreign languages of other groups, e.g., Slavic, Germanic, Ugric etc. There seems no doubt that also in our times of immense change in the life of mankind, languages will keep on maintaining their particular characters, even if this becomes more and more complicated.

1.2. We must admit that our bright technical period, making possible most accurate research in many fields of human activity casts its shadows, by: the constant decrease of respect for and appreciation of everlasting human cultural achievements, deterioration of moral standards and of the sense of harmony and beauty (to consider just what is accepted for music and dance nowadays).

1.3. The positive and negative sides of our era have unavoidably touched the field of language as well: while sophisticated technical devices facilitate minute language research and language teaching/learning, quite shocking negative tendencies -far from being grounded on deeper scientific research, rather disregarding language specificity - are developing in many countries; e.g., strange speech fashions, adoption of unnecessary foreign terms, frequent use of wrong grammatical forms etc. As a rule, they are also far from being the result of natural interference of languages in contact. Most likely, the native speakers sustaining them are similar to the superficial "enfants terribles" of our days, who seek publicity for their extravagant originality at any price.

1.4. Undoubtedly, improper alterations of everyday speech have occurred in the past too, without influencing standard speech at all, as fashions come and go. But about that time mass-media possibilities of spreading such speech fashions were not at hand. Now they are and language misuse can influence the speech of millions of listeners and can become a danger to the specificity of any language.

1.5. Let us concede that we, phoneticians, also deserve unfavourable criticism, as we have not at all done our best for the extension of useful pronunciation knowledge to the masses of language speakers. In general, even school and university handbooks still lack quite important information on the specificity of language, mainly in the field of pronunciation. Phonetics continues to be underestimated in the study and practice of language.

1.6. Under the circumstances, it seems to be high time for linguists and especially for phoneticians, joined in international associations, to unite their efforts, at least, to protect standard speech from deliberate alterations, chiefly in programs to be broadcast for radio or television all over the world. Our paper is meant to draw attention to this apparently overlooked problem of our days.

2. Main Features of German, English and Rumanian Accent in Contrast

2.1. For space restriction, we shall pick out just a few main word/vs. sentence accentual characteristics of these languages.

2.2. Similarities

2.2.1. The accentual pattern of these languages is free, in the sense that the accent is not constantly placed on the same syllable in all words. But it is also fixed, as their accent has a fixed position in the lexical pronunciation of each word; e.g., G.[1*] Vàter, getàn,[2*] E. àbbot, abòve, R. pasàj, pàsăre.

2.2.2. They have a dynamic or intensity stress, but it is usually associated with pitch variation. More often, the stressed syllable is higher in pitch than the unstressed syllables. Thus, the term "accent" can be used to define the correlation of stress and pitch (the same in this paper); e.g., G. Gàttin,[3*] E. wòman, R. càrte

2.2.3. Their accent can have a distinctive function, a trait which increases the importance of learning each word of these languages with its proper accent; e.g., G. zu mächen, zùmachen, E. fòrecast, forecàst, R. còpii, copìi.

2.3. Differences

2.3.1. a) Most G. and E. words have their accent on their initial part - as the Germanic tendency is to place the accent to the beginning, not to the end of words; b) whereas most R. words have their accent on the final part (the penult) - because the Romanic tendency is just the reverse.

2.3.2. a) G. and E. polysyllabic words often have a primary and a secondary accent; b) but this is a mere exception in Rumanian, where words normally contain a single accent. Compare the: G. Schèinwèrfer and E. phòtogràph with the R. fotogràf.

2.3.3. a) Certain G. and E. compounds can have two primary accents (a double accent) but b) this is a rare case in R. compounds; e.g. G. hàarschàrf, E. ìll-advìsed , R. reàua-voìnţă but rea-voìnţă .

2.3.4. a) Although in all three languages form words are generally unaccented in the sentence, the G. and E. prepositions are frequently stressed and get a considerable high pitch. b) Since almost all R. prepositions are unaccented, even in emphatic speech, they are never given the high pitch of main words. Compare: G. Blèibe bèi mir, E. Còme with me, R. Vìno cu mìne.

2.3.5. a) Variation of pitch , length and vowel quality are frequent in G. and E. unstressed syllables. Thus, their unstressed syllables are often shorter or even elided and sometimes undergo a vowel quality change; e.g., E. hè, he is, dìctionary, G. mèine, mèinen. b) But nothing similar happens in Rumanian: it does not have long and short vowels - like G. and E. - and the relative medium length of its vowels is not perceptibly shortened by the absence of stress. Moreover, unaccented vowels are not elided or replaced by other vowels in Standard pronunciation.

3. About Common and Different Means of Emphasis in G., E., R. Standard Speech

3.1. When addressing large audiences, especially over the radio, not only correct but also distinct speech is required to support intelligibility. For that aim, several means of emphasis can be used, provided that they correspond with the prosodic specificity of the given language.

3.2. It is quite incorrect to simply transfer any means of emphasis from one language to another, because some of them occur in many languages - e.g., to pronounce the most important words of the sentence slower, clearly, distinctly and with a stronger stress - while others are usual in one language but unusual in another language. It is thus common in R., but uncommon in G. and E., to pronounce the emphasized words not only slower and more distinctly, but also by loosening the junction of the pretonic syllables without raising their low pitch level; e.g., în-gri-jo-rèază

b) In accordance with their specificity, it is possible in G. and E. to add an emphatic first-syllable secondary accent on certain polysyllabic words; e.g., G. die individuèlle Oppositiòn can be: die ìndividuèlle Òppositiòn, likewise E. the infèriòrity còmplex , can be: . ìnferiòrity còmplex. c) But the same units can be in R. only: opozìţia individuàlă, complèxul de inferioritàte. If emphasized, the beginning of words can be uttered somewhat louder, but without a higher pitch. Otherwise we adopt the main error of the new speech fashion and would say: òpozìţia ìndividuàlă, còmplèxul de ìnferioritàte, what is quite opposed to the particular tune of the R. speech flow. As an exception, a secondary accent is possible in R. standard speech: 1) when emphasizing the contrast between two terms, e.g., Am zis: prèpozìţie nu pòstpozìţie; or 2) on very few prefixes of negation and repetition; e.g., nèînceput but necàz, a rè-cità (to cite again), but a recità (to declaim).

4. On the Development of the Word-initial Erroneous Accent of Rumanian Abroad

4.1. The foregoing contrastive analyses allow us to conclude that it is as bad to impose one's native stress and pitch shapes on the words of a foreign language, as to adopt foreign shapes for the native speech pattern. But it seems to be even worse to deliberately develop the adopted foreign prosodic shapes beyond any boundary, led by an extravagant originality.

4.2. To avoid misinterpretation, it must be underlined that no animosity against persons or broadcasting programs, just respect for my native language determined me to research and present this topic. Moreover, I located the registered examples in time only and the negative examples on the recording tape are confrunted with some positive examples of an older western radio speaker not yet "infected" with that dangerous "virus".

4.3. The new speech-fashion, adopted at present by most R. broadcasting speakers from abroad, has developed approximately during the last six years. Unfortunately it is not only more and more imitated, but - having reached its extreme degrees - it also alters all other R. prosodic features. Let us briefly examine its development:

Phase 1. A couple of western R. radio speakers occasionally began to add a first syllable secondary accent on some long R. words, e.g., còntinuitàţii for continuitàţii.

Phase 2. Meanwhile, the relatively rare emphasis accent becomes more frequent, being adopted by some more of their colleagues. It is now used on shorter words as well; e.g., còtitùra, for cotitùra.

Phase 3. The added secondary accent is intensified and becomes also primary. The words are thus double-stressed, with two similar high pitch levels. A very strange tune in R. indeed! e.g., lìberàli, còmbinà-

ție, stăbilităte, sǒliditâte (II.87).
Phase 4. The unfortunate development reaches its first extreme point, namely, the word initial accent is added even on such words which have their original accent on the second syllable. A sequence of two accented syllables is the result. Quite an absurd situation in R., where even at sentence level the succession of two accents is avoided, e.g., părèrile. But this change determines some other important change: two successive accents cannot be uttered within joined syllables, they must be disjoined, even separated. Thus the meaning of words is altered, e.g. rè|ăcțile(I.87) f. reăcțiile
Phase 5. The strong intensity of the added accent causes the decrease of the original primary accent, which becomes secondary and its pitch is lowered, e.g., cómbativitàte, spériàți, f. combativitâte, speriâți (II,87)
Phase 6. The original accent is dropped out while the added, word-initial accent is maintained. This second extreme point of the development is the cause of the low pitch level of the previously accented syllable; e.g., cónferința, próbleme, dánez (II.87) for conferința, problème, danéz. As a consequence, the intonation contours together with the rhythm are considerably changed. Even vowel change and reduction of syllables are resulting; e.g., soție (so-ți-e) becomes só-ție (6.86), reăcțiile (re-ăc-ți-i-le) becomes rè|ăc-ți-le (II.87).
Phase 7. The accentuation fashion has now reached its third extreme point: because of the tendency to stress any word-initial syllable, even form words, which are almost never stressed, become accented, being given a high pitch level. By this drastic change two wrong ways are open: either to introduce a pause between the accented form word and the following content word - if the accented syllables follow each other - or to maintain the strong accent on the form word and to make the content word unaccented, with a low pitch level, e.g., vór|cóntinuă (6.86) f. vor continuă, ó|realistă for o realistă, lă revedere f. la revedére (II.87).

4.2. Our division into seven phases of development is certainly subjective as their characteristics are co-existing and interchanging. But it helps us to briefly follow the negative influence of this prosodic error on all other prosodic features. Namely, it leads towards a new tune, opposed to the R. specificity. Thus, the relatively smooth speech flow, with harmoniously rising and falling pitch of voice and being more joined than disjoined, turns into a rather abrupt speech with many high pitch levels, sounding sometimes irritated or commanding.

4.3. No use to add that the radio speakers committing these prosodic errors pretend to be or even are Rumanians, who are thus very likely to be imitated by their fan-listeners. As a rule, a native is more often imitated by other natives than a foreigner and his unusual speech is not perceived as mistaken, even if it is so.

Remarks

1.* The terms: German, English, Rumanian occur very often in the text, so we use them abbreviated: G., E., R.

2.* The accent is marked right above the stressed vowel letter.

3.* Signs: ' = primary accent, ` = secondary accent, | = pause, ◡ = junction, ⌐¬ = high pitch, └┘ = low pitch.

Reference

1. Tătaru, A. (1981), "A prosodic contrastive analysis of prepositions in Rumanian, German and English connected speech" in the Study of Sounds, XIX, Tokyo; likewise: Tătaru, A. (1983), Romăna: Pronunțarea (Rumanian Pronunciation), Heidelberg, pp. 128-157.

FOREIGN LANGUAGE VOWEL PERCEPTION AND PERCEPTUAL SYSTEM OF VOWELS

G.N. LEBEDEVA

Dept. of Russian Language
Institute of Chemistry and Technology
Ivanovo, USSR 153460

ABSTRACT

The aim of this work was to study the mechanisms of a foreign language vowel perception by the native speakers of Russian and Spanish, to describe some universal and specific features of perceptual vowel system and new qualities of "phonological ear".

Introduction

There exist three opposite viewpoints on the perceptual abilities of a person. According to one of them, traditional for linguists, a perceptual space is identified with a phonological one. L.V.Scherba thought a person distinguished as many different vowels as there existed phonemes in his language, all other differences were not "in the light point" of his language conscience /1/. The second viewpoint has been formed as the result of psychophysiological investigations of person's perceptual abilities. According to this standpoint the ability to discriminate various classes of sounds (vowels in particular) is universal, a perceptual space, thus, being independent of a particular phonological system /2/.
On the basis of data obtained in phonetic experiments one can formulate a third approach to person's perceptual abilities. According to this viewpoint a person is able to distinguish more sounds than the number of phonemes in his native language system. This ability, however, is also conditioned phonologically /3/.
A description of a perceptual system requires, in our opinion, the solution of the following problems: a) exposure of those features by which the units of a system are discriminated and classified; b) establishment of correspondence between the relevant features of a phonological system and the meaningful features of a perceptual system; c) stratification of perceptual system units (the relation between the units of different levels is obviously most close here); d) description of both the universal features of a perceptual system and the specific ones dependent on a concrete language system.

This paper presents a description of a part of a perceptual system functioning in modern Russian literary language, i.e. the description of foreign language vowel perception mechanisms (by the native speakers of Russian).
Such an investigation would allow us to specify such general concepts as the supposed foreign language vowel identification with the native language phonemes, the unification of "more or less resembling", and non-differentiation of what is indiscriminative in a native language.
"A phonological ear" of the Russian language speakers is formed under the influence of an extremely interesting vowel system: with a comparatively small vowel phoneme inventory there is a tremendous variety of their phonetic realization. This is due to the following two basic reasons: the influence of the neighbouring soft consonants and a considerable reduction in unstressed syllables. The problem of main principles of different sound realizations' perceptual unification into something resembling is of paramount importance for the Russian vowel system. As far as general characteristics of the Russian vowel perceptual system are concerned the following is known: vowels are actually organized in some "space"; the number of discriminated sound units being more than the number of phonemes, and the nature of each concrete sound phonemic interpretation depends on such factors as the length of a phonetic context, the type of a task being solved by identification, the participation of higher language levels. The specific character of "the Russian phonological ear" undoubtedly reveals itself by the analysis of natural vowel identification. The substantiality of investigation of a foreign language vowel perception depends greatly on the fact what language is to be chosen as "foreign" and what in this case is a native one. We examine a perception of English (the British variant) and Spanish (the Cuban variant) vowels by the native speakers of Russian. In our opinion, this is one of the "advantageous" experimental situations, the following circumstances

determining its preference: 1) considerable differences in the number of opposed phonemes in Russian and English, and minimum differences in Russian and Spanish; 2) fairly systematic knowledge of the nature of a native language vowel perception by the Russians; 3) great significance of data about the perception of English and Spanish vowels by the Russians for teaching English and Spanish phonetics.

In the present paper we'll also use the data obtained in groups of Cuban listeners /4/ since "from a linguistic point of view, what distinguishes the speakers of different languages when they perceive the same natural vowels and what can be interpreted as the influence of language phonology on speech activity is of prime importance" /3/.

Let's examine Russian, English and Spanish vowels from the point of view of their acoustic characteristics. Figure 1 presents formant distributions of Russian, English and Spanish vowels used in experiments.

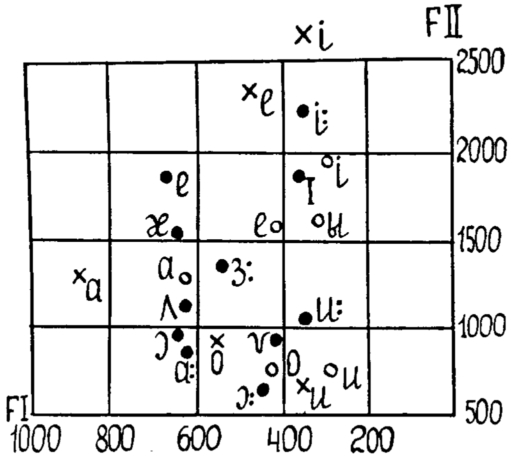


Fig.1 Position of Russian (o), English (•) and Spanish (×) vowels on a formant plane /5,6/.

Experimental Material and Listeners

Tape recordings of Russian, English and Spanish stressed vowels were used as a starting material. The vowels were cut out of the words in which they were pronounced by three male speakers of Russian, English (BE) and Spanish (Cuban variant). The listeners were 36 native speakers of Russian who didn't know either English or Spanish and 20 Cubans who were the beginners of Russian.

Stages of Experiment and Main Results

The first stage of the experiment was pair comparison of English and Russian vowels. The pairs included vowels which could prove to be potentially indiscriminative. Besides pairs including basic vowel allophones, also the pairs containing one of the "soft" Russian vowel allophones were compiled. Thus, each of the English vowels was presented in a pair both with different allophones of one Russian phoneme and with allophones of different Russian phonemes. The listeners had to judge each pair of vowels for perceptual similarity or dissimilarity and to write "plus" if they considered the vowels identical and "minus" if they thought them different.

Let's see how English vowels are placed in a perceptual space of Russian listeners capable of discriminating 18 allophones of the 6 Russian vowel phonemes /7/. The pair comparison revealed the following (see the Table): only English /i:/ is placed in the area of Russian /i/; English /i:/ and /I/ are placed in the area of /ы/; English /e/ - in the area of /e/; English /æ/, /a:/, /ɔ/, /ʌ/ are placed in the area of /a/; English /ɔ/ and /ɔ:/ - in the area of /o/; English /v/, /u:/ and /ɔ:/ are placed in the area of Russian /u/. Within these areas most similarity is found between an English vowel and two Russian vowel allophones -V and V' - what appears to be one more proof of their close proximity (with the exception of Russian /u/, allophones of which are not discriminated from English /u:/).

It's obvious that most differences being noticed between vowels presented in pairs are connected with i- like soundings of Russian 'V and 'V' allophones, therefore, comparison with such allophones is the best situation for perceptual discrimination of Russian and English monophthongs. Comparison of these data with the results of the analogous test carried out in a group of Cuban listeners shows the following: as a whole the Cubans discriminate the same vowels better than the Russians(see the Table). In contrast to Russian listeners they differentiate Russian [ы] and English /I/, Russian [a] and English /ʌ/, Russian [e] and English /e/, Russian [a], [a'] and English /ɔ/, Russian [o] and English /ɔ/. Common features revealed in a pair comparison test are: perceptual confusion of Russian [i] and English /i:/, Russian [ы] and English /i:/, Russian [a] and English /æ/ and /a:/, Russian [o], [o'] and English /ɔ:/, Russian [o] and English /u:/, Russian [u], ['u], ['u'], [u'] and English /u:/, Russian [u] and English /v/ and /ɔ:/. Both the Russians and the Cubans discriminate Russian /i/ from English /I/, allophones of Russian /o/ from English /3:/.

The results of this experiment testify to the fact that even in case of pairly presented vowels the formant characteristics of the latter are far from playing a leading part in vowel discrimination as it could have been expected. Speakers of different languages distinguish the same sounds differently /8/.

In another test the listeners were presented for identification only non-native vowels and were proposed to letter them either by means of their native alphabet or by means of transcription. Several different types of answers turned out to be possible in this test: 1) unanimous identification of an English vowel as one of the native language phoneme, as for instance, /i:/, /a:/, /ʌ/, /u:/, /v/;

TABLE

Position of non-native vowels in a perceptual space of Russian and Cuban listeners (results of 3 Tests: ○- pair comparison; △ - identification; □ - ABX-method)

a) the Russians

Russian / English	i	ы	e	a	o	u
i:	○△□	○□				
I	△	○△	△	△	△	△
e			○△□			
æ			△□	○△□		
a:				○△□		
ʌ				○△□		
ɔ				○△□	○△□	
ɔ:					○△□	○△
u:					○	○△
v						○△
3:			△	△	△	

b) the Cubans

i:	○△□	○				
I		□	△			
e				△		
æ				○△□		
a:				○△□		
ʌ				△□		
ɔ				△□		
ɔ:			△		○△□	○△
u:					○	○△□
v			△		△	○□
3:				△		

2) phonetic interpretation of a vowel corresponding to its articulatory and acoustic qualities and reflecting listeners' ability for a more subtle analysis than phoneme classification, for instance, designation of /e/ as [a'], /u:/ as [°u] or [ju], /ɔ:/ as [o^u], /æ/ as [a-æ], /I/ as [ы-i], /ʌ/ and /ɔ/ as [o-a]; 3) erroneous interpretation of a vowel testifying to the fact that a listener is not able to correlate a perceived sound with one of the Russian phoneme or with any single sound; it corresponds to /3:/, interpreting this vowel the listeners use 20 different signs, and to /I/ where they use 6 signs; 4) refusal to identify a vowel, the main motive being "there's no such vowel in Russian".

Identification of English vowels as Russian phonemes presents additional data for the characteristics of Russian vowel perceptual boundaries. It's evident that the area of Russian /a/ is characterized by the most extensive boundaries, including English /a:/, /ʌ/, /ɔ/, /æ/, partially /3:/. The areas of /i/ and /ы/ are characterized by the most narrow boundaries: only English /i:/ is identified as Russian /i/, and only in isolated instances the realizations of some vowels are classified as /ы/. The areas of Russian /e/, /o/ and /u/ occupy an intermediate position.

Let's examine, for comparison, the identification of the same English vowels by Cuban listeners. Their answers may also be divided in 4 groups: 1) unanimous identification of English vowels as one of the native language phoneme: for instance, /i:/, /I/, /e/, /æ/, /a:/, /ʌ/, /ɔ/, /3:/, /u:/; 2) phonetic interpretation of vowels, for example, identification of English /ɔ:/ as [uo], /v/ as [io], /ɔ/ as [ao]; 3) erroneous interpretation of vowels, for instance, /ɔ:/; 4) refusal to identify a vowel.

In order to extend our knowledge of Russian vowel perceptual boundaries one more test was carried out. Spanish vowels (the Cuban variant) separated from the context were presented for identification to the native speakers of Russian.

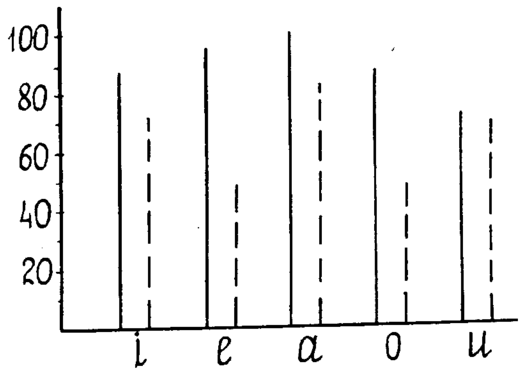


Fig.2 Identification of Spanish vowels by native speakers of Russian (dotted line in a figure) and Spanish (continuous line).

As shown in Figure 2, Spanish /i/, /a/, /u/ are identified by Russian listeners better than Spanish /e/ and /o/. In most cases Spanish /i/, /a/, /u/ are placed in the perceptual space of Russian /i/, /a/, /u/. It should be noticed that the Cubans identify their native /u/ vowel worst of all other vowels. Identification of /e/ and /o/ by Russian listeners is accidental. This fact is manifested in that, on the one hand, they are identified at the same time with several Russian phone-

mes, on the other hand, their phonetic interpretation is extremely various.
For further elucidation of the features of a perceptual vowel system ABX-method tests were carried out. In such experiments the stimuli are presented in triads. The listeners are asked to determine which of the first two vowels (A or B) the third vowel (X) is most like. As A and B stimuli we used only those English vowels which in previous tests were identified with one of the Russian X vowel. The results of this test are of prime interest in two respects: 1) to what extent the correlation of native and non-native vowel depends on the type of a task; 2) what new characteristics of Russian vowels are revealed in this case. Quite a number of facts shows that a perceptual estimation does not depend on the type of a task. Thus, it's revealed that Russian /i/ and /ɯ/ are close to English /i:/ and not to /I/ (it's also obvious from other tests). The listeners consider Russian /a/ vowel similar to English /æ/, /ɔ/, /ʌ/ and /a:/, i.e. extensive boundaries of a vowel area identified with Russian /a/ are also present here. When estimating /e/, /o/, /u/ sounds, the listeners' responses give some new knowledge (see the Table). The ABX-comparison does not reveal similarity between Russian [u] and English /ʊ/, also between Russian [u] and English /u:/ though in previous tests these vowels are identified. Comparing [ʼo] and [ʼoʼ] allophones with English /ɜ:/ the listeners consider them equally alike what is not observed in other tests. The same vowel triads were presented for ABX-comparison to the Cubans (see the Table). In contrast to the Russians the Cubans estimate as most resembling vowels /ɯ/ and /I/ (in a pair comparison test these vowels are also confused; English /I/ is classified as Spanish /e/ in an identification test).

Discussion

The study of foreign language vowel perception is only one of possible methods to obtain data for the description of a perceptual system. The results received are still insufficient for the presentation of this system in terms of quantitative correlations between perceptual and phonological units. However, one can draw quite definite conclusions as far as qualitative characteristics of the system are concerned: a) a perceptual system is more rich than a phonological one. The influence of a native language phonological system on non-native vowel perception is not absolute. The listeners always use the greater number of units than the number of native language vowels. Therefore, the phonology of speech hearing is not only the ability to identify a non-native sound with a native one, but also the ability to understand that it's not a native language sound; b) comparison of vowel perception results with vowel formant characteristics shows that vowel identification is far from being always explained only by their position on a formant plane. This testifies in favour of the fact that distances between the perceptual system units are determined by the properties of a mother tongue; c) comparison of both group results makes it possible to reveal certain universal and specific features of a perceptual system. The universal features are evident in that, first of all, the vowels located in the apexes of the cardinal vowel triangle (i-a-u) appear to be perceptually most "adapted" to this system; secondly, Russian vowel allophones with i-like transitions reveal perceptual independency: both the Russians and the Cubans are not inclined to identify English vowels with Russian 'V or 'V' allophones even in case when close acoustic proximity may be expected. However, this universal perceptibility to i-like transitions of Russian vowels reveals itself rather specifically when speakers of different languages identify Russian "soft" allophones /8/; d) foreign language vowel perception study gives an opportunity to expose those sound features which are alien to the perceptual system of speakers of a given language. Thus, English /I/, /ɜ:/ and partially /ʊ/ do not"go in"the perceptual system of Russian listeners.
The data obtained testify to the complexity of a process providing non-native phonological system vowel perception and to the importance of its further study and comprehension.

REFERENCES

/I/ Щерба Л.В.Русские гласные в качественном и количественном отношении. Л.,1983.
/2/ Чистович Л.А.,Кожевников В.А.Восприятие речи. - В кн.: Вопросы теории и методов исследования восприятия речевых сигналов. Л.,1969.
/3/ Бондарко Л.В. Фонетическое описание языка и фонологическое описание речи. Л., 1981, с.188.
/4/ Бондарко Л.В.,Лебедева Г.Н., Лейва Х. Восприятие гласных неродного языка и механизмы фонологического слуха.-В сб.: Фонетическая интерференция. Иваново, 1985.
/5/ Лебедева Г.Н. Восприятие гласных неродного языка. Канд.дис. Л.,1982.
/6/ Лейва Х. Фонологический слух и восприятие гласных родного и неродного языков
/7/ Бондарко Л.В., Вербицкая Л.А., Зиндер Л.Р., Павлова Л.П. Различаемые звуковые единицы русской речи. - В кн.: Механизмы речеобразования и восприятия сложных звуков. М.-Л.,1966.
/8/ Лебедева Г.Н. Универсальное и специфическое при обработке фонетической информации. В печати.

SOME OBSERVATIONS ON PHONETICAL HISTORY OF MELANESIAN PIDGINS

VLADIMIR I.BELIKOV

Dept.of Linguistics
Institute of Oriental Studies
Academy of Sciences of the USSR
Moscow, ul.Zhdanova, I2

ABSTRACT

The analysis of phonetic correspondences in two Melanesian pidgins (Tok Pisin and Bislama) against the background of English data is undertaken. The phonetic correspondences are shown to be regular for the modern state of the vocabulary, i.e. with all loans considered, and display almost no regularity for the early pidgin stage.

Linguistically Western Melanesia seems to be the most heterogenous part of the world. The total number of languages spoken here reaches nearly a thousand. Bi- and multilingualism was widely spread here, but there had been no linguae francae known on large territories up to the end of the nineteenth century. From that time on the official functions in the area have been hold by the metropolitan European languages: German succeeded by English in New Guinea, English in Papua and British Solomon Islands, English and French in New Hebridies. Yet the natives acquired almost no knowledge of the European languages, the function of interethnic communication media being gained by the English-based pidgins. All those pidgins descend from Bichlamar, a trade jargon spread during the middle of the last century on the Melanesian seashore, in Micronesia and Western Polynesia. In the second half of the nineteenth century Bichlamar arose as the only means of communication between the Melanesian labourers on the European plantations in Queensland, Samoa, Fiji, New Caledonia. Different variants of the jargon stabilized in different parts of the Pacific, thus leading to the resulting divergence.

These stabilized pidgins came into common use as linguae francae due to Melanesians returning home after the completion of their contracts. Beginning from the first decades of our century interethnic marriages resulted in the creolization of the pidgin in New Guinea, and later - in the New Hebridies and in the Solomon Islands. Meanwhile missionaries began applying pidgins in church and at school. Thus the process of lexical enrichment and sophistication of grammatical structure of non-creolized pidgin variants started.

The number of native speakers of the newly formed languages is not great, and up to now they exist chiefly in the forms of expanded pidjins (in the terms of Mühlhäusler /6/), nevertheless in the last decades, being used in press, radio, TV, and fiction, they began to acquire new communicative functions; in the I970's they got the official status.

Now a linguistic family consisting of three closely related Neomelanesian languages has formed, including Tok Pisin (Papua New Guinea), Bislama (Republic of Vanuatu and Neosolomonic or Pijin (Solomon Islands).

The report deals with some aspects of Neomelanesian comparative phonetics and is based on the data of Tok Pisin and Bislama, the languages that have representative dictionaries and a number of available texts.[I]

In the formation of Neomelanesian languages, English segmental forms were applied to the semantic system of Melanesian and underwent the influence of the aboriginal pronouncing habits. Mother tongues of the early pidgin speakers belong to the Oceanic branch of the Austronesian. Phonological structure of the majority of the Oceanic languages is rather simple. Usually a five vowel system is present: i, e, a, o, u. The opposition of voiced - voiceless stops is generally accompanied by the prenasalization of the former; labio-velar p^w and k^w are common; the phonological r/l opposition may be absent. Labials may have stop and fricative allophones p~b~β~v opposed to the sonorant bilabial w. Affricates are rare, fricatives are usually represented by s; h is often non-phonemic. Typical syllable structure is CV(C). Consonant clusters are rare, being usually impossible word-finally.

Comparative linguistics deals with regular phonetic correspondences of the inherent lexicon and interprets the irregularities, yet it is not easy to destinguish the inherent and the borrowed in pidgins and creoles. In spite of the obvious lexical similarity of Neomelanesian languages and of English[2], the latter cannot be regarded as their direct ancestor: "Protoneomelanesian" was an early trade jargon with unstable grammar and a scarce vocabulary of some three to four hundreds of items not necessarily of English descent. For many words of English origin it is difficult to define the exact period of their arising in the trade jargon / stabilized pidgin / expanded pidgin / creole, and to determine, therefore, whether these words can be treated as inherent in any sense. It seems doubtful if words of the German origin adopted at the begining of the stabilized pidgin stage, should be regarded as loans.

Let us first consider English-to-creole sound correspondencies taking into account all the creole lexics of the English origin indiscriminately.

The time-limit and the restrictions of exclusively written sources do not permit to dwell on the question of consonant cluster simplification and vowel epenthesis. Any standards seem hard to be found here, for, on the one hand, Neomelanesian languages exist in the form of different thnolects, and on the other, the degree of proximity to the English models varies greatly depending on the sociolect.[3]

Regular corresponcences are rather trivial and coincide in Tok Pisin and Bislama for the majority of the English phonemes. <u>Vowels</u>. I, i:>i; e, ei>e; æ, ʌ, ɑ:>a; ɒ, ou>o[4]; ʊ, u:>u; eə>ea; iə>ia; uə>ua; difference between the reflexes of the English diphthongs ai, oi, au exists only when written: ai, oi, au in Tok Pisin and ae, oe, ao in Bislama. English ə: is irregularly reflected as o, a, e; in particular items of the basic vocabulary, traceable back to the trade jargon, the reflexes in Tok Pisin and Bislama are identical: doti - toti 'dirty', tanim - tanem 'to turn', gel - kel 'girl', sket - sket 'skirt'. The final ə>a; non-final ə has different reflexes in orthography. Such variability (as in Bislama supos~sipos~sopos ~spos 'if') leads to the supposition that this is just means of coding in written form. In the vowel system the reflexes of the final -ɔ: should be pointed out (non final ɔ: in both languages becomes o). In Bislama final -ɔ:>o~oa (free variants?): sto~stoa 'store', lo~loa 'law'. In Tok Pisin after labial consonants -ɔ:>oa (for some words monophthong variants also exist): moa~mo 'more'; boa 'drill' (<bore), woa 'war'; after non-labial consonants -ɔ:>ua: plua 'floor', sua 'I. sore, 2. shore', stua 'store'. In late borrowings no diphthongization exists: lo 'law'.

<u>Consonants</u>. In Tok Pisin stops generally retain the distinction of non-final voiced and voiceless, final stops being always voiceless. The loss of voicing is registered also in other positions: kalap 'to gallop', dispela~tispela 'this'. In Bislama voiced stops are generally devoiced if not after the nasal. On the contrary, the sequence nasal - voiceless stop may result with the voicing of the latter: rapis 'rubbish', kampani~kambani 'company'. At the same time in the initial position not only can b retain its voiced characteristic, but p also can be voiced: bambu 'bamboo', baenap 'pineapple'[5]. Interdental ð and θ in both languages are reflected in the same way as dental stops. English s, z, ʃ, ʒ are substituted by the Neomelanesian sibilant s. In Tok Pisin tʃ and non-initial dʒ have the same sibilant reflex, initial dʒ retaining its quality. Affricates are preserved in Bislama, varying by voicing, however in the orthography j is chiefly used: fiuja 'future', haejin 'hygiene', safrej 'suffrage', jusum 'choose'. Labial consonants w and v are generally retained, but in Tok Pisin can merge with the resulting bilabial w. The phoneme f is optional in both languages, and can be substituted by the labial stop p. In Bislama substitutions f>v and f>b also occur: tevren 'different', binka 'finger'; there are some cases of hypercorrection as well: fikemap 'to pick up', foes 'voice'. Nasals, r, l, and y are retained in both Neomelanesian languages. The phoneme h in some ethno- and sociolects is optional and can be dropped. Some words in both languages are chiefly used in a hypercorrect form: hai - hai 'eye'. Some of Tok Pisin speakers pronounce initial hu- as wu-: huk ~wuk 'hook'.

Regular correspondences shown are found both in the vocabulary inherited from the "protopidgin" and in the new borrowings. However, in the basic vocabulary of the Neomelanesian languages many instances of other correspondences are found. Some of them are ideosyncratic, cf. æ>ia in giaman - kiaman 'to lie, be false' (<gammon) or ou>a in banara - banara 'bow' (< bow and arrow). But the essential part of the "irregular" correspondences is systematic enough.

These are the most important.

I. I>e, e.g. lewa~leva - leva 'liver', melek - melek 'milk'. 2. ɒ>a, e.g. stap - stap 'leave' (< stop), antap - andap 'above' (< on top). 3. ou>u, e.g. nus - nus 'nose', bun - bun 'bone'. 4. e>a banis - banis 'fence', salim - salem 'to sell'. 5. t, d>r intervocally, e.g. Sarere 'Saturday', kirap - krap 'to get up'. Significantly, in the earliest indigenous vocabulary the frequency of those "regular irregularities" is quite comparable to that of regular correspondences discussed above. Thus, in the Swadesh I00 word list the correspondences ɒ>o and ɒ>a are found twice each: Bislama tok 'dog', long 'long', hat 'warm' (< hot), wanem 'what' (< what name).

In such cases the principles of comparative linguistics presuppose the reconstruction of two distinct phonemes in the parent language, though it is obvious that different reflexes are traced back to the same English phoneme.

The percentage of "irregular" corresponcences in Tok Pisin is higher than that in Bislama, which could be a result of the complete absence of the English normalizing effect on Tok Pisin during stabilization and initial creolization period. Meanwhile, Bislama underwent the stage of regularizing sound corresponcences. Therefore, the following conclusions can be arrived at. In Neomelanesian, the English-based lexicon taken as a whole permits to establish regular phonetic correspondences. However, in the vocabulary arising from the trade jargon the seeming irregularities prove to be systematic. So in early pidgin Bichlamar phonetic correspondences display almost no regularity. It would be desirable to verify these findings on the data of other pidgins.

Notes.

I. Tok Pisin sources: Rev.F.Mihalic's dictionary /5/, texts narrated by speakers of different Tok Pisin variants /4, IO/, a play /7/ and poems /3/. Bislama sources: J.-B.-M.Guy's dictionary /2/, agricultural show booklet /I/, the book on current problems of Vanuatu /8/, poems /9/.

2. The number of coincidences in the Swadesh IOO word list is: Tok Pisin - English - 70, Bislama - English - 77, Tok Pisin - Bislama - 80.

3. Tok Pisin has standard orthography, but there are many deviations in printed sources. Orthographical practice in Bislama abounds in rough anglicisms. Even the text of national hymn in the book edited under the direction of the Vanuatu prime minister W.H.Lini includes the word klat 'to be glad' in two different anglicized forms: glat and glad /8, p.4/.

4. wɒ > wa in both languages: wasim - wasem 'to wash', was - waj 'to watch'. Here and further Tok Pisin word appears the first in a pair, Bislama word - the second.

5. In many Oceanic languages the prenasalization of the initial voiced is very slight or absent altogether. Besides, the phenomenon described can arise due to fricativization b > β.

Bibliography.

I. Akrikaja So, I & 2 Sept. 1978, s.l., s.a.

2. J.-B.-M. Guy. Handbook of Bichelamar, Canberra, 1964.

3. D.C.Laycock. Creative writing in New Guinea Pidgin, - S.A.Wurm (ed.) New Guinea area languages and language study, vol.3, fasc.I, Canberra, 1977.

4. D.C.Laycock. Materials in New Guinea Pidgin (Coastal and Lowlands), Canberra, 1970.

5. F.Mihalic. The Jacaranda dictionary and grammar of Melanesian Pidgin, Milton, 1971.

6. P.Mühlhäusler. Structural expansion and the process of creolization, - A.Valdman, A.Highfield (eds.). Theoretical orientations in creole studies, N.Y., 1980.

7. J.Tokome. Oli kam na paulim yumi, Port Moresby, 1973.

8. Vanuatu. Twenti wan tingting long taem blong independens, Christchurch, 1980.

9. A.Wendt (ed.). Lali: A Pacific anthology, Auckland, 1980.

IO. S.A.Wurm. New Guinea Highlands Pidgin: Course materials, Canberra, 1971.

PHONETIC PECULIARITIES OF THE FAR EAST PIDGIN RUSSIAN

FATIMA YELOYEVA

Dept. of General Linguistics
Leningrad State University
Leningrad, USSR 199036

YELENA PEREKHVALSKAYA

Dept. of Indoeuropean Studies
Institute of Linguistics
Leningrad, USSR 199053

ABSTRACT

While analysing phonetic interference that took place in the process of formation of the Far East Pidgin Russian the following types of interference were discovered: phonemic substitution, underdifferentiation, overdifferentiation of phonemes, differentiation by non-relevant features.

INTRODUCTION

Taking part in the field studies of the Udihe language in the Ussuri region the following fact was discovered. A number of elderly Udihe speakers, obviously unable to express themselves in Russian,use as a second language a Russian-based pidgin. Far East Pidgin Russian (PR), as it appears, was in a wide use in the areas along the Amour-river course and also in the Ussuri region in the end of the XIX - beginning of the XX centuries. Nowadays it is still used by a limited number of speakers from the oldest generation of Udihe, Nanai and other Tungus-Manchu nationalities. The PR is definitely a filiation of the Russian-Chinese trade language formed in the process of Russian-Chinese trade relationships, but used also by other Far East ethnoses (Mongol, Tungus, Manchu etc) while being in the contacts with Russians. The fact that it was also used in their own communication is not excluded.

The basic data for the present report were obtained in the Ussuri region mainly from Udihe native speakers. The PR seems worth being investigated as an interesting case of interaction between typologically dissimilar language systems. It is a Russian-based pidgin, therefore, its vocabulary is mainly of Russian origin, but the latter is realized exclusively by means of the Udihe phonetics. The fact mentioned together with an unusual word order, reduced grammar changed word meaning and native Udihe intonation, makes this language hardly understandible for an unused Russian speaker.

The aim of the present work is to demonstrate the regularity in the substitution rules, functioning while PR speakers pronounce words of Russian origin, to trace back interference processes that took place during the formation of the PR.

The picture is complicated by the secondary influence of Russian, the first language of the young Udihe generation.

CONSONANTS

One of the general principles of consonant functioning in Udihe is the weakening in the intervocalic position. Thus voiced plosives /b/ and /g/ are realized as fricatives [β] and [ɣ] between vowels. In the same way unvoiced plosives /p/ and /k/ that are realized as aspirated in the ini-

tial position, loose this feature between vowels (/k/ also appears as a velar [q] near /a/, /o/).
Table I. presents a comparison of the Udihe (row I) and Russian (row II) consonant systems. The main phonemic variants relevant for the interference processes are in the brackets.

(row I - Udihe) (row II - Russian)		Row	Bi-labial	Labio-dental	Dental	Pala-tal	Velar	Uvu-lar	Pharyn-gal	Glot-tal
Plosives, Non-sonorants	Pure	I	pʰ(p') b(b')		tʰ(t) d		kʰ(k') g(g')	(q)		ʔ
		II	p p' b b'		t t' d d'		k k' g g'			
Plosives, Non-sonorants, Affricates	with one focus	I			(ts) (dz)					
		II			ts					
	with the second middle focus	I			tś dź					
		II								
	with the second back focus	I			(tʃ)(dʒ)					
		II			tʃ					
Nasal sonorants		I	m (m')		n	ɲ	ŋ			
		II	m m'		n	ɲ				
Fricatives, Non-sonorants	with one focus	I	φ (β)			j	(ɣ)	χ	h	
		II		f f' v v'	s s' z z'					
	with the second middle focus	I			ś					
		II								
	with the second back focus	I								
		II			ʃ ʃ' ʒ ʒ'					
Fricatives, Sonorants & Semi-vowels	Semi-vowels	I	w							
		II				j				
	Lateral	I			l					
		II			ɫ l'					
Rolled sonorants		I			(r)					
		II			r r'					

Analysing the interaction of consonant systems in the languages, the following cases of phonetic interference can be pointed out.

I. The substitution of Russian phonemes by the articulatory similar but not fully identical Udihe sounds, e.g. Russian non-palatalized unvoiced plosives are represented as corresponding Udihe consonants: /p/ → [pʰ-], [-p-]; /t/ → [tʰ-], [-t-]; /k/ → [kʰ-], [q](near /a/,/o/), [k](after /s/), cf. [pʰo'to](<потом) 'then', ['pʰapa] (<папа) 'father', [qa'puśta] (<капуста) 'cabbage', [u'pala](<упала) '(he) fall', [qam'pap] (<кампания) 'together with (postposition)', [tʰə'rawa] (< трава) 'grass', [tʰo'qoj](<такой) 'so', ['tʰutə](<тута) 'here', ['qaśa] (< каша) 'gruel', [qo'rotə](<огород) 'garden', [kʰə'də](<когда) 'when', [kʰəp'kən] (<капкан) 'trap', [do'loqo](<далёко) 'far away', [bośtoqə](< Восток) (place name).

Russian palatalized unvoiced plosives /p'/ and /k'/ correspond to phonemic variants functioning in Udihe: [p'ej](<пей '(he) drinks', [k'i'tajśk'i](< китайский) 'Chinese', [kʰu'p'il] (< купила) '(he) bought', [k'i'p'i](< кипи) '(he) boils'. As for palatalized /t'/, Udihe lacks the corresponding sound, thus it is regulary replaced by the affricate /tś/, e.g. [tśur'ma](< тюрьма) 'prison', ['detśi](< дети) 'child'.

2. One of the most characteristic types of interference in the PR is the underdifferentiation of features. Thus, Russian unvoiced fricative /s/ (with one focus) acquires the second middle focus → [ś]: [śᵊ'p'i] (< спи) '(he) sleeps', [qa'śa] (<коса) 'sand bank', [iś'kaj] (< искай) '(he) seeks'. The same rule is detected for the palatalized /s'/ → [ś]: [śi'di] (<сиди) '(he) sits', [śo'r'em'] (<всё время) 'all the time', 'waśa (< Вася) (personal name). The result of the process described is the neutralization of the opposition /s/ - /s'/ in the PR.
At the same time Udihe ś represents Russian unvoiced sibilants [ʃ] and [ʃ'] which are not found in Udihe, e.g. [śej] (<шей) '(he) sews', ['waśa](< ваша) 'you, your (pl.), [uśu'waj](< вышивай) '(he) embroiders', [i'śo](<еще) 'still, yet', ['b'eśi](< вещи) 'things, belongings'. So, four Russian phonemes turn to be represented by a single Udihe phoneme.
The parallel rule can be deduced for Russian voiced [z], [z'] and [ʒ], which are not found in Udihe. The rule of their representation is the following: in the word initial position and after /n/ they are substituted by Udihe affricate /tź/ (with the second middle focus), e.g.: ['dźapᵊtra](< завтра) 'tomorrow', ['dźubᵊrᵊ] (< изюбрь) 'roe', [dźi'ma] (< зима) 'winter', [dźi'na] (< жена) 'wife', ['dźepśip] (<женщина) 'woman', ['pʰandźa] (<фанза) 'house of a Chinese type'.
In the other positions they are represented by the voiced member of the opposition /ś/ - /ź/, that seems to be chracteristic only for the PR, e.g.: [śu'źoj] (< чужой) 'strange', [maga'źin] (<магазин) 'shop', [qol'χozᵊ] (<колхоз) 'kolkhoz'.
Another case of the underdifferentiation presents the way of representation of Russian liquids /l/, /l'/, /r/ and /r'/. Udihe does not distinguish these phonemes, here one finds only one sound of the group in question - lateral /l/ (not velarized as Russian [ɫ]). Thus, Russian word русский was adopted in Udihe as [lu'śa], this word must be considered as a word of Udihe origin. In the PR the same word appears in the form ['ruśku]. PR speakers pronounce both /l/ and /r/ (tapping), but still mix them in the intervocalic position. In this aspect the PR follows the general trend of consonant weakening in the middle position, functioning in Udihe. In the initial position /l/ and /r/ never mix in the PR, while their mixture between vowels leads to hypercorrection, e.g.: ['luqa] (<лук) 'onion', [śᵊli'poj] (< слепой) 'blind', ['riśa] (<рис) 'rice', ['riba] (<рыба) 'fish', ['maliśa] (< малец) 'baby', [p'ir'otᵊ] (< вперед) 'forward', ['b'erəj] (<белый) 'white', ['pʰarqam] (< палкой) 'with a stick', [aś'tara] (< осталась) '(he) remained'.

3. There are also cases of phonemic differentiation by non-relevant features. Thus, voiced labio-dentals /v/ and /v'/ which in Russian form an opposition based on the palatalization feature, are substituted in the PR in the following way: non-palatalized /v/ is replaced by the semi-vowel /w/, while palatalized /v'/ is substituted by palatalized bi-labial plosive [b'] or by its fricative variant [β'] in the intervocalic position. The processes described result in annihilation of the former opposition, e.g.: ['wani] (<Ваня) (personal name), [da'waj] (<давай) 'let', [tʰᵊra'wa] (< трава) 'grass', [dᵊ'wa] (< два) 'two', ['b'idələ] (<видела) '(he) saw', ['b'eśirᵊm] (< вечером) 'in the evening', [dźi'β'i] (<живи) 'life', [aś'taβ'i] (остави) '(he) leaves'.

VOWELS

As for the vocalic system, Russian and Udihe differ greatly. One of the most prominant chracteristics of Udihe is the vowel harmony. E.R.Schneider distinguishes three harmonic series:

I.	a	ae	e	+ i u
II.	o	e	y	
III.		ə		

On the other hand, Russian vocalic system is characterized by the vowel reduction in the unstressed position. Thus, one can state that the vowels [a] and [o] neutralized in the unstressed position in Russian, in Udihe belong to different harmonic series. The interaction of such systems results in the following: the vowel set of a word in the PR depends on the quality of the stressed vowel: in other words, in the PR Russian reduced [ə] can be substituted in the two ways, by [a] or [o] according to the quality of the vowel stressed, e.g.:

[tʰam də'rawa qo'loj] (<там дрова колой) '(he) cleaves wood yonder', [ro'botəj 'noɣo i'wo] (<работай много его) 'he works much', [adi'naqawə] (<одинаковый) 'alike, as (postposition)', [χo'roši] (< хороший) 'good, well'.

Vowels /e/ and /i/ in Udihe belong to the same harmonic series, these sounds form nc opposition based on harmonic rules.

In Russian /e/ and /i/ are neutralized in the unstressed position. One must take into account that The Russian language acquisition was performed in its oral form, therefore [e] is registered inthe PR only in the stressed position; in other positions one invariably finds [i], e.g.: [tši'wo tši'wo] (< чего-чего) 'different', ['deda] (< дед) 'old man, husband', [ɲiši'wo ɲi 'šej i'wo] (< ничего не шей его) '(she) sews nothing', [ɲi u'm'ej] (< не умей) '(he) cannot'.

One of the main characteristics of the Udihe vocalic system is the phonological opposition based on vowel length. The vowel length in the penultimate syllable defines the processes of reduction in the final position , cf. Udihe [bīɲi] -- [bīɲ]; [lǣši] -- [lǣš] ; [ətətə̄m'i] -- [ətətə̄m'] 'I worked' (compare with the present tense [ətətəm'i] 'I work' where the last vowel /i/ is always preserved). So, the reduction happens only if the vowel of the penultimate syllable is long.

In the PR the following facts are registered: /i/ in the word final position is sometimes reduced as if some vowels were perceived as phonologically long. This phenomenon may be treated as overdifferentiation of phonemes, cf. these pronunciation variants: [wāɲi] - [wāɲ] (< Ваня) (personal name), [wāša] - [wāš] 'you, your'. In such cases the non-phonological lengthening of the stressed vowel seems to have been interpreted by PR speakers as the phonological vowel length.

CONCLUSION

As it can be seen, even from the purely phonetic point of view the PR should not be regarded as a mere Russian dialect or as a spontaneously corrupted form of language. Here one evidently deals with a pidgin, an auxiliary means of communication between communities that do not share a common language. It seems essential for the formation of a true pidgin that more than two mutually unintelligible languages in contact are required. This condition is fulfilled in the area, where the main PR data were obtained - some 50-40 years ago in the Ussuri region it was used by Udihe, Nanai and Chinese ethnoses. The study of the PR is relevant for the general interpretation of the pidginization theory, for it is an example of an European-based and in the same time independent creation. No monogenetic theory can be applied to this language and nevertheless it shares many similarities, considered usually as prooves of the monogenetic origin of European-based pidgins. These are elimination of inflections for number, gender and case; identity of adverb and adjective, use of iteration for intensification. All these peculiarities are also chracteristic for the PR, though the latter by no means can be defined as a product of a relexification process. Being "independent" the PR turns to be an example of a "classicle type" pidgin. Material discussed above shows that the PR lexicon presents a wholesale adoption from Russian, while its phonology can be described as a truly "interlinguistic" (it shares definite peculiarities from both languages in contact plus some arisen in the PR itself). Descending to the phonetic level one finds that practically the whole set of the concrete phonemic realizations owes to the vernacular "substratum" language.

These observations seem important for the determination of the contribution of different languages participating in the process of a pidgin formation.

Below we present a bibiography of the Far East Pidgin Russian, that was not included in the J.E.Reinecke's Bibliography:

I. Г.Шухардт. Маймачинское наречие.-Русский филологический вестник. I884, т.ХП, № 4, сс.3I8-320.

2. А.Г.Шпринцын. О русско-китайском диалекте на Дальнем Востоке.- Страны и народы Востока, вып.УI, М.,I968,с.86-I00.

3. И.Ш.Козинский. К вопросу о происхождении кяхтинского (русско-китайского) языка. - Генетические и ареальные связи языков Азии и Африки. Тез.докл.М.,I974,сс.36-38.

4. Ф.Елоева, Е.Перехвальская. К характеристике Дальневосточного контактного языка. - Тез.докл. XXIX сессии ПИАК (PIAC) (Ташкент), т.П, М., I986, сс.54-56.

PROSODY AND CREOLIZATION IN TOK PISIN

NICHOLAS FARACLAS

Language and Literature Dept.
University of Papua New Guinea

ABSTRACT

The results of this sociophonetic study of stress and intonation in Tok Pisin suggest that a new conception of creolization, pidginization, and decreolization as stages on a single continuum, rather than as distinct processes is in order.

INTRODUCTION

Sample

One hour of spontaneous speech was collected from each of 30 members of the Boiken and Olo ethnolinguistic groups living in Wewak town, East Sepik Province, Papua New Guinea. The speakers were chosen by age, sex, years of formal schooling, and whether Tok Pisin was learned as a first or second language to represent a balanced sample of Tok Pisin speakers in the Sepik area. Every rise in pitch was counted over the first 2,000 words transcribed for each speaker, after eliminating the first 500 words of each taping session. High and low pass filters and an oscilloscope were used to measure pitch levels, where necessary.

Pitch Patterns in Tok Pisin

Wurm /1/ observes that for speakers of Tok Pisin from the Eastern Highlands Province of Papua New Guinea: a) stress and intonation account for all pitch patterns; b) affirmative declarative statements normally bear an intonation contour beginning at mid-low pitch, rising to high, then falling slightly over each stressed syllable, falling to a low pitch at the end; and c) stress is signalled mainly by high pitch and is not reduced under declarative intonation contours.

Affirmative declarative statements in the Sepik data collected for this study often bear similar intonation contours to those described by Wurm for the Eastern Highlands, except that most words lose their stress in connected Sepik speech, with a flat intonation contour, beginning at mid-low and gradually falling to low resulting when all of the word stresses are reduced.

ANALYSIS

The total number of pitch rises for each speaker was divided into four environments: a) rises over objects of prepositions (calculated as the percent of the total number of prepositional objects in the sample for each speaker); b) rises over di- and polysyllabic words (calculated as the percent of the total number of di- and polysyllabic words in the sample for each speaker); c) rises over monosyllabic words (calculated as the percent of total declarative intonation contours in the sample for each speaker, since monosyllabic words were present in nearly every phrase); and d) high or rising pitch at the beginning of a contour (calculated as the percent of total declarative contours in the sample for each speaker). Prepositions and other words which are normally not stressed in any of the lects of Tok Pisin were not counted. Special nonfinal intonation contours were excluded and only emotionally neutral statements were considered.

RESULTS

Pidginization and Creolization

The results in Table 1 show that: a) words are stressed substantially less often in Sepik speech than in Eastern Highlands speech (Wurm would have predicted 100 percent or more in each environment); b) ethnolinguistic background is the social factor which best predicts stress reduction rates (the Olo group consistently reduces stress more often than the Boiken group); and c) the differences in the stress retention rates between first language (L1) speakers of Tok Pisin and those who speak Tok Pisin as a second language (L2) are not significant. Given the fact that in many Sepik languages (Boiken and Olo included) stress is not phonemic and can be reduced by intonation phenomena, while most Eastern Highlands languages have phonemic word stress or word tone which is not normally reduced, substrate language influence appears to be the most reasonable explanation for the differences between Sepik and Eastern Highlands pitch patterns. These substrate language influences, moreover, persist from second language (pidginized) to first language (creolized) speech, as well as to the speech of children of Olo-Boiken mixed marriages, who have spent all of their lives in Wewak and who know no other language except Tok Pisin (represented by a cross x on Table 1).

Creolization and Decreolization

The evidence in Table 2 indicates that rates of stress retention can be predicted to some extent by the sex of the speaker, especially contour initially. In Tables 3a and 3b, the stress retention rates for individual speakers are plotted on separate curves for each sex, first by age and secondly by years of formal schooling completed. From the results shown in these tables, it would not be unreasonable to postulate that: a) a tendency for males to reduce stress more often than females has become more pronounced and has spread to new environments over the last 20 years or so; and b) this tendency is exaggerated among speakers with the most exposure to Standard English in school (who can be assumed to speak decreolized varieties of Tok Pisin), with female speakers moving closer to Standard English word stress patterns and male speakers distancing themselves from these same patterns.

CONCLUSIONS

The data presented here can only be accounted for by a theory that views pidginization, creolization, and decreolization as parts of a single continuum, with no clear break between one process and another. Substrate language patterns can play an important role in determining not only the varieties of Tok Pisin used by second language speakers but also the creolized varieties used by first language speakers. Rather than introducing new patterns, decreolization (in this case, at least) merely accentuates tendencies already present in the speech patterns of first language and even second language speakers. There is no need to invoke a theory of linguistic universals or some 'bioprogram' to account for some hypothetical disruption in communication here. All that is necessary is to study the speech patterns which typify the other languages with which speakers of Tok Pisin are or have been familiar and to trace the natural and gradual modification of these patterns in different lects of Tok Pisin.

REFERENCE

S.A. Wurm, 1984, "Phonology: Intonation in Tok Pisin. In S.A. Wurm and P. Mühlhäusler, eds., 1984, Hanbook of Tok Pisin (New Guinea Pidgin). Canberra, Australian National University, 309-344.

TABLES

Abbreviations: L1- first language speaker of Tok Pisin; L2- second language speaker of Tok Pisin; B- Boiken speakers: O- Olo speakers; M- male speakers; F- female speakers.

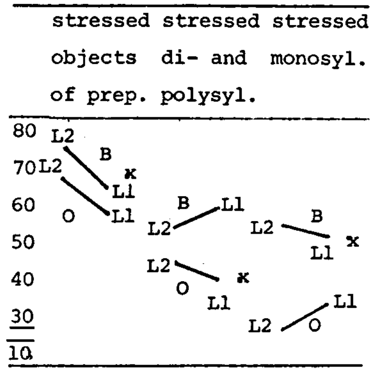


Table 1. Stress by substrate for L2 and L1 speakers

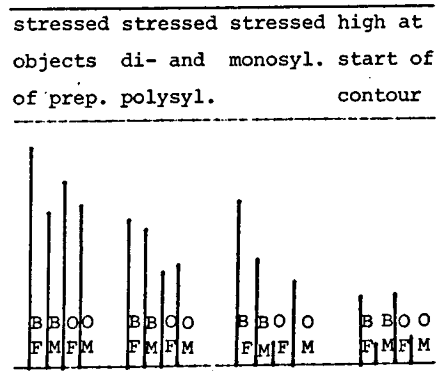


Table 2. Stress by substrate for male and female speakers

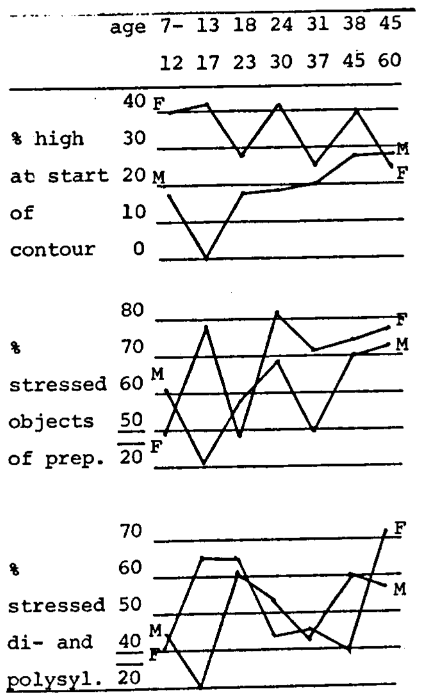


Table 3a. Stress by age

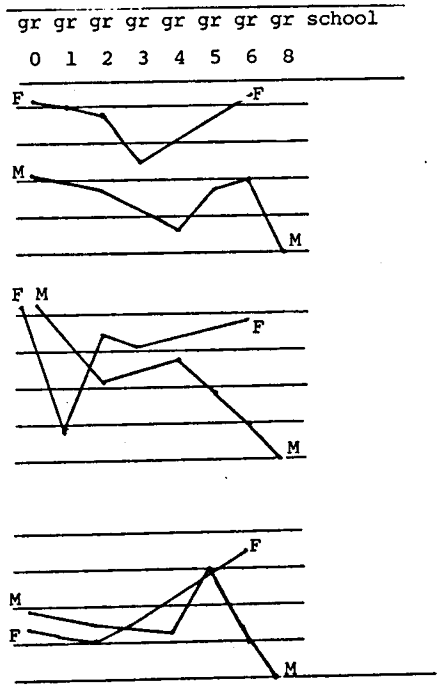


Table 3b. Stress by years of school

THE CONVERGENCE OF CONTACTING PHONOLOGICAL SYSTEMS: THE ALEUT DIALECT OF COPPER ISLAND

Yevgeni V. Golovko, Nikolai B. Vakhtin

Institute for Linguistics
Academy of Sciences
Leningrad, USSR 199053

ABSTRACT

The Aleuts of Copper Island speak a creolized version of Aleut-Russian Pidgin that formed by the end of the XIX c. The phonology of the language, as well as all other levels of its structure, formed as a result of lingustic contact and interference. The consequences of this contact can be traced in the change of phoneme system, in higher variativity of consonants, and in certain phonotactical processes in clusters.

INTRODUCTORY REMARKS

The Aleut language has at present two main dialects, both being subdivisions of what was known in the XIX century as Western Aleut: Atkan dialect (ATK) and Attuan dialect (ATT), named according to the main islands where the dialects are spoken, resp., Atka and Attu (see /8, p. 49/ for details and further references).

In 1826 the administration of Russian-American Company transfered several dozens of Aleut workers that were in its service to previously uninhabited Commander Islands in order to have constant access to the Bering Island and especially Copper Island seal furs. During the 150 years that have passed since that time the new-comers have developed two different dialects: Bering Island dialect (BI) that is a conservative form of ATK, and a creolized Copper Island dialect (CI) of which ATT is considered to be the "maternal" dialect /1/, /6/, /7/.

In this paper we adhere to the hypothesis according to which certain social and historical conditions led to formation by the second half of the XIX century of a peculiar social group of Russian-speaking "creols". This social group later created "Pidgin Aleut" as a means of communication with the Aleut-speaking population of the islands. The newly created language went in course of time through creolization and subsequent relexicalization on the Aleut basis resulting in the modern version of CI (see /7/ for details and further discussion).

The most conspicuous grammatical feature of CI is its system of verbal inflexion for person, number, tense, and mood, as well as negative forms that have definite Russian origin /6/, /7/. All the rest grammatical subsystems are typically Aleut: nominal inflexion, derivation of both verbs and nouns, NP syntax, a.o.

The aim of the present paper is to outline the phonological consequences of creolization, of converging sound systems of Russian and Aleut that resulted in formation of a rather peculiar CI phonological system.

There are at least three points in CI phonology where we can suspect Russian interference: 1. the phoneme inventory; 2. the corruption of velar/uvular opposition; and 3. consonant clusters.

THE PHONEME INVENTORY

Professor Knut Bergsland called the absence of labial obstruents /p,b/ the most conspicuous feature of Aleut phonology, and considered it to be of diachronic character /3, p.69/. However, W.Jochelson who did his field work on Attu at the begining of the XX century, used letters b and p to mark "bilabials... pronounced with a very slight closure of the lips so that we have transient sounds between b and v, p and f" /4, p.1/; see also /5, p. 131/. K.Bergsland who transcribed Jochelson's ATT texts rendered his p and b as v /2, p.9/, i.e. he rightly treated them as two allophones of one and the same phoneme.

Unlike all other Aleut dialects, CI has both /p/ and /b/. CI /p,b/ obviously originated from ATT bilabial /v/ (corresponding to ATK /m/, /M/, /mg/, or sometimes to labiovelar /w/, cf. ATT qavya=, CI qabya=, ATK qamḋa= to be deep; ATT čavluꜧ, CI čabluꜧ, ATK čamluꜧ floor; ATT kaviꜧ, CI kabiꜧ, ATK kamgiꜧ head; ATT avčul to tell, CI apcuꜧ tale (ATK aMas ask?); ATT ava=, CI aba=, ATK awa= to work, etc.

The choice between /p/ and /b/ is definitely positional: ATK /m/∽ CI /b/ before sonants, voiced fricatives or vowels, and ∽/p/ before voiceless consonants: ATK čamluukaꜧ, CI čabluukaꜧ chin; ATK imḋu=, CI ibyu= to roll (a stone); ATK hamraꜧ, CI habraꜧ sleeve; but ATK čamčxiꜧ, CI čapčiꜧ fishline; ATK umsuꜧ, CI upsuꜧ or umsuꜧ tongue; ATK umta=, CI upta= to smell. Before sonants even the voiceless ATK /M/ corresponds to CI /b/, not /p/: ATK saaMlaꜧ CI saablaꜧ egg; ATK čiMignuꜧ, CI čibinuꜧ big toe, etc.

The immediate ancestor of CI /b,p/ is evidently ATT /v/, not ATK /m/: while we have a consistent correspondence of CI /b, p/∽ ATT /v/, in ATK we have here a variety of sounds: /m, w, M/ and a cluster /mg/.

It seems highly probable that the two separate sounds appeared in CI as a result of the splitting up of /v/ that was triggered by strong Russian influence. It was quite natural for the Russian-speaking makers of CI creole to treat the bilabial fricative /v/ as bilabial stops /p/ and /b/ depending on its position.

CORRUPTION OF VELAR/UVULAR OPPOSITION

In Aleut (and Eskimo) there are two distinct interlingual (as well as morphonological) rows of correspondences: /g-k-x/ and /r-q-ꜧ/ that never mix up: there is no such thing as alternations /g-r/ or /k-q/. The only dialect that breaks the rule is CI.

CI loses uvular sounds in many cases, substituting them for corresponding velars, cf. ATT (txin) irata= get frightend, ATK iratu=, CI igatu= to be afraid; ATT hiꜧta=, ATK hiꜧta=, CI hixta= to speak, etc. Uvular sounds in general are rather unsteady in CI: there are words that are pronounced with /ꜧ/ or with /x/ alternatively by one and the same speaker, e.g. axsa=/aꜧsa= to die, a.o.

In two cases the substitution is reversed: where in ATT we have /g/, in CI we find /r/: ATT qaglaꜧ, CI qarlaariꜧ raven; ATT qagliꜧ, CI qarliꜧ shoulder.

Finally, there are even cases when the voiced fricative /g/ is pronounced as a voiced stop /g/ - a sound unknown to the Eskimo - Aleut phonology, e.g. tiꜧ anagasal I hit myself with smth.; tin ačigait he learns (cf. ATK txin ačixal), etc.

We think that these facts can be best explained by direct Russian influence: the Pidgin ancestor of CI must have had a highly Russianized phonetics, and one of the first things to do was to eliminate from the Pidgin all the sounds that are not found in Russian and are thus hard to pronounce for Russian-speaking comunity, and to simplify ("to correct") the pronunciation of the rest. Consequently, the uvular sounds were forced out, the velar/uvular opposition was ruined, and later re-established itself in a corrupted form.

CONSONANT CLUSTERS

Phonotactical peculiarities of CI or, more specifically, order of consonants in typical clusters can be of two types: 1. those developed by CI during the 150 years of its independent existence - that is, independent from its "maternal" language ATT, and 2. those developed as a result of close contact with BI (=ATK). The borderline between the two types of changes is naturally rather unsteady. We shall consider cases when CI differs from both ATK and ATT as independent development and cases when CI differs from ATT but is identical with BI - as BI interference, though such a decision is of course quite arbitrary.

Let us consider changes in consonant clusters in CI compared with ATK and ATT, i.e. metatheses. Metatheses are frequent as interdialectal correspondences in Aleut; they are also found within any of the Aleut dialects. In CI metatheses are also numerous. The following classification can be suggested as regards the possible sources of metatheses. If a CI cluster differs from ATT, it means that there is a development of some kind. If CI is identical with ATT, it is the ATT "heritage". Now, if CI differs from ATT, but is identical with ATK, it may be ATK influence. Finally, if CI is different from both ATT and ATK, it must be independent development. Consider:

1. CI ≠ ATT, ATK. /yg/∽ /gy/: ATT aygags, CI agyagait, ATK aygahs to walk, to go; /sꜧ/ ∽ /ꜧs/,/xs/: ATT asxinuꜧ, ATK asxinuꜧ, CI axsinuꜧ girl, daughter. /tx/, /tk/∽/ꜧt/: ATT atkiyaꜧ, CI aꜧtiyaꜧ, ATK atxiḋaꜧ cod. /gl/∽/lg/: ATT ilga=, CI igla=, ATK ilga= to seek, look for. /rn/∽ /nr/: ATT suganriꜧ, CI sugarniꜧ, ATK suganriꜧ young person.
2. CI, ATK ≠ ATT. /rn/∽ /nr/: ATT hunrutaꜧ warmth, CI hurnayait, ATK hurnas to be warm. /gn/∽ /ng/: ATT qingaꜧ, CI qignaꜧ, ATK qignaꜧ fire.
3. CI, ATT ≠ ATK. /gl/∽ /lg/: ATT igluꜧ, CI igluꜧ, ATK ilguꜧ grandson. /ks/, /xs/∽ /sx/: ATT ixsaꜧ, CI iksaꜧ, ATK isxaꜧ place, bed.

As can be seen from the given examples, CI seems to have developed a phonotactical pattern of its own, at least as far as consonant clusters are concerned, namely, it follows a rule that places velars and uvulars before any other consonant. In all three groups of metatheses CI tends to place velars /k,x,g/ and uvulars /ꜧ,r/ before sonants, glides and fricatives (and in one case before a stop). The pattern is consistent and purely CI, i.e. it is independent of ATT and ATK clusters. If ATT had the same order of consonants, CI preserves it, resisting the BI influence. If ATT had the reverse order of consonants, CI develops its own order either independently or perhaps yielding to BI influence.

The tendency of cluster organization can also be treated as a result of language contact; most probably, its source can be found in the Russian influence which is very strong in CI. CI seems to be the only Aleut dialect that developed so consistent a tendency of cluster shape.

BI INFLUENCE

There are several cases when changes in CI compared to ATT probably took place as a result of direct BI borrowing, cf.: ATT /ŋ/ ∽ CI /n/: ATT haaŋuẖ, CI haanuẖ, ATK ẖaanuẖ salmon; ATT /v/ ∽ CI /m/ (not the usual /b/): ATT kiv=, CI kim=, ATK kim= to descend, to walk down.

CONCLUSION

The phonology of CI differs noticeably from that of other Aleut dialects. These differences are most likely due to the specific position of CI among other Aleut dialects as a creolized Pidgin Aleut. Many changes that occured in CI compared to maternal ATT can be explained by Russian interference. Other changes may have occured as a result of permanent contact with BI. The possibility of independent development should be also taken into consideration.

REFERENCES

/1/ А.С.Асиновский, Н.Б.Вахтин, Е.В.Головко. Этнолингвистическое описание командорских алеутов. - "ВЯ", 1983, №6.

/2/ K.Bergsland. Aleut Dialects of Atka and Attu. - Transactions of the American Philosophical Society, New series, vol. 49, pt. 3, 1959.

/3/ K.Bergsland. Comparative Eskimo-Aleut Phonology and Lexicon. - Journal de la Societe Finno-Ougrienne, vol. 80, Helsinki, 1986, pp. 63-137.

/4/ W.Jochelson. Essay on the Grammar of the Aleut Language. - American Philosophical Society, Boas Collection, ms.

/5/ В.И.Иохельсон. Унанганский (алеутский) язык. - В кн.: Языки и письменность народов Севера, ч.3, М.-Л., 1934, с.129-148.

/6/ Г.А.Меновщиков. К вопросу о проницаемости грамматического строя языка. - "ВЯ", 1964, № 5.

/7/ Н.Б.Вахтин. Некоторые особенности русско-алеутского двуязычия на Командорских островах. - "ВЯ", 1985, № 5.

/8/ A.C.Woodbury. Eskimo-Aleut Languages. In: Handbook of American Indians. Vol. 5: Arctic (Vol.ed. D.Damas). Washincton, Smithsonian Institution Press, 1984.

THE PROBLEM OF BILINGUISM AND PHONETIC PECULIARITIES OF RUSSIAN SPOKEN BY THE KAZAKH

Abuov Zhumagaly

Kzyl-Orda Pedagogical Inst., Kzyl-Orda, Kaz.SSR, 467014

Shcherbakova L.P.

Leningrad State University, Leningrad, RSFSR, 199164

ABSTRACT

The paper presents the results of a phonetic analysis of the segmental and suprasegmental characteristics of the Russian language of Kazakh speakers. It shows a general tendency for the variation of standard Russian pronunciation and describes dialect features specific for the inhabitants of South Kazakhstan, North-Central Kazakhstan, West Kazakhstan and East Kazakhstan.

The Russian language of the USSR is not only the national language of the Russian people but it is also a means of international communication. It is taught at schools and institutes of Kazakhstan alongside the mother tongue. The interaction and interpenetration of the two languages, Russian and Kazakh, result in specific features of the bilinguism /I/. From the linguistic point of view the main problem, in terms of bilinguism, is to describe both language systems, state the difference in them and thus predict the probability of interference. And we must bear in mind that the phonetic systems need not represent languages related to, or different from, one another. The degree of genetic kinship of two interfering languages is not the decisive factor for mastering the pronunciation of a foreign language, the most important factor being the peculiarities of phonetic realization of the sound systems of these languages /2/.

The present paper describes the characteristics of the Russian speech of the Kazakh inhabiting the Southern, Western, North-Central and Eastern regions of Kazakhstan. In this connexion, local phonetic features of native speakers should be taken into account. The first results of this kind of research appeared but recently, during the 1950s. New Kazakh linguists produced a number of monographs dealing with the problems of local dialects, dialect vocabularies, questionnaires, subject collections of scientific papers, the first volume of an experimental atlas of the Kazakh language, and other works. They all register and describe the characteristics of the speech of the inhabitants of a certain region; the boundaries of the expansion of these characteristics are defined and scientifically interpreted. Thus due to the painstaking efforts in gathering the relative data and its detailed description, the Kazakh dialectologists S.Amanzholov and Zh.Doskarayev proved the existence of specific local characteristics in the speech of the Kazakh.

S.Amanzholov describes three main regional dialects in the Kazakh language, and he believes that the North-Eastern dialect is the basis of the modern Kazakh language /3/.

Zh.Doskarayev presents his own point of view based on the phonetic principle and he believes in the existence of two vernaculars: the so-called "Chock" vernacular (S-E) and the "Shock" vernacular (N-W). His classification gives no information on lexical or grammatical characteristics of these vernaculars /4/.

Sh.Sarybayev presents quite a different point of view on the dialect division of the Kazakh language. The research was carried out on all linguistic levels and the results made it possible to distinguish four regional groups of dialects: South, West, North-Central, East /5/.

We believe this distinction to be the most convincing one.

The aim of this research, made on the basis of reception analysis, was to describe the phonetic peculiarities in Russian spoken by the Kazakh population in the regions mentioned above. Twenty native Kazakh speakers from each region were involved in the research.

The material of the research is a text prepared at the Laboratory of Experimental Phonetics at Leningrad University, containing 200 most frequent syllables of the Russian language. The text contains about 3000 phonemes in standard transcription. Sound duration is about five minutes.

The text is a story including monologues and dialogues, abounding in all kinds of orthoepic difficulties. For example,

forelingual noise consonants have 538 realizations which fall into three groups: occlusives (279), fricatives (241), affricates (48). The vowels occupied the positions of various degrees of reduction, both quantitative and qualitative. The research was carried out with the help of native speakers from the regions mentioned above. They were teachers, students of the first and the fifth years, and people on the staff at various institutes in these regions. Before they were recorded they had an opportunity to see the text. The recording was done in a specially equipped studie on "a Reporter" tape-recorder at a speed of 19.5 cm/sec.
Every person gave his surname, name and patronimic, the date and place of birth, qualification, profession, place of study or work. The data could be useful for explaining some phonetic peculiarities of the speakers.
After the reception analysis of the recorded material the results were divided into two groups: a) the speech of those who have a poor knowledge of Russian, b) the speech of those who have a good or medium knowledge of Russian.
The recordings of 80 people were several times listened to and analysed. In accordance with the aim of the research special attention was paid to the analysis of consonants and vowels according to their differential properties, to the realization of word stress and intonation. Minutes were taken of every recording and all the deviations from standard pronunciation of the Russian language were registered, as well as all extra-linguistic deviations, such as pauses, emotions and hesitation.
The reception analysis of 80 Kazakh speakers showed the following general peculiarities in stressed vocalism: In stressed vowels, I) Instead of standard /e/ we have a more open /ɛ/ after a palatalised consonant, with the preceding consonant insufficiently palatalised, or non-palatalised. 2) The closed vowel /i/ after non-palatalised consonants gives a more open /ɪ/. 3) The non-uniform /e/ with a narrow u-like beginning and an open end changes into a uniform vowel which is more open than it ought to be. 4) The Russian /ы/, which in standard Russian pronunciation comes after a non-palatalised consonant, in the Kazakh speech is front-retracted /ɪ/ with the preceding consonant palatalised. 5) Instead of standard open /e/ after a non-palatalised consonant we get a less open /e/ with the preceding consonant slightly palatalised.
In unstressed vowels, the peculiarities are as follows: I) The unstressed /e/ is used in all positions of the orthographic "e", i.e. there is no alternation of o/a is an unstressed syllable. 2) Orthographic pronunciation of "a" and "я" replaces the standard literary ы / ʲ -variants, i.e. there is no alternation of a/ʌ in an unstressed position. 3) To pronounce the orthographic "e" in an unstressed syllable a more closed vowel is used with the preceding consonant palatalised, whereas in standard literary pronunciation we have / ы / ɛ / ə with the first variant preferable; there is no alternation of e/ы is an unstressed position. 4) The standard literary i/ь for the orthographic "e" is replaced by /e/, i.e. there is no alternation of e/i. 5) Instead of the standard ы / ə after a non-palatalised consonant (for the orthographic "и"/"ы") we get a front-retracted vowel /ɪ/ with the preceding consonant palatalised.
6) To pronounce the orthographic "e" after a palatalised consonant a more open vowel is used, /ɛ/, instead of the standard i/ь, with the preceding consonant insufficiently palatalised. 7) The vowel i/ь after a palatalised consonant is replaced by a more open /ɪ/ with the preceding consonant insufficiently palatalised. 8) There is a sharp reduction in the post-stressed vowels nearing the "zero" mark in the sound volume. 9) If there are vowels in a word which need different shades of reduction, a Kazakh speaker may pronounce them with one and the same degree of reduction. 10) The interfering influence of synharmonism is observed, i.e. all vowels in a word are assimilated to one another.
The peculiarities of pronouncing Russian consonants by Kazakh speakers are as follows: I) The palatal consonants are not sufficiently palatalised a) before the front vowels, b) at the end of a word after central and back vowels, c) with the consonants /č, š:/ invariably non-palatalised. 2) The non-palatalised consonants /š, ž, c/ are palatalised. 3) The sound /j/ is not pronounced in combinations of the type c + j + r. 4) As a rule, the labio-dental fricative consonant /v/ is replaced by the labial-labial /w/. 5) /b/ is mispronounced as /w/. 6) The plosive voiced back lingual /g/ is replaced by the voiced occlusive /ɣ/. 7) There is a weaker plosive in the final consonant.
8) The back lingual voiceless /k/ is replaced by the plosive voiced /g/. 9) A pronounced contiguity of final consonants. The back lingual fricative /x/ is replaced by the uvular-fricative /χ/.
10) The noise occlusives are insufficiently voiced. II) The velarised /ł/ is replaced by the apical /l/.
Deviations from standard Russian pronounciation occur in pronouncing combinations of consonants: I) There is no palatalisation of preceding consonants, so that /s/ and /z/ are pronounced as non-palatalised sounds before the palatalised /t'/ and /d'/ without any assimilation taking place. 2) Before the noised voiced consonants we get the noise voiceless sound, i.e. there is no assimilation in the volume of voice.
3) There is no replacement of the backlingual occlusive by the corresponding fricative, i.e. a non-standard dissimilation occurs.
Deviations in rhythmical organisation are as follows: I) The word stress may be misplaced. 2) The number of syllables in a word may be increased due to an extra vowel introduced.
Of this total list of phonetic peculiarities, several characteristics may be ascribed to specific regions: I) In the South Kazakh region, especially in Chimkent district, the voiced occlusive backlingual /g/ is replaced by the voiced fricative uvular /ɣ/, which was not observed in any other region. The final consonants are "less contiguous", which is another specific feature of the Russian language spoken by the Kazakh population in that region. 2) In the East Kazakh region, especially in Semipalatinsk and Taldy-Kurgan, the post-stressed syllables are greatly reduced; the reduction may be either quantitative (even in the 2nd syllable before the stress) or qualitative (even in the Ist syllable before the stress), but not both. 3) In the South and East Kazakh regions, the rhythmical organisation of the Russian language as spoken by Kazakh speakers distinguishes them from those in other Kazakh regions. For example, the words are frequently misstressed. Besides, a) In Chimkent, Kzyl-Orda, Djambul districts, extra vowels are introduced so that the number of the syllables is increased. b) In Taldy-Kurgan and Semipalatinsk districts, the number of the syllables is decreased due to dropping some vowels.
Thus not only the general peculiarities of the Russian speech of the Kazakh but also the peculiarities of the phonetic system of the Russian language as it is spoken in each particular region have been analysed.
The description of various intonation groups is no less important than the description of the segmental characteristics of the Russian language of Kazakh speakers. It is well known that intonation is one of the most important means of carrying information. So it is essential that students should be taught both standard pronunciation and correct intonation. Wrong or distorted intonation makes communication difficult and prevents one from absorbing information.
Some linguists, discussing the problem of bilinguism and the interaction of languages, believe that prosodic interference is the most stable, regardless of the type of bilinguism. L.V.Shcherba introduced the term of syntagm as a minimal sense-group "which is a phonetic unit expressing a sense of unity in speech or thought" /6/. In order to be able to describe the syntagm and define its structure (its melodic, dynamic and temporal contours) it is essential to know the syntagmatic division of the text.
The most universal means of division is a pause. The number of pauses as compared between the model speaker and the Kazakh speakers from different regions characterizes the intonation of the speech. When reading the text for the experiment, the model speaker made 180 pauses while Kazakh speakers, as a rule, made from 220 to 270 pauses. It means that Kazakh speakers make pauses between words rather than between syntagms, and it makes their speech more monotonous and abrupt.
It is also known that "the elements of a syntagm, i.e. words it consists of, may become a unit due to one general stress uniting them and due to their melodic, dynamic and temporal contours"/7/.
The aim of the reception analysis of the experimental material has been to describe the prosodic organisation of a syntagm in the speech of the model speaker and Kazakh speakers, respectively. Certain characteristics were taken into account, i.e. the boundaries of a syntagm, the word under syntagmatic stress, the centre of a syntagmatic stress, the direction of fundamental frequency on a syntagmatically stressed vowel and the fundamental frequency in before-stressed and after-stressed elements of a syntagm. Attention was paid to temporal and dynamic characteristics of vowels of a syntagmatically stressed word.
The results of the reception analysis are as follows: I) The intonation of an incomplete speech: The melodic contours of an incomplete syntagm as pronounced by a model speaker may be expressed in different ways, a) by rising melodic contours and b) by falling-rising melodic contours. The Kazakh speakers in the test were using only one intonation, the falling-rising one. In their speech, the direction of fundamental frequency of the syntagmatically stressed word, and of before-stressed and after-stressed elements of a syntagm, is much weaker and more difficult to recognise.
2) The intonation of a complete speech: The model speaker pronounces a complete syntagm with a falling tune, and the fall of the fundamental frequency begins on a syntagmatically stressed vowel and continues on after-stressed syllables. The Kazakh speakers seldom use the falling tune to complete the syntagm.

Very often the falling tune of a syntagmatically stressed vowel is not enough to describe this syntagm as completed; the after-stressed vowels in Kazakh speech are a little longer, there is no falling tone, sometimes we come across a slight rising tone, which makes the syntagm sound incomplete. It is especially common in the spoken Russian of the West Kazakh region. 3) There are two types of interrogative intonation in the text: the intonation of a general question and the intonation of a special question. There are also sentences which are special questions in form but they are used one after another and so the intonation of enumeration is used instead. For Kazakh speakers, the characteristic features of the intonation of a special question are the rising tone on a syntagmatically stressed vowel and the absence of a fall on after-stressed syllables, so the intonation of a special question may be mistaken for the intonation of incompleteness. As to the general question, the melody is also rising, with the tune usually falling over the last word in the sentence. The model speaker uses various interrogative intonations: the centre of a syntagmatic stress both in a special question and in a general question is not necessarily the last word in a syntagm, the syntagmatic stress is often emphasized by a logical stress. The Kazakh speakers seldom if ever use a logical stress. 4) The intonation of emphasis in the experimental text is represented by intonations of adress and apposition. The Kazakh speakers use the intonation of adress frequently, whereas they use the intonation of apposition extremely rarely. The direction of the tone on a stressed vowel is the same as that used by the model speaker but the dynamic characteristics of a syntagmatically stressed vowel are much weaker. Thus the reception analysis, on prosodic level, of the Russian language spoken by the Kazakh shows the following peculiarities: I) The intonation of incompleteness is characterized by a falling-rising tone, with a slight rise or no rise on after-stressed syllables. 2) The intonation of completeness is characterized by a falling tone (in the West region, falling rising or rising) on a syntagmatically stressed vowel with longer after-stressed syllables and the absence of a falling tone on them. 3) The interrogative intonation is as follows: The general question is characterized by a rising tone on a syntagmatically stressed vowel and the absence of a falling tone on after-stressed syllables. The special question is characterized by a falling tone on a syntagmatically stressed vowel of the last word in a syntagm. 4) The intonation of adress is characterized by a falling-rising tone with a stronger emphasis on the stressed vowel. A study of segmental prosodic characteristics in the Russian language spoken in Kazakhstan may be of both theoretical and practical value. From the theoretical point of view, it helps to define the standard pronunciation, with dialect and regional variants. From the practical point of view, it may be used to improve the teaching of Russian in non-Russian schools.

REFERENCES

/I/. L.V.Bondarko. The phonetic descriptions of a language and phonological descriptions of speech. L., I98I.

/2/. L.A.Verbitskaya, M.V.Gordina. Peculiarities of phonetic interference in languages with different degrees of relationship. In: Problems of philological studies. L., I980.

/3/. S.A.Amanzholov. Problems of dialectology and the history of Kazakh language. Alma-Ata, I959.

/4/. Zh.Doskarayev. On problems of dialectology and the history of the Kazakh language. M. V.Ya. No.2, I954.

/5/. Sh.Sh.Sarybayev. On dialects in the Kazakh language. Baku, I963.

/6/. L.V.Shcherba. The language system and speech activity. M., I974.

/7/. N.D.Svetozarova. The intonation system of the Russian language. L., I982.

THE PROSODY OF A WORD IN RUSSIAN SPOKEN BY MONGOLS

SVETLANA KASHCHAYEVA

Dept. of the Russian Language
Polytechnical Institute
Irkutsk, USSR, 664074

ABSTRACT

The phonetic distortions in the speech of Mongol learners at the level of the prosody of a word are prompted by the differences in the prosodic construction of a word in Russian and in Mongolian. They can be foreseen and foretold as the result of the comparative analysis of the prosodic means of both languages.

INTRODUCTION

The problem of the linguistic interpretation of various breaks in the prosody of a word in the Russian speech pronounced by Mongols seems to be important both in the theoretical meaning (the Russian speech of Mongols is not yet investigated from the point of the complex analysis of the suprasegmental phonology) and in the practical one (in teaching Mongols the Russian pronunciation it will give the chance of the explanation and foreseeing the pronunciation mistakes).

The prosodic means of Russian at the level of the prosody of a word is accent. By its phonetic nature it is quantitative, qualitative and dynamic. A stressed vowel is characterized by a set of phonetic means: the length, the tamber, the strength. According to these characteristics it is realized in communication depending on its functions: culminative (the accent makes a word a phonetic unit), constitutive (helps to identificate words or their forms), distinctive (provides the differentiation of words and their forms, for example: пла́чу - плачу́, стра́ны - страны́).

A word-stress is traditionally considered to be one of the prosodic means of the Mongolian language. The first reference of it in Linguistics is to be found 150 years ago, and up to now the prosody of a word is the least investigated branch of the Mongolian phonetics. In fact this problem has not yet been the subject of the linguistic research and was studied only from the point of the rhythmical nature of the Mongolian versification. There are various, sometimes contradictory opinions in reference to the word-stress in the Mongolian language, its phonetic nature and its seat in the word. Thus, J.Schmidt (1832), Y. Ramstedt (1908), A.B.Rudnev (1913), B.J. Vladimirtsov (1929), G.B.Sanzheyev (1960), Sh.Luvsanvandan (1967), G.Galsan (1975) say that a word-stress in the Mongolian language is expiratory (force) and always falls on the first syllable of the word. On the contrary, O.M.Kovalevsky (1835), A.A.Bobrovnikov (1848), A.M.Pozdneyev (1880), V.L.Kotvich (1902) think that the stress falls on the last syllable and is not expiratory (force).

Such contradictory opinions can be explained by the fact that the seat and the character of the stress were investigated only with the help of the auditory analysis without thorough phonetic research based on experiments.

The experiment results/1/ show that any syllable can have a greater force, especially when it contains a long diphthong We couldn't find any regularity in the correlation of vowels in different positions either in their length or in the change of the tone. The same is about the qualitative differences between them. The experiment and a number of psycholinguistic tests on the perception of a stress by Mongols show that in the Mongolian language there is no word-stress. This language can be related to unaccented ones.

In this case the length of a vowel in the Mongolian language is not the characteristic of the prosodic means but a phonematic indication of the segmental units - long vowels. It is the length of vowels that carries out the distinctive function characteristic of the Russian word-stress It helps to differentiate words (судал - pulse; суудал - a seat; цас - snow, цаас - paper, etc.) and their forms (эвлэ - the imperative of "to reconcile", эвлэ́э - the one who reconciles, etc.).

Some clearness of the pronunciation of the vowels in the initial syllables (traditionally taken for stress) can be explained differently: the effect of the law of synharmony. Synharmony, that is the regulation of the succession of vo-

vels in a word - in other words - assimilation of vowels in prefixes to those ones in the root. Functionally synharmony is the main way of making the phonetic unity of a word (a culminative function), identification of a word in a number of other words (a constitutive function). The problem of the delimitative function of synharmony is not quite investigated in Linguistics. To study this problem we made the statistic analysis of the poetical and prosaic texts in Mongolian.
The differentiation of words according to the law of synharmony coincided with the semantic articulation in the poetical text for 73% of words. The received statistic data show that synharmony is a quite safe way of the differentiation of words.
The comparison of the prosodic construction of a word both in Russian and Mongolian show some likeness and some very essential differences. The likeness is to be found in the functional meaning: the culminative and constitutive functions characterize the prosody of a word both in Russian and Mongolian. The main difference is in the ways of realization of these functions. In Russian it is word accent, and in Mongolian it is synharmony. Besides, both the Russian language and the Mongolian language have different functions. Only the Mongolian prosodic system has a delimitative function. Only the Russian prosodic system has a distinctive function. In the Mongolian language the function is realized not with suprasegmental ways but with segmental ones - long vowels.
Thus we can say that in the situation of the subordinate bilinguism in the Russian utterances of Mongols the following probable errors can be expected:
1.Mongolian learners of Russian can construct a Russian word according to the law of synharmony, making its vocal structure close to the synharmonic models of their native language.
2. In the Russian utterances of Mongols we can expect the errors connected with the absence in Mongolian of the qualitative distinctions of vowels in different word-positions. As a result Mongols can find it difficult to give in their utterances a special tamber quality of a Russian stressed vowel, they are apt to see no difference between stressed and unstessed vowels with reference to the presence or absence of the qualitative reduction.
3.The qualitative reduction of vowels in the non-initial syllables can be expected in Russian utterances, and the further from the word-beginning is the syllable, the stronger will be a qualitative reduction of vowels in it.
4. We can suppose that having difficulties with the proper word-stress in Russian, Mongols will prefer to stress the initial syllables.
5. The length as a component of a Russian word-stress can associate with the length of vowels in their native language. Clusters of consonants in Russian words can be comprehended and reproduced as stressed ones.
6. As a consequence of the functional lack of coincidence of the prosodic means in both languages there can be errors in the segmentation of the auding: Mongolian students can differenciate the unknown words in accordance with the law of synharmony.
This is the brief description of a theoretically possible accent of Mongols in the prosody of a Russian word, which gives us a chance to find out difficulties for Mongolian learners of Russian.
The experiments in comprehension of the accent-rhythmical structure of a Russian word by Mongolian learners, the analysis of their errors prove that the foretold deviation actually take place in the Russian utterances of Mongols, in the prosodic structure.

/1/ Gerasimovich L.K. Mongolian versification: Experimental phonetic study, Leningrad, 1975.

PHONETIC INTERFERENCE IN THE CONDITIONS OF RUSSIAN-UKRAINIAN BILINGUISM

L.F. TSIPTSURA

Dept. of the Russian Language
Institute of National Economy
Kiev, Ukraine, USSR 252057

The report is devoted to the nature, tendencies and specific features of phonetic interference, inevitable in the conditions of the active related bilingualism. The conclusions are based on the author's field research, as well as on the data of the oscillographic, palatographic and radiographic analysis of the Russian speech in the Ukraine.

By interference we understand any linguistic phenomenon resulting from interaction of two languages, two communication systems. Each phenomenon of this kind should be treated not in isolation, but in system. Thus the notion of interference includes not only the given linguistic fact, as, for example, a new quality of an allophone, but also the totality of system relations, either modified or qualitatively new, emerging from a breach (modification) of structure.

According to Soviet and foreign authors two types of interference are realized in morphology and syntax-latent and overt, but only the latter is represented on the phonetic level. This phenomenon is based on the specific features of the phonetic level: rigid interrelation of all the elements of the structure, their strict hierarchy, practical absence of synonymy and variation which make exclusion from speech of any unit, even more so a group of units impossible.

The present linguistic situation in the Ukraine is characterized by active Russian-Ukrainian bilingualism, which means that the overwhelming majority of Slavic population in the Ukraine uses both systems of communication - the Ukrainian and Russian languages. The measure and character of their use are undoubtedly different and the range of difference is wide enough: from genuine command of both languages in their codified form to the use of a peculiar intermediate variant that differs from both systems.

The genetic community, which brings about structural, articulation and acoustic proximity of the related languages, Russian and Ukrainian, creates especially favourable conditions for their transposition and interference in the process of interaction.

Phonetic interference which is registered in the Russian speech in the Ukraine is characterized by exceptional stability. Even in the cases when the system of the Russian language is mastered sufficiently enough and is practically identical in functioning with the native language, the Ukrainian influence in pronunciation is quite strong.

It is necessary to emphasize that the quality of the Russian speech in the Ukraine is noticeably varied, depending not only upon age, education and profession of speakers, but also upon the district and region, i.e. there are a few local variants of Russian in the Ukraine. However, alongside with local peculiarities there exist common features resulting from interaction of the two related languages.

The results of active bilingualism are evident in practically all the sections of the phonetic structure: the set of phonemes, phonological correlations and oppositions, quality and quantity of an allophone, articulation and acoustic characteristics of separate sound units, accentual and rhythmic organization of words, sentence intonation etc.

The system of vowels. The allophones /O/ and /a/ are characterised by a more backward articulation, and the phoneme /e/ by a more open and backward articulation. There is a considerable range of deviations from the norm in the pronunciation of the phoneme /ы/. According to the majority of authors, the Russian phoneme /ы/ is a central high vowel (in some cases its articulation may be somewhat backward). Instrumental analysis shows that in the Ukrainian language this is a front higher-mid vowel. According to palatographic analysis the most typical for Russian speech in the Ukraine is the central /ы/, but noticeably lower, than the normative sound.

The universal peculiarity of the Russian speech in the Ukraine is the different principle of changes of unstressed vowels. In the overwhelming majority of cases the sounds pronounced instead of /o/ and /a/ in the first pretonic syllable do not coincide in quality: the sound pronounced instead of /a/ is approaching /ʌ/, which is more open and long, than the normative Russian /ʌ/.

A somewhat fronter and higher /e/ is pronounced after palatalized consonants, while in the Russian literary pronunciation the sound approaching /и/ - /и^е/ is the norm.

Thus the system of vocalism of the first pretonic syllable in the Russian speech in the Ukraine is represented by the sounds: /ʌ/, /a/, /ʌ°/, /e/, /и/, /ы/, /у/. It obviously differs from the realization in the same position of vowel phonemes in normative Russian: in the Russian speech in the Ukraine vowels in the first pretonic syllable in their quantity and especially quality approach stressed vowels. Though these vowels are reduced, the character of reduction is different: the sounds do not lose their main quality. The influence of the Ukrainian norm here is obvious.

There are other deviations from the Russian orthoepy in unstressed syllables in the Russian speech in the Ukraine.

In modern Russian the second pretonic syllable allows considerable (though not maximum) reduction of a vowel. The system of vocalism in this position is represented by only five vowels: /ы/, /и/, /у/, /ъ/, /ь/.

The system of vocalism in the second pretonic syllable is different too. The short /a/ is pronounced instead of /o/, /a/, and often /a/, pronounced instead of /o/, is somewhat labialized, as in the first pretonic syllable: / маг з'ин /, /лаб/\°рант/ etc.

Vowels /e, a/ that follow palatalized consonants also differ in their pronunciation in the second and other pretonic syllables: /в'ент'ил'а'тъ°р /, /н'езн/\комыј /, /м'аса̂фупка/.

Thus, the reduced sounds /ъ/, /ь/, so typical in this position in the Russian orthoepy, are practically non-existent in the Russian speech in the Ukraine.

It is interesting to note that after /ж/,/ш/,/ц/ in pretonic syllables other than the first, the sound approaching /ы/ is pronounced instead of the vowel e: /шыс'д'ес'ат /, /шырст'аној /, /жыр'еб'ец/, /цыл/\°вала /. The use of the vowels и,ы,у generally corresponds to the norm.

The system of vowels in the third and fourth pretonic syllables is the same as in the second pretonic syllable. On the contrary, in the absolute beginning of the word all the vowels are pronounced very distinctly, especially the vowels /a/and/o/. This distinction is preserved even in those cases when in other pretonic syllables these vowels coinside: /очев'идно /, /отблагадир'ит' /. So the difference between the system of the first and other than the first pretonic syllables, so prominent in the Russian orthoepy, is less noticeable in the Russian speech in the Ukraine.

In modern Russian the degree of reduction of vowels in posttonic and pretonic syllables is of graduated character: there are syllables with a stronger and weaker reduction.
The maximum reduction is registered first of all in the first posttonic syllable. In this position the complete reduction is also the most frequent.

In the Russian speech in the Ukraine the general tendency of vowel reduction in a posttonic syllable is normally the same. This reduction is mainly quantitative.

Of special interest are the data characterizing the principles of accentual and rhythmic organization of the word. The comparison of accentual and rhythmic structure of the word in Russian, Ukrainian and Byelorussian, originally related but now functioning as separate languages, makes it possible to show common and specific typologic tendencies in East-Slavic accentuation. The common tendency is, above all, the centralizing function of stress. In all the languages stress is the distinctive feature of the word, which determines its accentual and rhythmic structure. However, the realization of the dominating influence of a stressed syllable on the other parts of the word are not the same in Russian and Ukrainian. Oscillographic analysis shows that the principles of accentual and rhythmic organization of the Russian speech of the Ukrainian population does not fully coinside with any modern normative East-Slavic language.

Preserving length as the main correlative of stress, the phonetic word in the speech of bilinguals is normally characterized by entirely different quantitative correlations between stressed and unstressed vowels and between unstressed vowels of different degree of reduction, which is one of the reasons of the specific Ukrainian accent, i.e. a more noticeable degree of vocalization of speech, its melodiousness.

The system of consonants. The system of consonants in the Russian speech of the Ukrainian population is close to the norm with the exception of the backlingual consonants. Thus atpresent /к'/, /х'/, /г'/ in the Russian speech of the Ukrainians do not have the status of phonemes but are variants of non-palatalized /к/, /г/, /х/. This peculiarity finds its explanation in the fact that in Russian orthoepy palatalized /к'/, /г'/, /х'/ became phonemes only recently and are still positionally limited in comparison to other palatalized consonant phonemes.

One of the most common and noticeable deviations from the orthoepic norm in the Russian speech in the Ukraine is the quality of the plosive voiced backlingual . Observations show that in the absolute majority of cases this consonat is substituted with the fricative /γ/. In its articulation and acoustic characteristics the sound is approaching the South-Russian dialectal /γ/, but it is noticeably different from the corresponding consonant in the Ukrainian/h/.

As a result the fricative, voiced in the Russian speech of the Ukrainians, is devoiced not into /к/, as in the normative speech, but into /х/. This deviation from the norm is even a more constant peculiarity than the fricative /γ/.

There are certain peculiar features in the system of labial consonants too. There is only one labiodental consonant in the Russian normative speech - the consonant b. In the Russian speech in the Ukraine the phoneme /в/ is very often realized in two variants: labiodental /в/ and bilabial /w/. It should be noted that the confusion of /в/ and /w/ may occur quite frequently in the speech of the same person and it is difficult to trace any tendencies here.

In the system of labial consonants in the Russian speech of the Ukrainians the consonant /в/ often remains unpaired in the opposition of voiced and voiceless consonants, because it does not alternate with /ф/ in the word-final position and in the end of the syllable, followed by a voiceless consonant, thus differing from other voiced phonemes.

In word-final positions the labial consonants are as a rule velarized: /кроw/, /сып/, /с'ем/, /голуп /, /руп/, /стаw/, /вос'им/. It is known that of all lax labial consonants in Russian and in some Russian dialects only the sonorous /м/ is velarized. In the other dialects and also in the Ukrainian and Byelorussian languages the process received further development, and all the labial consonants in the absolute word-final position were velarized. So the tendency is not new, what we see in this case is a somewhat specific development of the long-standing Russian (East-Slavic) tendency.

Experimental phonetic investigation of the Russian speech in the Ukraine testifies to the fact that in the conditions of the related bilingualism the tendency is not towards the mixture of the sound systems but towards their genuine interaction resulting in emergence of a qualitatively new language phenomenon, which has no absolute analogue in the contacting languages.

In the conditions of the active related bilingualism the interference is realized not only in the system of sounds but also in their distribution.

The specific result of such interaction is enrichment of the basis of articulation. We assume that the main reason of this phenomenon, its certain precondition, is the genetic community of the languages, creating the community of the articulation basis.

ФУНКЦИОНИРОВАНИЕ ЗВУКОВОЙ СИСТЕМЫ РУССКОГО ЯЗЫКА В УСЛОВИЯХ ТУРКМЕНСКО-РУССКОГО ДВУЯЗЫЧИЯ

ДЖЕМИЛЯ САРДЖАЕВА

Туркменский государственный университет
Ашхабад, Туркменская ССР, 744000

Общелингвистическая проблема языковых контактов особенно актуальна в нашей многонациональной стране (СССР) в связи с функционированием русского языка в качестве языка межнационального общения. Динамическая структура современного русского языка оказалась способной принять огромное количество вариантов живой речи представителей разных национальностей. Тенденция роста туркменско-русского двуязычия дает основание заключить, что функциональное взаимодействие языков будет в дальнейшем углубляться и ускоряться.

Численность туркмен, проживающих в Советском Союзе - по данным статистики /I/ - 2028000, причем на территории Туркменской ССР проживает I892000, т.е. 93,3%. Наибольший удельный вес лиц, проживающих на территории своей республики, характерен также и для грузин, литовцев, латышей, эстонцев (от 96-93%). Компактность расселения туркмен влияет и на этнолингвистические процессы. 98,7% туркмен при переписи I979 года родным языком называли язык своей национальности. Этноязыковые процессы в нашей стране характеризуются развитием двуязычия - широким использованием наряду с родным языком - русского языка как языка межнационального общения. Распространение туркменско-русского двуязычия в известной мере зависит от этнической среды /2/. В городах и поселках городского типа, население которых характеризуется разнонациональностью, число двуязычных туркмен значительно больше, чем в селах, где проживают в основном представители коренной национальности. Существует также прямая связь между расширением туркменско-русского двуязычия и социально-профессиональной деятельностью туркмен. Двуязычие менее распространено среди сельского населения республики. По данным ЦСУ СССР свободно владеют русским языком как вторым языком 31% городского населения туркмен и 7,4% сельского населения /3/.

Чрезвычайно убедительны данные возрастающей распространенности туркменско-русского двуязычия, приводимые за межпереписной период I970-I979 гг.

Туркменское население		Родным языком считают		Свободно владеют вторым языком	
	общая численность	туркменским	другие языки	русским	другими языками
I970 I979	I4I6700 202800	99,3% 98,7%	0,7% I,3%	I5% 25,4%	2,27% I,6%

Объектом настоящего исследования является фонетическая интерференция при взаимодействии русского и туркменского языков, языков с контрастными фонологическими системами. При взаимодействии этих языков фонетическая интерференция определяется как на уровне систем, так и на уровне статистических структур и вариативности фонем в потоке речи.

Система туркменского литературного языка включает I8 гласных и 2I согласную фонему, в русском языке - 6 гласных и 36 согласных.

	Русский язык	Туркменский язык
согласные	п,п',б,б',г,г',д,д', к,к',м,м',н,н',л,л', с,с',з,з',ф,ф',в,в', ш,ш',ж,ж',х,х',й,ч, ц,р,р',щ	p, b, g, к, t, d, m, n, ŋ, l, j, w, ʃ, ʒ, θ, ð, h, r, ф, tʃ, dʒ.
гласные	а,о,у,ы,е,и	а,а:, о,о:, и,и:, ы,ы:, i,i:, y,y:, е,е:, æ,æ:, ɛ,ɛ:

Приведенный количественный состав фонем, свидетельствует о вокалической доминации туркменского языка и о большей консонантной наполненности русского языка.

Анализируя консонантный состав исследуемых языков, мы придерживаемся следующих принципов классификации:

I. По типу преграды и способу образования шума.

2. По активно-артикулирующему органу и месту образования преграды.
3. По участию голоса и шума.
4. По работе голосовых связок.
5. По положению мягкого неба.
6. По окраске или наличию - отсутствию смягчения.

Согласно первому принципу классификации консонанты в туркменском и русском языках подразделяются на три однозначные группы: смычные, щелевые, дрожащие (дрожащий туркменский /r/ малоударный сравнительно с русским /р/).

Количество подразделений согласных по активно-артикулирующему органу произношения зависит от участия активных органов речи. В связи с этим следует различать две основные и две факультативные группы согласных: I. губные, 2. язычные, 3. увулярные, 4, глоточные (фарингальные). Если к первым двум группам относятся согласные обоих языков, то последние две группы представлены только звуками туркменского языка [q, ɢ, ħ], представляющие позиционные аллофоны фонем /k, g, x/. В русском языке группа губных согласных разбивается на две подгруппы (губно-губные и губнозубные), согласным туркменского языка нетипична губно-зубная артикуляция, она встречается только в заимствованных словах.

По участию голоса и шума согласные туркменского и русского языков разделяются на две равноценные группы: сонорные и шумные. По работе голосовых связок согласные обоих языков делятся на глухие и звонкие, но звонкость согласных явление в фонетическом отношении неоднородное. Разную степень звонкости в туркменском и русском языках характеризуют позиционные аллофоны, туркменскому языку особенно свойственны полузвонкие согласные. По положению мягкого неба согласные анализируемых языков делятся на носовые и ртовые. В системе туркменского языка имеется заднеязычный носовой согласный /ŋ/, отсутствующий в русском языке.

С фонетической точки зрения явление палатализации отмечается в большинстве языков мира, в том числе и в туркменском и в русском языках. Артикуляторно этот признак выражается в поднятии средней спинки языка к твердому небу. Но в русском языке в отличие от туркменского этот признак выполняет не только фонетическую, но и фонологическую функцию, являясь дифференциальным признаком согласных, благодаря которому одни согласные фонемы противопоставлены другим. Палатализация или ее отсутствие различает смысл русских слов и их форм.

Сводная таблица русских и туркменских согласных подтверждает превосходящий количественный состав русских согласных за счет фонологического противопоставления мягкость-твердость, отсутствия в инвентаре туркменского языка круглощелевых шумных /с, з/, губно-зубных щелевых /ф,в/ и аффрикаты /ц/ и наличия нескольких фонем, которые не имеют своих соответствий в русском языке: плоскощелевые интердентальные фонемы /θ,ð/ губно-губные щелевые /ф, w/, заднеязычный смычный сонант /ŋ/ и аффриката /dʒ/.

Туркменский вокализм представлен I8 фонемами (количество их до сих пор спорно), девятью парами, строго разграниченными по долготе-краткости. Долгота туркменских гласных имеет функциональную нагруженность. Гласные звуки русского языка как и гласные туркменского языка делятся на идентичные классификационные подгруппы.

Подъем \ Ряд		Передний ряд	Средний ряд	Задний ряд
Высокий	узкий	И y y: i i:	Ы ы ы:	У и и:
	широкий			
Средний	узкий	е е:		
	широкий	Е æ æ:		О о о:
Низкий	узкий			
	широкий	ɛ ɛ:		А а:а

Судя по таблице, где туркменские фонемы обозначены мелким шрифтом, усвоить изолированные аллофоны 6 фонем русского языка не представляет особой трудности для носителей туркменского языка, исходя из присутствия равноценных фонем в системе родного языка. По ряду, подъему и огубленности туркменские гласные противопоставлены более дифференцировано. Каждая из трех степеней подъема представлена двумя разновидностями (широкой и узкой), некоторые гласные реализуются также разнообразно, как "передне-среднерядное сверхкраткое" /i/, лабиализованный /o/ низкого подъема, более широкий средне-нижнего подъема, передне-среднего ряда /e//4/. Основной же трудностью для носителей туркменского языка является позиционная обусловленность по отношению ударения. Сильной позицией для русского гласного является ударный слог. В безударных слогах гласные в русском языке подвергаются качественной и количественной редукции.

Диаметральной противоположностью и особенностью туркменского вокализма является чрезвычайно малая степень редукции безударных гласных. С функциональной точки зрения для гласных сильной фонетической позицией оказывается первый слог многосложного слова. Безударные гласные в начале слова характеризуются относительно большей интенсивностью и высотой частоты основного тона, по сравнению с серединой слова. Длительность туркменских гласных в безударных положениях в зависимости от характера слоговой структуры или увеличивается (в словах с открытыми слогами), или же уменьшается (в словах с закрытыми слогами) к концу слова. Гласные туркменского языка в начале слова, по сравнению с серединой и концом, сохраняют наиболее свойственное им качество и реализуются в виде основного аллофона. Безударные гласные туркменского языка больше изменяются в количественном отношении, чем в качественном.

Вариативность же русских безударных гласных в речевом потоке чрезвычайно велика и многообразна. В связи с малой длительностью безударных гласных, они часто не имеют так называемого "стационарного" участка, т.е. такого участка звучания,где частоты формант были бы постоянными.Поэтому, несмотря на малое табличное количество гласных фонем, в речи их насчитывается до I8 эталонов /5/.

Трудности овладения звуковой системой оказываются в основном связанными с необходимостью усвоить новые или несколько отличные от привычных способов реализации.В туркменском языке максимально возможное количество согласных в абсолютном конце слога только два, причем первый из них обычно сонант, второй - шумный смычный или щелевой. Сопоставительный анализ конца слов туркменского и русского языков показал, что в русском языке возможны I33 конечных сочетаний из двух согласных, а в туркменском только 23. Отличительной специфической особенностью туркменского языка является также разрешение системой употребления звонких согласных в конце слова.Характерной для слогов, составляющих единое фонетическое слово в туркменском языке,является и роль слогообразующих гласных, определяющих под влиянием сингармонизма характер согласных, входящих в данные слоги.

Материальная реализация языка обладает не только артикуляторно-акустическими характеристиками, но и статистическими свойствами. Реальность этой статистической организации специфична для каждого конкретного языка. Для системы русских согласных характерно одинаковое количество твердых и мягких согласных (по I8 фонем в каждой группе), смычных и щелевых (также по I8), почти одинаковое количество глухих (I6) и звонких (20, включая сонанты), сонорных в системе согласных в 3 раза меньше (всего 9) по сравнению с шумными (27). Переднеязычных в русской системе согласных - I9, губных - I0, заднеязычных - 6, а среднеязычных - I. В туркменском языке также одинаковое количество смычных и щелевых фонем (9:I0). Что же касается количества согласных фонем, коррелирующих по признаку глухости-звонкости, то звонкие согласные, включая сонорные,в 2 раза превосходят (I4:7) глухие. В туркменском языке сонорных в 2 раза (7:I4) меньше, в русском же языке, их в 3 раза меньше. Следующие цифры определяют противопоставление по активно-действующему органу: переднеязычных в туркменской системе - I0, губных - 5, заднеязычных - 4, среднеязычный - I, фарингальный - I.

Определение встречаемости фонем в реальных речевых условиях, представленное в данном случае анализом письменных текстов,явилось следующим этапом статистического представления. Статистические данные, касающиеся русского языка, используются из известных в лингвистической литературе источников /6/. Данные по статистике туркменского языка были получены на основании анализа текстов общим объемом I00000 звуковых единиц. Тексты подбирались с учетом трех основных стилей: художественной прозы, научной прозы и газетной лексики, учитывались все основные случаи ассимиляции, стяжения и выпадения звуков. В процессе работы выявились некоторые трудности единого транскрибирования, спорные вопросы лингвистического характера.

Действующая в туркменском языке губная гармония предполагает наличие огубленности в последующих слогах. Однако в туркменской лингвистике остается спорным вопрос о том, одинаковы ли гласные по степени огубленности во всех слогах слова. А.Моллаев /7/, анализируя акустические параметры гласных в двусложных словах, делает вывод, что /о/ во втором слоге не является чистым огубленным гласным, а имеет лишь оттенок огубленности, поэтому транскрипционно представляется символом /а^{о}/. Полученные же Дж.Гокленовым /8/ данные по восприятию губных гласных убедительно показывают, что лабиализованность вокалических единиц под влиянием губной гармонии достаточно выражается и в последующих слогах. Выделенные гласные опознаются носителями языка вне контекста в 80% случаев и в фонетическом контексте в I00% случаев.

Спорным моментом является и транскрибирование полузвонких согласных типа /t^{d}/, /p^{b}/. В данном случае представлено сочетание двух взрывных звуков, разделенных краткой паузой смычки, причем первый звук имплозивный, а второй - эксплозивный. "Поскольку эксплозия смычных звуков в туркменском языке выражена очень резко, то во многих случаях эксплозивный характер второго звука, следующего за акустически довольно отличным от него имплозивным, заставляет говорящих ошибочно принимать эксплозию согласного звука за его голосность" /4/.

Почти все туркменские согласные могут быть не только краткими, но и долгими. При транскрибировании в первую очередь принималась во внимание морфемная принадлежность, но и одновременно учитывалась двухпиковость и однопиковость изображения согласных на осциллограммах.

Подсчет частоты встречаемости фонем текстов на туркменском языке дал данные, основываясь на которые мы смогли выстроить туркменские фонемы в ряды с убывающей по значениям последовательностью: /l,а,n,r,e, i,ы,о,m,d,t,g,а:,в,ŋ,s,j,h,к,tʃ,ʃ,y,z,и, æ:,i:,р,æ,æ:,w,о:,ы:,dʒ,y:,и,æ,ф,ʒ/.

Большая частотность употребления кратких гласных отмечается как общая закономерность, выявленная как при анализе текста, так и при анализе словаря /9/. Как качественные характеристики гласных, так и количественные четко отражаются на статистическом материале: сочетаемость согласных с гласными заднего ряда составляет 52%, переднего ряда - 42%, среднего ряда - 6%. Согласно общелингвистической тенденции статистически более вероятными являются слова от одного до 3-х слогов. Статистика туркменского языка подтверждает эту тенденцию, причем выделяет двусложные слова, как слова наиболее употребительные в речи. В туркменском языке односложные слова встречаются в I5,6%, двусложные - 40,4% трехсложные 27,7%, четырехсложные I0,5% и далее в порядке резкого убывания. Полученный статистический материал предполагает составление достоверного отрезка речи на русском языке, включающего основные особенности структурной организации звуковых последовательностей туркменского языка.

Слуховой анализ русской речи туркмен был проведен на материале записи текста, составленного в ЛЭФ ЛГУ, включающего 200 наиболее частотных слогов русского языка /I0/ десятью информантами-туркменами. Основные орфоэпические отклонения от русской нормы связаны с системными различиями. Неумение правильно произносить русские редуцированные формы безударных гласных фиксировалось повсеместно. Отмечалось полное, сильное произнесение первого безударного слога, что свойственно произносительным нормам родного языка информантов. Характерная особенность туркменского вокализма - частичная и даже полная редукция узких гласных - прослеживается по всему тексту.Туркменский язык относится к языкам с ярко выраженной вокалической структурой не только в количественном отношении. В силу закона сингармонизма показателем качества слога является гласный. Если слогообразующий гласный относится к заднему ряду, то сочетающийся с ним согласный тоже сдвинут назад. Заднеязычные согласные /к,g/ в речи туркмен передаются их увулярными позиционными аллофонами /q,G/ в комбинации с гласными заднего ряда. Закон гармонии гласных определяет единство тембра гласного и согласного в одном слоге и слове. В сочетании с гласными переднего ряда как предшествующий, так и последующий согласный будут в некоторой степени смягчаться. В сочетании с гласными заднего ряда оба согласных будут веляризованы. Эта отработанная механика речи туркмен хорошо прослеживается на их произношении слов экспериментального текста. Важным условием овладения системой гласных русского языка является консонантная аккомодация на уровне слога, ибо слог, будучи минимальным отрезком речи, заключает в себе характерные взаимодействия гласных и согласных звуков. В русском языке, в зависимости от качества окружающих согласных, гласный может быть продвинут вперед на протяжении всего звучания. В речи информантов-туркмен четко фиксируются свистящие круглощелевые /s,z/, хотя русской речи старшего поколения туркмен свойственна устойчивая замена круглощелевых на интердентальные плоскощелевые звуки /θ,ʒ/, что прогнозировалось системой родного языка. Заимствованная из русского языка лексика может дать материал для появления в системе новых корреляций, уже наблюдаемых в парах /θ-s,ʒ-z/ /I0/. Замена аффрикат на щелевой, тенденция упрощения свойственна речи туркмен. Следует отметить, что системность и универсальность некоторых фонетических реализаций достаточно часто не срабатывает. В речевой деятельности наиболее автоматизированными, а следовательно наиболее труднопреодолимыми являются ошибки на произносительном уровне. Артикуляторный уклад туркменского языка, его механизмы, значительно противопоставлены русскому. Свойственное русскому языку передне-верхнее выпуклое положение языка свидетельствует, что палатализованность является одной из важнейших особенностей русской фонетики. Глубокое и плоское положение языка туркмен позволяет сделать вывод о большей степени бемольности туркменской артикуляторной базы. Следует отметить и активное участие губ, увулы и гортани в речи туркмен.

Приводимые данные могут быть использованы как в теоретическом плане при осмыслении типологических проблем фонологии, так и в практическом отношении при обучении русскому языку учащихся - туркмен.

Список использованной литературы

I. И.П. Зинченко. Национальный состав населения СССР. В кн.: - Всесоюзная перепись населения I979 г. М., I984.
2. О. Назаров. Функционирование русского языка в Туркмении. - В кн.: Русский язык в национальных республиках Советского Союза. М., I980.
3. Итоги Всесоюзной переписи населения I970 года, т.4. М., I973.
4. А.П. Поцелуевский. Фонетический строй туркменского языка. Диалекты туркменского языка. В кн.: Избранные труды. Ашхабад, I975.
5. Л.А. Вербицкая. Русская орфоэпия. Л., I976.
6. Л.В. Бондарко, Л.Р. Зиндер, А.С. Штерн. Некоторые статистические характеристики русской речи. - В кн.: Слух и речь в норме и патологии. Л., I977.
7. А.Моллаев. Акустическая характеристика ударных и безударных гласных в двухсложных словах туркменского языка. Ашхабад, I980.
8. Дж. Гокленов. Безударные гласные в многосложных словах туркменского языка. Ашхабад, I984, кандидатская диссертация.
9. Орфоэпический словарь туркменского языка. Ашхабад, I978.
I0.Динамика структуры современного русского языка. Л., I982.

БОЛГАРСКИЙ АКЦЕНТ В РУССКОМ ПРОИЗНОШЕНИИ У НОСИТЕЛЕЙ РУССКОГО ЯЗЫКА, ПОСТОЯННО ПРОЖИВАЮЩИХ В БОЛГАРИИ

ВАЛЕРИЯ Г. СМИРНОВА

I Кафедра русского языка
Софийского государственного
университета,София,Болгария

РЕЗЮМЕ

Приводятся наблюдения над интерференцией в области русского произношения у диктороз - носителей русского языка в условиях русско-болгарского двуязычия. Отклонения от произносительной нормы у русских дикторов совпадают с основными трудностями, которые испытывают болгарские русисты при изучении русской фонетики и интонации.По мнению автора,с данными отклонениями /болгарским акцентом/ можно и нужно бороться путем создания периодических корректировочных курсов по русскому произношению,хотя бы для русских переводчиков /устный перевод/ и дикторов,а также путем увеличения контактов с живым русским словом /кино,радио,телевидение, грамзаписи, посещения СССР/

ВСТУПЛЕНИЕ

Общеизвестно,что если долго или постоянно живешь в иноязычной среде /т.е. в условиях двуязычия/,родной язык останавливается в своем развитии или подвергается интерференции в той или иной степени и прежде всего в области произносительных норм. Зачастую иностранный акцент в речи русского или болгарина,проживающего за рубежом, раздражает ухо сильнее, чем далеко не совершенная речь иностранца на русском или болгарском языках, которая подчас звучит даже "мило" и экзотично /за исключением разве что оперного певца, исполняющего арию Ленского с немецким акцентом. Здесь любое отклонение от произносительных норм языка грозит разрушить все очарование как от прекрасной музыки, так и от прекрасного голоса/.

СОБСТВЕННЫЙ ОПЫТ

На протяжении многих лет автору данного сообщения приходилось сталкиваться с проблемами болгарского акцента,сначала как преподавателю Софийского университета в работе с болгарскими русистами по русской фонетике и интонации /экспериментальный контрастивно-сопоставительный курс,руководимый М.А.Георгиевой на основе ее же и М.Поповой учебника [1],а впоследствии как консультанту по русскому произношению при записи учебников русского языка для русских и болгарских школ в Болгарии на Софийском радио. В результате проведенных исследований /акустических и аудитивных/ и наблюдений автор приходит к выводу о том, что основные ошибки в русском произношении у болгарских студентов-русистов и у дикторов - носителей русского языка совпадают:

I. На уровне гласных

1. Назализация гласных перед н в абсолютном конце слова или в сочетании с другими согласными /вагон, блин, талант, институт, оранжевый, конкурс/;
2. недостаточная лабиализация у в абсолютном начале слова /университет, уподобление/;
3. отсутствие дифтонгоидного элемента, слабая губная реализация у о;
4. отсутствие редукции э в абсолютном начале слова /энергия, экскурсия/;
5. ъ вместо ʌ на месте о и а в первом предударном и в абсолютном начале слова /Москва, толпа, они/;
6. ʌ вместо ъ на месте о и а во втором предударном /погоди, паровоз/;
7. отсутствие редукции безударного э - "эканье" /ветерок, передать/.

II. На уровне согласных

1. Недостаточная степень смягчения /"полусмягчение"/ согласных, в особенности н', л', р', т', д'; отсутствие свистящего призвука в т' и д'во всех позициях, в особенности,в абсолютном конце слова /птицы, дети, брать/;
2. болгарские ш и ж вместо русских твердых ш и ж /Миша, Сашка, жестокий/;
3. болгарское твердое ч вместо русского мягкого ч'/честь, точка/;
4. отсутствие или ослабленная артикуляция йот /белеет, уезжает/;
5. неправильная диссимиляция - шн и шт вместо чн и чт /конечный,точно,почти/;
6. болгарское дж вместо русского сочетания дж /пиджак, таджик/;
7. произношение без выпадения т и в сочетаний стн, стл, стск, вств/радостный, совестливый, марксистский, чувство, здравствуй/.

III. На стыке слов и морфем

1. Оглушение в перед сонорными и гласными /в Ленинград, в университете/;
2. и на месте ы/днем и ночью/;
3. сдвиг ударения со знаменательного слова на служебное /уж очень, без ошибок/, с глагола на отрицание /не спутаешь/;
4. побуквенное произношение сочетаний сч, зч и зж /с чашкой, возчик, без жизни/.

IV. На интонационном уровне

1. Отсутствие понижения тона до нуля звука в повествовательном и вопросительном предложениях с вопросительным словом;
2. недостаточное повышение тона ударного гласного предиката вопроса в вопросительном предложении без вопросительного слова и, наоборот, повышение тона на заударных слогах, в особенности, в конце фразы;
3. сдвиг интонационного центра в неполном вопросительном предложении с союзом "а" и в незавершенных синтагмах, содержащих препозитивное определение или обстоятельство, а также причастный или деепричастный обороты;
4. немотивированная паузация /нарушение компенсационного закона Пешковского [2]/;
5. отсутствие эмоционально-экспрессивной окраски и повышения тона в восклицательном предложении /срв. "восклицательное" предложение и "удивително" изречение - болг./.

Таким образом в работе с дикторами - носителями русского языка так же, как и в работе с болгарскими студентами-русистами, подтвердилась истина о высокой устойчивости болгарского акцента [3] , объясняемой родственной близостью русского и болгарского языков, когда происходит механический перенос из одного языка - болгарского в другой - русский. Тип данного двуязычия есть нечто среднее между координативным и субординативным билингвизмом [4]. С точки зрения языковой нормы - это, скорее всего, ненормативный билингвизм /здесь нарушаются некоторые произносительные нормы родного языка при далеко не совершенном "практически-бытовом" владении вторым языком/. Надо отметить, что дикторы представляли собой группу носителей русского языка, закончивших высшее образование в СССР и проживающих в Болгарии от 15 до 30 лет /двое артистов, журналистка, инженер - руководитель русского театрального кружка/. Их основной "рабочий" язык - болгарский, однако в течение многих лет они принимают активное участие в передачах на русском языке на телевидении и радио в Болгарии. Очевидно, бо́льшая напряженность, энергичность и "твердость" болгарской артикуляции оказывает влияние на их артикуляционный аппарат, хотя поначалу они не отдают себе в этом отчета. Оттого, вероятно, на начальном этапе работы дикторы допускали гораздо больше отклонений от произносительной нормы, чем впоследствии, когда им были объяснены "правила соотнесения контактирующих языков" /В.Ю. Розенцвейг [5] /, т.е. сопоставлены или противопоставлены - на основе учебника М.А. Георгиевой и М. Поповой - русская и болгарская фонетика и интонация. Это обстоятельство еще раз подтверждает высказывание А.Е. Супруна [6] и других ученых о том, что "реальным источником интерференции является именно совмещенный механизм порождения текстов на двух языках, недифференцированность установок".

Думается, что болгарский акцент в русском языке можно и необходимо преодолевать с помощью периодических корректировочных курсов по русской фонетике и интонации, основной целью которых является противопоставление, дифференциация механизмов контактирующих языков. Огромную роль играют здесь также контакты с живой русской речью /советское кино, радио и телевидение/, а также самоконтроль и самокоррекция.

ЛИТЕРАТУРА

[1] М.А. Георгиева и М.Попова. Русская фонетика и интонация. София, 1974.
[2] Т.М. Николаева. Фразовая интонация славянских языков. Москва, 1977.
[3] Ю. Г. Лебедева. Звуки, ударение, интонация. Москва, 1975.
[4] К.Г. Бабов. Проблемы интерференции в процессе обучения русскому языку в болгарской школе. София, 1974.
[5] В.Ю. Розенцвейг. Языковые контакты. Ленинград, 1972.
[6] А.Е. Супрун. Лингвистические основы изучения грамматики русского языка в белорусской школе. Минск, 1974.
[7] Д. Тилков и Т. Бояджиев. Българска фонетика. София, 1977.
[8] Ст. Ив. Стойков. Увод във фонетиката на българския език. София, 1966.

THE EFFECTS OF FOREIGN LANGUAGES PROFICIENCY AND COGNITIVE PROCESSING ON TEMPORAL STRUCTURE OF SPEECH

E.L. NOSSENKO

Dept. of English Philology
Dniepropetrovsk State University
Dniepropetrovsk, Ukraine, USSR 320625

ABSTRACT

Oral proficiency in a foreign language can be automatically evaluated by using as an interface to computer an electronic speech analyzer designed for detecting unfilled pauses in speech. Proficiency evaluation is based on a psycholinguistic approach to the analysis of cognitive processing in speech through its reflection in the temporal structure of oral discourses.

INTRODUCTION

A number of studies of the temporal structure of oral discourses produced by the speakers in their mother tongue showed the relevance of pauses in spontaneous speech to its cognitive processing [1]. The analysis of the location of these pauses in speech has revealed that they tend to precede relatively unpredictable lexical items [2] and occur with greater than chance frequency at the beginning of phonemic clauses [3]. It has also been shown that pauses are associated with intuitively determined "idea boundaries" in speech text [4]. On the basis of these findings it has been concluded that pauses in spontaneous speech are used for lexical selection, holistic planning of phonemic clauses, suprasegmental ideational planning.
Further research of the temporal organization of spontaneous speech showed that cognitive nature of pauses predetermines a universal pattern of their distribution in speech of the native speakers of different languages with "hesitant" and "fluent" phases following each other in cycles [5]. The hesitant phases are characterized by longer and more frequent pauses and it was hypothesised that during these pauses speakers make anticipatory decisions of what to say next. The fluent phases of sequentially alternating cycles reflect the execution of semantic plans formulated in hesitant phases. These phases are not hesitation free either, but, unlike hesitant phases, in which pauses may be distributed randomly, fluent phases are characterized by the localization of pauses at grammatical boundaries and at the points of lexical selection, where this selection is guided by a preformulated semantic plan. Inspite of the predominantly random distribution of pauses in hesitant phases their cognitive nature leaves no doubt. It has been demonstrated experimentally that the mean hesitancy of hesitant phases cannot be diminished without causing a decrement in the quality of the utterance (its ideational content) [6]. Attempts to make speakers consciously modify the number and length of pauses in spontaneous speech under experimental conditions resulted in the substitution of pauses by other hesitation phenomena: "filled pauses" (non-phonological vocalic sequences "er", "hm" and the like), repeats, false starts etc. It proves that pauses are necessary for speech planning and that there are no modifiable noncognitive pauses in spontaneous monologue.
This conclusion is also supported by the finding that there is a positive relationship between the duration of speech pauses and difficulty or abstractness of experimental tasks involving speech production [7]. As F. Goldman-Eisler demonstrated, oral descriptions of concrete events (cartoon-stories) are accompanied by shorter pauses than the formulation of their meaning.
As a result of the above reviewed studies it became clear that pauses in speech can be regarded as manifestation of the more general blocking of activity, which occurs when organisms are confronted with situations of uncertainty, i.e. when selection of the next step requires an act of choice. When a person speaks his mother tongue, the situations of uncertainty arise mainly on the ideational level of speech production. The speaker of a foreign language, whose vocabulary diversity is naturally much more limited than that of the native speaker and who, unlike the native speaker, doesn't feel the so called "transition probabilities", i.e. words combinability

rules, is expected to pause longer at the points of lexical selection. For him the syntactical structuring of speech also remains for a long time a conscious time-consuming act of recollecting appropriate grammar rules. Therefore his total pause time in speech will include not only the time, necessary for semantic planning, but that, required for lexical selection and grammatical processing of speech, i.e. for fulfilling the operations, which are performed by native speakers to a great extent automatically.
In addition to the above formulated assumptions we also claim that the degree of continuity of speech utterance as measured by lengths of phrases pronounced without hesitation pauses (i.e. pauses, exceeding in length 250 ms) is positively related to the level of foreign languages proficiency. The higher is the level of oral proficiency in a foreign language the less frequently the speaker has to interrupt his speech by pauses and thus the higher is the degree of continuity of his speech utterance.
This claim is based on the following data. While the difference of level of verbal planning in spontaneous speech of the native speakers of a language, as exemplified, for instance, by the description of concrete events and the formulation of their meaning, is clearly reflected in the length of pauses, it is n o t reflected in the degree of continuity of speech utterance as measured by lengths of phrases, uninterrupted by pauses [7]. This experimental finding made it possible to conclude that phrase length is an indicator of how automatic the utterance of speech sequence has become and to which extent its production ceased to be a conscious effort.
The importance of this observation for the solution of the task of automatic objective evaluation of oral proficiency in a foreign language cannot be overestimated. It is just the degree, to which speech utterance in a foreign language gradually ceases to be a conscious effort, that we are after, when trying to evaluate the progress, made by the learner in the process of mastering a foreign language.

HYPOTHESES

1. Since all pauses displayed in spontaneous monologues of the native speakers of a language are of cognitive nature, i.e. necessary for speech planning (except, perhaps, brief pauses at grammatical junctures), their length can be regarded as reflecting the degree of concious efforts on the part of the speakers in the process of speech generation.

2. Since the difference in levels of verbal planning in spontaneous speech of the native speakers of a language is clearly reflected in the length of pauses and not reflected in the degree of continuity of speech utterance as measured by lengths of phrases uninterrupted by pauses, the latter can be regarded as indicating habit strength entering into production of speech.

3. Since different levels of foreign languages proficiency involve either essentially automatic or mainly cognitive lexico-grammatical processing of oral discourses, temporal patterns of these discourses will also differ.

EXPERIMENT

The hypotheses were tested by comparing temporal structures of spontaneous and preliminary planned oral discourses produced by speakers with different levels of foreign languages proficiency.
The experimental procedure involved asking the subjects to speak on a suggested topic first spontaneously and then after preliminary preparation. The speakers were not restricted in time and speech was a monologue. The subjects were chosen in such a way as to form distinctly different proficiency groups (foreign languages instructors - university students majoring in a foreign language - university students doing a foreign language on a non-professional level).
Tape-recorded speech was subjected to computer analysis through an interface designed to classify as "pauses" interruptions of the vocal utterance of not less than 250 ms. Breaks less than 250 ms were not counted and were included into vocalization time.
The choice of 250 ms time interval as a border-line between pause/non-pause sequences was predetermined by earlier observations, reported by F. Goldman-Eisler in her above-reviewed papers, that pauses longer than 250 ms cannot occur within a word boundary without violating its integrity as a speech unit.
Temporal structure of oral discourses recorded in the process of the experiment was described in the following terms:

- mean phonation time/pause time ratio;
- mean phonation time/frequency of pauses ratio;
- percentage of pauses exceeding 1000 ms (number of pauses exceeding 1000ms/number of all pauses in speech);
- mean phonation time/pause time ratio variance (per cycles of 15 s).

Mean phonation time/pause time ratio was chosen as an indicator of the degree of hesitancy in speech. It was expected to be the measure of concious efforts on the part of the speaker related to semantic planning of speech and to its syntactical structuring (if the latter is not automatized).

Mean phonation time/frequency of pauses ratio denoted the continuity of speech utterance, or the mean length of a word sequence, uttered continuously (without pauses equal to or exceeding 250 ms). This was expected to indicate habit strength, entering into production of speech. Percentage of long pauses (exceeding 1000 ms) was expected to be an additional measure of difficulties, experienced by speakers in the process of encoding. These difficulties were believed to be of linguistic rather than extralinguistic nature, since mean pause length in spontaneous speech of the native speakers of a language rarely exceeds 800 ms if they do not experience anxiety or emotional stress [8]. Mean phonation time/pause time ratio variance was chosen to designate sequential temporal patterning of speech. The higher variance was alleged to be an indicator of spontaneity of speech and the lower - of its previous preparation. The interval of 15 s was chosen not randomly, but on the basis of the observation that "hesitant" and "fluent" phases in spontaneous speech make cycles lasting approximately for 10-20 s.

RESULTS

Oral discourses of the speakers, included into a high-level-of-foreign languages proficiency group, appeared to be characterized by the highest mean phonation time/pause time ratio; longest continuity of speech utterance and lowest percentage of pauses, exceeding 1000 ms, as compared with the corresponding data for other groups. High level of foreign languages proficiency also finds reflection in a distinct sequential temporal patterning of spontaneous discourses and in a more even distribution of pauses/phonation ratio throughout a preliminary planned discourse.

The lower is the level of foreign-languages proficiency, the more deteriorated is sequential temporal patterning of spontaneous speech, the lower is a mean phonation time/pause time ratio, the shorter are phrases, sandwiched between pauses, the higher is the percentage of long pauses in speech. The analysis of visual transformations of oral speech synchronized with the verbal content of the records showed that the lowest level of foreign languages proficiency is associated with the incidence of hesitation pauses within the boundaries of phonemic clauses not only in hesitant phases, but in what is expected to be a "fluent" phase. Pauses of the same duration at grammatical junctures, found in the speech of the subjects with higher level of oral proficiency, are less detrimental to the textual cohesion of speech. The gain in fluency of preliminary planned speech, as compared with spontaneous one, is also characteristic of the level of foreign languages proficiency.

The more pronounced is the difference in temporal structures of spontaneous and preliminary planned discourses, the higher is foreign languages proficiency.

It is worthwhile mentioning that in the preliminary planned (thought over) speech of the foreign language learners, who have not yet achieved a sufficiently high degree of proficiency, the duration of hesitation pauses may decrease in the "hesitant" phases, during which the speakers make decisions of the "w h a t -to- say" type, while in the "fluent" phases pauses remain fairly long. The thing is the speakers with low degree of foreign languages proficiency cannot perform lexico-grammatical processing of their oral discourses automatically enough. That is why the variance of the mean duration of hesitation pauses in their speech is less pronounced than in the speech of the learners with higher degree of foreign languages proficiency.

The above described data revealed a statistically significant difference in the temporal structures of oral discourses, produced by the speakers with considerable difference in foreign languages proficiency. It is evident that foreign languages instructors have undoubtedly better command of the language, they teach, than the students, who are being taught, particularly those, who do not major in a foreign language and who take it as a minor subject.

Will temporal structures of oral discourses differ significantly if the levels of foreign languages proficiency of the speakers are not as strikingly different, as in the case referred to above? To answer this question we have carried out an additional experiment. This time the subjects were university applicants, seeking admission to the foreign languages division of the University.

Their proficiency in a foreign language was naturally not known beforehand and had to be evaluated on the basis of the results of computer-assisted language testing.

Tests were designed to evaluate reading comprehension, grammatical competence and diversity of the active vocabulary of the prospective students.

Those of them, who were admitted to the university on the basis of the results of entrance examinations, were later on subjected to another foreign languages proficiency testing, which involved listening comprehension tests and computer-assisted analysis of the temporal structure of oral discourses of the subjects.

In accordance with the results of proficiency testing the subjects were split into three proficiency groups: "A", "B" and "C". The subjects, included into group "A" (50 persons), fulfilled 92% of the total number of examination assignments correctly. Those, included into group "B" (also 50 persons), scored lower. The average percentage of correct answers in this group was equal to 77%. The subjects, who made up group "C" (50 persons), managed to fulfil correctly only 55% of all examination assignments suggested.

The difference between the results of group "A" and "B" turned out to be statistically significant at $p < 0.003$ and of groups "B" and "C" - at $p < 0.247$.

Similar distribution holds true of each separate type of examination tests. For example, average reading comprehension test scores for the subjects of group "A" were equal to 90.1%, for group "B" - 78% and for group "C" - 47.5%. Average grammatical competence test scores displayed a similar picture with 89.2% of correct answers in group "A", 78.2% - in group "B" and 50.7% - in group "C". The same distribution pattern is also true of vocabulary test scores with 92.2% of correct answers in group "A", 71.6% - in group "B" and 54.8% - in group "C".

Auditory comprehension tests also revealed statistically significant difference among the above mentioned proficiency groups. The subjects, who made up group "A", followed 85% of the units of information, conventionally singled out in the text for auditory comprehention, of group "B" - 66.3% and of group "C" - 55.1%.

Thus the difference in levels of foreign languages proficiency among the above described groups of subjects left no doubt.

When we analyzed the temporal structures of oral discourses, elicited from the same subjects upon their admission to the University, we obtained the following data. Mean phonation time/pause time ratio in spontaneous oral discourses of 50 subjects, included into group "A", appeared to equal 1.34. The subjects, making up group "B", paused longer. Their mean phonation time/pause time ratio turned out to equal only 0.63. Group "C" ranked still lower - with 0.21 mean phonation time/pause time ratio.

The continuity of utterance (mean phrase length as measured by phonation time/frequency of pauses ratio) also revealed difference among the three proficiency groups mentioned above. It ranged from 0.73 s for the group "A" subjects to 0.61 s - for group "B" and 0.39 s - for group "C". Average percentage of long pauses (exceeding 1000 ms) in oral discourses of different proficiency groups also appeared to be significantly different: 24.4% - for group "A", 34.7% - for group "B" and 69.7% - for group "C". It is noteworthy that the corresponding data characterizing the temporal peculiarities of oral speech of the same subjects in their mother tongue revealed practically no difference among the above mentioned three groups of subjects. Suffice it to mention that mean phonation time/pause time ratio in oral discourses of the group "A" subjects appeared to equal 1.98; of group "B" - 1.82; and of group "C" - 1.83. Accordingly the continuity of utterance is equal to 0.84 s - for group "A"; 0.79 s - for group "B"; 0.83 s - for group "C". Percentage of long pauses varies from 18.3 to 18.7 for all experimental groups.

CONCLUSION

Temporal structure of oral speech as analyzed in terms of continuity of speech utterance, phonation time/pause time ratio and frequency of pauses of various length is indicative both of the level of foreign languages proficiency, gained by the learner, and of spontaneity of his speech utterance. Since temporal peculiarities of oral speech can be easily subjected to computer analysis the method of foreign languages proficiency evaluation, based on the assessment of temporal structure of oral discourses, appears to be most economic.

REFERENCES

1 F. Goldman-Eisler, Hesitation, Information and Levels of Speech Production. - "Ciba Foundation Symposium", L., 1964, p. 96-111.

2 F. Goldman-Eisler, Speech Production and the Predictability of Words in Context. - Quart. Journ. Experim. Psychol., 1958, v. 10, p. 2.

3 D. Boomer, Hesitation and Grammatical Encoding. - Lang.Speech, 1965, v.8, p. 3.

4 B. Butterworth, Hesitation and Semantic Planning in Speech. - Journ. Psycholing. Res., 1975, v. 1, p. 75-78.

5 F. Goldman-Eisler, Sequential Temporal Patterns and Cognitive Process in Speech. - Lang. Speech, 1967, v. 10, p. 122-132; F. Goldman-Eisler, Pauses, Clauses, Sentences. - Lang. Speech, 1972, v. 15, p. 103-113.

6. G. Beattie, R. Bradbury. An Experimental Investigation of the Modifiability of the Temporal Structure of Spontaneous Speech. - Journ. Psycholing. Res., 1979, v. 8, p. 225-248.

7 F. Goldman-Eisler. Continuity of Speech Utterance, Its Determinants and Its Significance. - Lang. Speech, 1961, v. 4, p. 4.

8 K. Scherer, Speech and Emotional States. - In: "Speech Evaluation in Psychiatry", N.Y., 1981, p. 189-220.

ARTIFICIAL INTELLIGENCE APPROACHES TO COMPUTER-BASED ENGLISH TEACHING: A TUTORIAL SPEECH RECOGNITION

YELENA Y. DASHKO

Dept. of Foreign Languages
Mining Institute
Leningrad, USSR, 199026

ABSTRACT

With the increasing role of computers in teaching, there is little doubt that we will eventually want them to talk to us and allow us to speak to them. This review presents an experimental approach of voice I/O techniques to computer-based teaching English on the basis of Expert Type System. Advanced applications of the speech processing technology and some special linguistic/information problems are discussed.

INTRODUCTION

The (micro)computer in education has both stimulated research on linguistic database and has provided a more precise experimental vehicle for controlling the presentation of instruction and measuring responses /1/.

The basic hypothesis is that the computer offers an opportunity:

(i) to provide the teacher with a powerful resource to manage individual learning within the terminal room;
(ii) to enable the student to follow learning procedures which incorporate step-by-step feedback and to stimulate individual attention and to gain assistance discreetly;
(iii) to make it possible to the teacher to observe and monitor the progress of the student in detail.

The objectives of the preliminary study are, firstly, to investigate the feasibility of the approach and the appropriateness of the software facilities being employed.

AUTOMATED LEARNING SYSTEMS

Evidence indicates that it is most productive to teach grammatical and lexical bases in context. In order to carry out the exercise the student must thus draw upon reading (knowing) English letters, vocabulary.

In computer-based English teaching the student receives all of his training from the display device including tests and performance feedback. Variations and combinations of instructional arrangements are not uncommon. The computer-based study management model usually employs existing materials and the student spends only a part of his study time interacting with the computer; the material used in the initial study provides 20-40 minutes work for students. Testing and teaching may be done on the computer via keyboard input by students. Responses may be saved on diskette and at the end of testing the student's score can be displayed on the screen immediately, or a printed evaluation can be reproduced depending on the computer programme and the test construction. Questions to be printed are chosen from a master list and their descriptors (up to 1000 items).

Each Automated Learning System (ALS) comprises several learning volumes: Training volume; Control volume with its priority scoring due to three levels of complexity, grammatical and lexical databases; Reference volume with its grammar and lexicon retrieval; Encouragement volume.

Potential linguistic problems are worked out before they are translated into hardware and software. Needs, goals, constraints are described first, thus helping a complex problem to be divided into a hierarchy of simpler problems to understand/control the whole process of learning. Each training step is related to previous and subsequent stages, and misunderstandings among students are avoided. Questions/answers from the packet of exercises in Control volume should be typed with a 'minimum-energy' solution.

ALS AND DIALOG INTELLECTUAL EXPERT SYSTEM (DIS-332)

Due to the technological development of voice recognition systems voice input can be embodied in ALS. Voice input gives the chance to relieve the overloaded visual/manual channel and may achieve a natural form of human-computer communication.

A major problem is how to integrate the different sources of knowledge in such a way as to exploit their interaction. One can identify the following conceptually distinct sources of knowledge as important in determining the interpretation of a spoken utterance/2/:

1. Segmentation, Feature Extraction and Labelling - processes of detecting acoustic-phonetic events in the speech signal and characterizing the nature of individual segments of the signal.

2. Lexical Retrieval - a process of retrieving candidate words from the lexicon base that are acoustically similar to the labelled segments.

3. Word Matching - a process of determining some measure of the goodness of a word hypothesis at a given point in the speech signal.

4. Syntax - the ability to determine if a given sequence of words is a possible subpart of a grammatical sentence and to predict possible continuations for such sentence fragments.

5. Semantics - the ability to determine if a sentence is appropriate to the context in which it is uttered, and what has been said previously in the discourse.

A variety of different approaches have been explored on the basis of the Dialog Intellectual Expert System (DIS-332)/3/. Generally, they fall into top-down and bottom-up strategies, where a single network parser combined syntactic, semantic and pragmatic components on the basis of Hidden Markov Models, is presented.
In these syntactically constrained tasks performance results ought to be reported with the following information: a) complete description of the domain grammar including full specification of the vocabulary in the form of scripts, b) frequencies of lexical units transitions from each task state to successive states and c) average branching factor.

It is important to distinguish between the total vocabulary capacity and the branching factor, the average number of words which must actually be discriminated at each stage of the task (sometimes referred to as the size of the average active subvocabulary). DIS-332 includes a study of extensive database collections of both isolated and connected utterances, spoken by ten Russian speakers.

Vocabulary size = 500 words, performance of the recognizer = 98-99%, recognition time = real, branching factor $\leqslant$ 10.

Above mentioned configuration provides a large scale of opportunities while using in ALS:

1. Russian lexical vocabulary and corresponding English items Input;

2. Printed text Visualization;

3. Impartial Control upon English words learning;

4. **Sounding for each input word (speech synthesis);**

5. Voice input ($\leqslant$500 words) for English spelling correction of phonemic baseforms in training and recognition mode;

6. Voice input of English sentences (sentence length $\leqslant$15) in training and recognition mode;

7. Impartial Control upon learned grammatical/lexical English level;

8. Reference information output due to the error rate or to inquiry.

LINGUISTIC AND INFORMATION PROBLEMS OF VOICE I/O

Finally, prediction models of recognizer performance can be used in determining optimal operating conditions for voice input /4/. A major problem is the identification of those factors having significant influence on the performance of the speech recognizer. W.A.Lea (1982) has compiled an extensive list of more than 80 variables including language and task factors (number of training passes, reject threshold, size of the active vocabulary, inter-word confusability), human factors (sex of the speaker), algoritmic factors, channel and environmental factors (microphone type and position), performance factors (type of feedback, error correction).

Ergonomic aspects for improving recogni3 tion performance should include:

a) a short speaker training of several hours is necessary;

b) if possible, a 3-5 repetitions in system training should be carried out;

c) equally-positioned phonemes of the vocabulary should be out of different articulation types;

d) vocabularies should be splitted down even in smaller subvocabularies than specified;

e) system training must be performed with the operational noise at minimum.

There was found an improvement for DIS-332 from 96% to 98-99% when three instead of one word repetition was chosen. Summing up, it provides a sufficient variety of coarticulatory environments. 5 - 6 word repetitions brought no further improvements which is not surprising in view of the high level of the recognition rate. Generally, it is suspected that the necessary number of word repetition during system training mode between about 3 to 6 depends positively on vocabulary size, complexity and confusability. The results show that equally-positioned phonemes can be better distinguished from among one another when different articulation types are used (e.g. plosive - fricative). Such features as voicing, nasality, affrication, duration and place of articulation are the primary channels of the intelligibility-relevant information. Different vocals tend to be good features of words to be recognized if they do not belong simultaneously to the high vowels "e" and "i" and to the deep vowels "o" and "u". We believe that for each doubling of the vocabulary size, the recognition accuracy tends to decrease by a fixed amount, which is different for each talker.

Yet there can hardly be any more important task in speech recognition than determining now well algorithms or devices work. Thus the error rate as a performance measure conveys no information about performance except the relative number of errors made on a given task. It tells nothing about the distribution of errors and the costs of making particular errors; depends on vocabulary size and doesn't reflect large vocabulary difficulty, the inherent acoustic confusability, the difficulty of the speaker, or the environment. A new information-theoretic performance measure is based, in part, on the idea that automatic as well as human speech recognition systems can be modelled as communication channels. A more meaningful measure, called the Relative Information Loss (RIL) would normalize the amount of information lost in a recognition process with the amount transmitted/5/. Woodard and Nelson /6/ propose combining the 'Human Equivalent Noise Reference' (HENR) method with a RIL method. HENR is based on the confusions between speech sounds by humans listening in noise. The model predicts the percentage word recognition rate, and the confusions at any signal to noise ratio for any vocabulary which has been defined in phonetic terms. This combined approach may be used to relate device performance to task difficulty.

Grammatical constraints whether they will be stochastic or deterministic, have the effect of decreasing entropy, increasing redundancy and hence decreasing error rate (entropy is, of course, a statistical property). Each natural language requires that some assumption be made about the likelihood of occurence of trained difficulty at a given point in sentence.

Two reasonable assumptions are that the difficulties are equiprobable or distributed to maximize entropy. Under the medium entropy assumption,

$$H_{eq} = \frac{\log_2 |L(G)|}{E_1(G)\{W\}}$$

so that entropy in bits/word is the base-two logarithm of the size of the language divided by average sentence length. For ALS redundancy is to be increased up to 20% against existing Expert Systems.

CONCLUSIONS

The development of faster microprocessors, larger memories, better printers and storage devices, together with pricing competition, will play roles, too. But the factor likely to be judged most significant in the academic microcomputer revolution will probably be the rate at which these recognition systems have gained widespread acceptance by humans in serving their diverse educational needs on the basis of ALS.

REFERENCES

/1/W.D.Fattig, "Microcomputers in Academia", Journal of the Alabama Academy of Science, vol.55, No.1, Jan.1984, pp.31-37

/2/"Computer Speech Processing" ed.by F.Fallside, W.A.Woods (UK) LTD, London, 1985

/3/Петров А.Н., Туркин В.Н. "Система речевого диалога", Автоматическое распознавание речевых образов, ч.II, Каунас, 1986, с.97-99

/4/H.Mutschler, "Ergonomic aspects for improving recognition performance of voice input systems", IFAC Analysis, Design and Evaluation of Man-Machine Systems, Baden-Baden, 1982, pp.261-267

/5/Th.M.Spine, B.H.Williges, J.F.Mainard, "An economical approach to modeling speech recognition accuracy", Int.J. Man-Machine Studies, 1984, 21, pp.191-202

/6/J.M.Baker, D.S.Pallett, J.S.Bridle, "Speech recognition performance assessments and available databases", Proceedings ICASSP-83, Boston, pp.527-532

ИСПОЛЬЗОВАНИЕ АВТОМАТИЗИРОВАННЫХ ОБУЧАЮЩИХ СИСТЕМ (АОС) ДЛЯ ОВЛАДЕНИЯ ЗВУКОВЫМ СТРОЕМ НЕРОДНОГО ЯЗЫКА

Д.Н.АНТОНОВА

Подготовительный факультет для иностранных граждан ,МГУ, Москва, СССР, 117234

Л.С.КРАСНОВА

Филологический факультет МГУ, Москва,СССР, 117234

В настоящее время все больше внимания уделяется изучению звуковой стороны языков. Актуальным при обучении неродному языку становится требование достижения произносительных навыков, приближающихся к уровню навыков носителей языка.

Обучение орфоэпически правильному, безакцентному произношению сопряжено с целым рядом трудностей, главная из которых – звуковая интерференция. Иноязычный акцент в русской речи иностраннных учащихся представляет собой систему устойчивых навыков неправильного произношения, сложившихся под влиянием родного языка, что определялось А.А.Реформатским как подчинение чужой фонетики фонологическим навыкам родного языка. Таким образом, характер акцента зависит от контактирующих языков, что позволяет говорить о болгарском, немецком, венгерском, японском и других акцентах.

Эффективность фонетического курса во многом определяется отбором учебного материала, учитывающим данные сопоставительного анализа звуковых систем взаимодействующих языков, а также анализ ошибок учащихся и трудностей, связанных с усвоением правильного произношения.

Условия обучения, особенно при отсутствии языковой среды таковы, что не позволяют обеспечить фонетически правильное оформление речи. Как показывает практика, устоявшиеся приемы преподавания русского языка как неродного недостаточны. Требуется поиск новых путей преподавания и большее,чем раньше, внимание к техническим средствам обучения (ТСО) и, в первую очередь, компьютерным.

В основе успешного применения компьютерных средств в обучении звуковому строю неродного языка лежат те же общеметодические принципы, что и при традиционном обучении, а именно: системность, всесторонний учет родного языка, коммун икативность, наглядность, и другие. Однако компьютерные средства дают качественно новую ступень развития некоторым из них. Так принцип коммуникативности воплощается здесь в реальное диалоговое взаимодействие обучаемого с компьютером, важным следствием которого является индивидуализация обучения. Помимо различных видов наглядности , которые при этом могут органично сочетаться, компьютерная техника позволяет осуществить процесс визуализации звучащей речи и обеспечить эффективный контроль за постановкой нормативного произношения на всех этапах обучения.

Для обеспечения прочного усвоения формируемых произносительных навыков необходима достаточная повторяемость материала и обратная связь. Наличие обратной связи которую дают компьютерные средства (ответ принимается как правильный или не принимается при наличии ошибок), выводит на принципиально новый уровень возможности развития речевого слуха.

Качественное улучшение произношения невозможно без развития речевого слуха. Речевой слух, представляющий сочетание разных форм абсолютной и дифференциальной чувствительности (собственно речевой слух и тональный), является основой формирования правильного произносительного навыка. Специфические качества речевых звуков первоначально неразличимы в силу отсутствия в рецептирующей системе моторного звена, адекватного отражаемому качеству звука (А.Н.Леонтьев,1972). Тренинг с обратной связью способствует становлению слухового образа, который играет двоякую роль: роль эталона и артикуляторной программы.

Широкое использование магнитофонов и лингафонных кабинетов в настоящее время стало традиционным и принесло свои плоды. Однако следует учитывать сложность биомеханической структуры продуцирования речи, быстротечность артикуляции звуков, осуществляемой в миллисекундные интервалы, что создает часто непреодолимые препятствия для обучаемых при воспроизведении требуемых произносительных движений с целью достижения того или иного слухового эффекта. Это обстоятельство обосновывает необходимость привлечения новых модальностей ощущения и разработку соответствующих методов и средств вычислительной техники.

Разрабатываемый автоматизированный фонетический курс русского языка с использованием ЭВМ предлагает наличие таких внешних устройств, которые смогут в значительной степени взять на себя тренировочно-контролирующие функции и, что особенно важно, функции слухового анализатора. В настоящее время слуховой анализ речи обучаемых проводится преподавателем, ведущим занятия, фонологичность речевого слуха ограничивает возможности самостоятельного слухового анализа речи обучаемыми, особенно на уровне звука и ритмики слова. По той же причине нельзя полностью положиться и на результаты слухового анализа зарубежного преподавателя -русиста. Даже носитель языка не всегда может хорошо справиться с этой задачей.

В связи с этим представляется целесообразным поручить ЭВМ следующее:

1. регистрацию и ведение каждого обучаемого;

2. входное тестирование на продвинутом этапе обучения: его анализ, выдачу результатов и их хранение до окончания обучения;

3. обучающе-контролирующая и контролирующе-обучающая тренировка по всем трем уровням: звуковому, акцентно-ритмическому, интонационному - с моментальным контролем правильности/неправильности выполнения;

4. текущий контроль: выдача сведений после проведенного занятия о количестве проработанного учебного материала, качестве и количестве попыток, степени его усвоения;

5. тренировочно-контролирующие упражнения в фонетической транскрипции (для филологов);

6. поэтапный контроль, позволяющий получить представление о прохождении курса;

7. итоговый контроль (при корректировочном курсе повторяющий параметры входного тестирования) и выдача сведений по каждому учащемуся о результатах прохождения всего курса.

Как видно из приведенного перечня, ЭВМ будут поручены все виды контроля, наиболее трудоемкая и утомительная часть работы преподавателя над произношением. Это позволит более продуктивно использовать учебное время и уделить большее внимание развитию различных видов речевой деятельности.

Для успешной реализации автоматизированного курса по разделу фонетики необходимо предусмотреть в АОС следующее.

1. Компьютер должен быть совмещен не только с дисплеем, но и с магнитофоном, видеомагнитофоном, а также с печатающим устройством.

2. Клавиатура компьютера должна иметь не только строчные, но и прописные буквы, "ъ", "ё", а также все виды скобок, надстрочных и подстрочных знаков, используемых в русской фонетической транскрипции.

3. Экран дисплея должен передавать изображение тональных контуров предложений звучащего текста, интенсивность и длительность гласных звуков, характеризующих ритмическую структуру слова, определять и высвечивать на экране качество контролируемого звука. Экран дисплея должен допускать возможность демонстрации артикуляционных и других схем,таблиц,видеофильмов.

4. Разработка "автоматического транскриптора" позволит обеспечить фонетическую транскрипцию необходимых отрезков звучащей речи и "будет способствовать порождению правдоподобной с точки зрения орфоэпии звуковой последовательности" (Л.В.Бондарко, 1986).

Значительное увеличение эффективности обучения произношению ожидается не только за счет активной самоконтролируемой деятельности обучаемого, но и в значительной степени благодаря подкреплению слухового восприятия зрительным.

Преимущество обучения с использованием АОС заключается в том, что объективно оцениваются результаты работы обучаемых, выдаются оптимальные рекомендации для дальнейшей самостоятельной работы.

Для реализации программы эффективной АОС назрела необходимость создания единого банка фонетических данных нормативного русского произношения.

THE ABSOLUTE SEMITONE SCALE

KULLO VENDE

Department of English
Tallinn Teacher Training Institute
Tallinn, Estonia, USSR 200102

ABSTRACT

The absolute semitone scale is a scale combining the properties of both physical and perceptual units. It is derived from a modified Fletcher's formula,

$$P\ (st) = 12 \log_2 F_o\ (Hz),$$

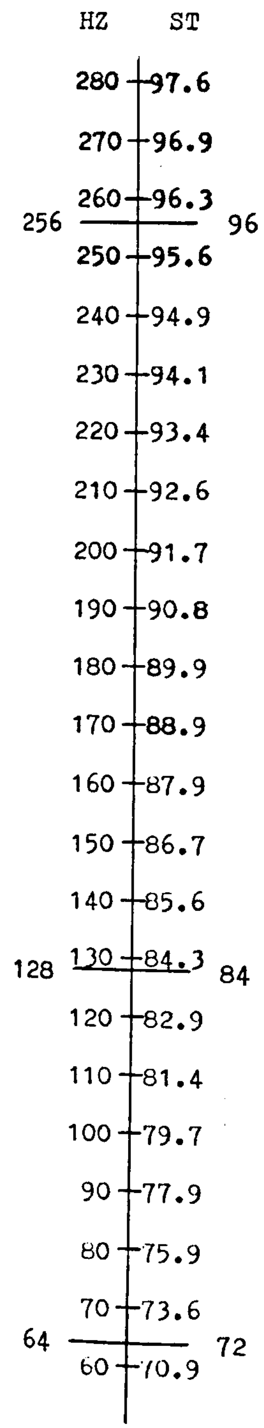
to relate fundamental frequency to its correlate perceptual units of pitch, viz., semitones above 1 Hz (1 Hz = 0 st). The absolute pitch units are much more convenient than Hz for the presentation, comparison (i.e., calculation of relative pitch differences) and other processing of raw data obtained in instrumental prosodic research.

In prosodic research the presentation of fundamental frequency in cycles per second is admissible only as far as raw measurement data are concerned. Any further manipulation and discussion or interpretation of the data should be carried out in units of perception. Even the graphs of F_o movement applying the linear frequency scale are perceptually misleading: they give a wrong idea of an extensive pitch movement which is never perceived by listeners as such. The logarithmic scale is a solution for graphs, although not very convenient for plotting unless one has special charts where every cps (Hz) can be plotted accurately.
The comparison and statistical processing of raw data in Hz in terms of perception leads to distortions even in case of one speaker, let alone speakers with different F_o ranges. The perceptually relevant comparison of two tones can be carried out by calculating their ratio, which further may be converted into semitones. Thus, given two measured frequencies, 150 Hz and 100 Hz, it is useless to state, in a discussion of their perception, that their difference is 50 Hz: the perception of the 50 Hz difference here is quite unlike the perception of a 50 Hz difference, say, between 250 Hz and 200 Hz. Instead, one can state that the r a t i o of the first pair of frequencies is 3:2 whereas that of the second pair is 5:4. For an untrained imagination, however, it is still clearer to state that the (musical) i n t e r v a l between the first two frequencies is 7 semitones, or a fifth, and that between the other two frequencies is 4 semitones, or a third.
But calculating average F_o values in Hz is of very doubtful value, as is drawing conclusions from differences between such values or such averages. To say that one F_o contour individually or on an average differs from another by a 10 Hz difference between their peaks is quite meaningless.

As long as we believe that the perception of the fundamental frequency of speech is logarithmic in the same way as it is for pure tones, the only possibility to process F_o data mathematically is in linear units on the logarithmic scale to which the data should be converted. The basic unit of pitch is the octave. The convenient unit for the analysis of fundamental pitch is the semitone. Proceeding from FLETCHER (1929) who introduced a scale of octaves and centioctaves above 1 kHz for the whole of the audible pitch range, it is possible to modify Fletcher's formula for calculating the pitch of the voice fundamental in a b s o l u t e s e m i t o n e s a b o v e 1 Hz:

$$P\ (st) = 12 \log_2 F_o\ (Hz \approx cps).$$

According to this formula,
1 Hz = 0 st, 2 Hz = 12 st, 4 Hz = 24 st, 64 Hz = 72 st, 512 Hz = 108 st (Fig. 1).
That is, instead of operating with figures in the F_o range of (roughly) 64..512 Hz, we can operate in the pitch range of 72.. ..108 st. The figures of the latter scale are suited for any kind of mathematical processing without notably violating the perceptual reality. Considering these two pairs of numerical data (depicted in Fig. 2), we can easily find the average pitch of the latter pair to be 90 st; the difference (interval) between the lower pitch and the average as well as between the higher pitch and the average is 18 st (1.5 octaves). The result is perceptually informative, unlike the average of the two former figures, 288 Hz, where the lower interval would be about 26 st as against 10 st of the upper interval.
Data in Hz can easily be converted into absolute semitones by means of a table where every Hz is given its correlate



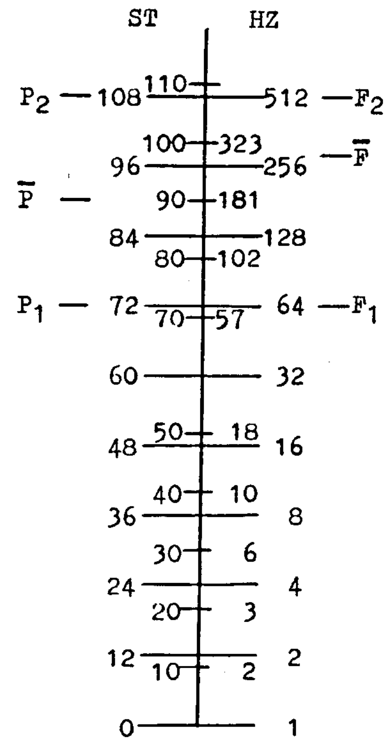


Fig. 1. (Left.) Linear frequency scale in Hertz (left) and the correlate pitch values in semitones (right).

Fig. 2. (Right.) Linear pitch scale in st (left) and the correlate frequency values in Hz (right). Plotting of two fundamental frequencies, F_1 = 64 Hz and F_2 = 512 Hz, on the linear pitch scale and the correlate logarithmic frequency scale. $\bar{F}$ is the mean frequency, $\bar{P}$ is the mean pitch of the two signals.

value in st with the accuracy of .1 st (higher precision is unnecessary in phonetics). This table is printed on the 4th page of the present paper. In computer-assisted F_o extraction the conversion can be done automatically, applying the above formula. For a programming language applying natural logarithms (such as BASIC used for computing the given conversion table) the formula will be

$$P = 12 \times 1.442695 \times \ln F_o.$$

It would be highly advisable to present even raw data in these absolute perceptual units. The investigator himself could immediately estimate the perceivable differences between the measured parameters of pitch and carry out all kinds of mathematical operations with the data without the ad hoc calculation of ratios or finding of logarithms. Intervals could be calculated by simple subtraction. Pitch contours and other graphs can easily be drawn on ordinary square paper.
The reader, too, could at once see what the measured pitches, intervals and ranges mean in terms of perception. Also, a reader of publications applying the absolute semitone scale could easily compare the data of different authors without the need to convert the Hz into ratios and then back into the traditional but unnecessary Hz if he wants to publish his results of comparison. One could combine and process the results and initial data for further generalizations, compute averages of pitch contours of different authors (including one's own), etc.
For example, the paper of LIN et al. (1984) includes two tables representing the average pitch in 2- and 3-syllable tone groups of Chinese. Pitch is expressed in Hz. Let us consider a line of their Table 1 - tone 4 + tone 3:

	1st syll.	2nd syll.
male speaker	196-110	104-82-114
female speaker	242-152	143-82-156

The figures are given as averages. Although it is wrong in phonetics to average hertzes (what can be averaged is their logarithms), let us regard these figures as representing single speech acts. All we can see is that both speakers pronounce the first syllable with falling pitch and the second with a fall-rise which is steeper for the female speaker. Now let us convert the hertzes into semitones:

	1st syll.	2nd syll.
male speaker	91.4-81.4	80.4-76.3-82.0
female speaker	95.0-87.0	85.9-76.3-87.4

Here the extent of pitch movement is at once obvious: the male speaker appears to make a 10-st fall in the 1st syllable against the female speaker's 8 st; the 2nd syllable starts 1 st below the end of the 1st; the subsequent falls in the 2nd syllable are 4 st and nearly 10 st, respectively, and the final rises about 6 and 11 st, respectively. Further comparison with the other tone groups in the table may show to what extent these findings are relevant.

Another aspect. In order to average the two pitch contours of the above tone group, with their parameters expressed in Hz, we would have to draw both of them on logarithmic paper and calculate the average contour geometrically (Fig. 3). Yet it is much simpler to average the parameters expressed in st arithmetically, (m+f)/2 93.2-84.2 83.2-76.3-84.7, and draw the resulting contour of the same shape on square paper.

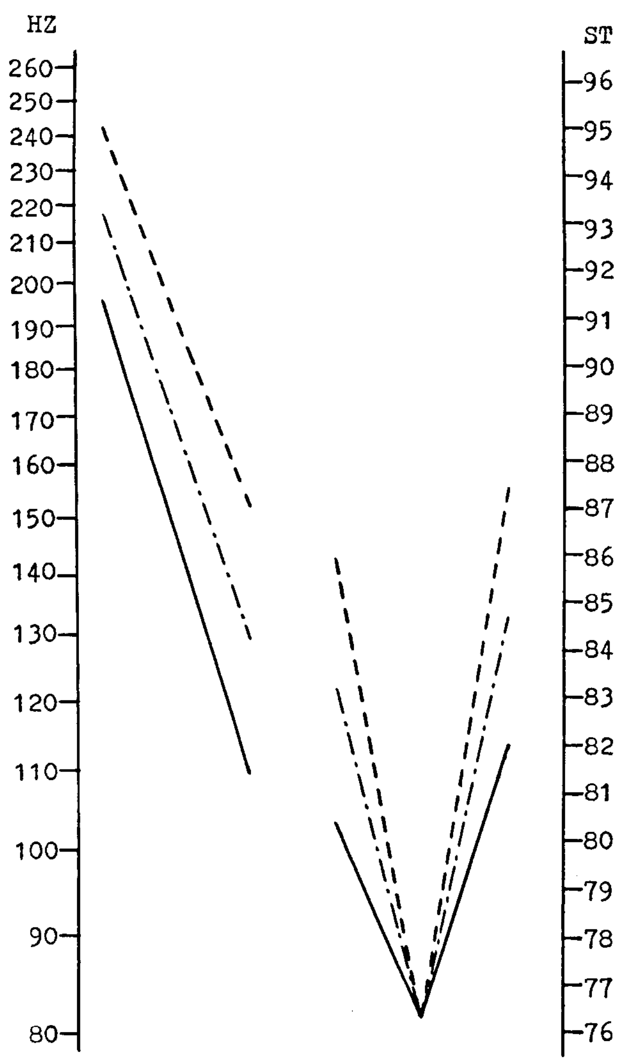


Fig. 3. Plotting and averaging of two contours of a Chinese tone group on the logarithmic frequency scale and absolute semitone scale.

——— male speaker
------ female speaker
-.-.-. the average contour

When synthetic speech is used in prosodic research, it is expedient, with a view to their subsequent mathematical/statistical analysis, to make up the tonal contours for the synthetic stimuli in absolute pitch units, varying the pitch of certain contour points by steps of m st instead of n Hz. It will considerably facilitate, for instance, correlation analysis between the input pitch data and the listeners' responses when the former are expressed in semitones on the absolute linear pitch scale; and it is equally easy to interpret the results of such analysis.

The absolute semitone scale was first introduced in Tallinn in 1972 (VENDE 1972) and has since been successfully applied here (e.g., PIIR 1985).

The other existing pitch scales, such as the mel scale or the Bark scale, are efficient for plotting psychoacoustic data for frequencies above 500 Hz, i.e. for spectrum analysis, but apparently not sensitive enough and too clumsy to handle (otherwise why should prosodists have avoided them?) in the range of fundamental frequency. It remains to hope that the absolute semitone scale, which is likewise both physical and perceptual, will gradually break through the hitherto dominating hertz tradition, take root and spread in prosodic research.

REFERENCES

Fletcher, H. 1929. Speech and Hearing. London.

Lin, M.C., Yan, T.Zh., Sun, G.H. 1984. The stress pattern and its acoustic correlates in Beijing Mandarin. - In: The Proceedings of the Tenth International Congress of Phonetic Sciences. - Dordrecht: Foris Publications, p. 504-514.

Piir, H. 1985. Acoustics of the Estonian diphthongs. - Estonian Papers in Phonetics: EPP 1982-1983. - Tallinn, p. 5-96.

Vende, K. 1972. Intrinsic pitch of Estonian vowels: measurement and perception. - Estonian Papers in Phonetics. - Tallinn, p. 44-108.

TABLE

Conversion of Hertzes into Semitones

Hz	st	Hz	st	Hz	st	Hz	st
1	0.0	60	70.9	119	82.7	178	89.7
2	12.0	61	71.2	120	82.9	179	89.8
3	19.0	62	71.5	121	83.0	180	89.9
4	24.0	63	71.7	122	83.2	181	90.0
5	27.9	64	72.0	123	83.3	182	90.1
6	31.0	65	72.3	124	83.5	183	90.2
7	33.7	66	72.5	125	83.6	184	90.3
8	36.0	67	72.8	126	83.7	185	90.4
9	38.0	68	73.0	127	83.9	186	90.5
10	39.9	69	73.3	128	84.0	187	90.6
11	41.5	70	73.6	129	84.1	188	90.7
12	43.0	71	73.8	130	84.3	189	90.7
13	44.4	72	74.0	131	84.4	190	90.8
14	45.7	73	74.3	132	84.5	191	90.9
15	46.9	74	74.5	133	84.7	192	91.0
16	48.0	75	74.7	134	84.8	193	91.1
17	49.0	76	75.0	135	84.9	194	91.2
18	50.0	77	75.2	136	85.0	195	91.3
19	51.0	78	75.4	137	85.2	196	91.4
20	51.9	79	75.6	138	85.3	197	91.5
21	52.7	80	75.9	139	85.4	198	91.6
22	53.5	81	76.1	140	85.6	199	91.6
23	54.3	82	76.3	141	85.7	200	91.7
24	55.0	83	76.5	142	85.8	201	91.8
25	55.7	84	76.7	143	85.9	202	91.9
26	56.4	85	76.9	144	86.0	203	92.0
27	57.1	86	77.1	145	86.2	204	92.1
28	57.7	87	77.3	146	86.3	205	92.2
29	58.3	88	77.5	147	86.4	206	92.2
30	58.9	89	77.7	148	86.5	207	92.3
31	59.5	90	77.9	149	86.6	208	92.4
32	60.0	91	78.1	150	86.7	209	92.5
33	60.5	92	78.3	151	86.9	210	92.6
34	61.0	93	78.5	152	87.0	211	92.7
35	61.6	94	78.7	153	87.1	212	92.7
36	62.0	95	78.8	154	87.2	213	92.8
37	62.5	96	79.0	155	87.3	214	92.9
38	63.0	97	79.2	156	87.4	215	93.0
39	63.4	98	79.4	157	87.5	216	93.1
40	63.9	99	79.6	158	87.6	217	93.1
41	64.3	100	79.7	159	87.8	218	93.2
42	64.7	101	79.9	160	87.9	219	93.3
43	65.1	102	80.1	161	88.0	220	93.4
44	65.5	103	80.2	162	88.1	221	93.5
45	65.9	104	80.4	163	88.2	222	93.5
46	66.3	105	80.6	164	88.3	223	93.6
47	66.7	106	80.7	165	88.4	224	93.7
48	67.0	107	80.9	166	88.5	225	93.8
49	67.4	108	81.1	167	88.6	226	93.8
50	67.7	109	81.2	168	88.7	227	93.9
51	68.1	110	81.4	169	88.8	228	94.0
52	68.4	111	81.5	170	88.9	229	94.1
53	68.7	112	81.7	171	89.0	230	94.1
54	69.1	113	81.8	172	89.1	231	94.2
55	69.4	114	82.0	173	89.2	232	94.3
56	69.7	115	82.1	174	89.3	233	94.4
57	70.0	116	82.3	175	89.4	234	94.4
58	70.3	117	82.4	176	89.5	235	94.5
59	70.6	118	82.6	177	89.6	236	94.6

237	94.7	306	99.1	375	102.6	444	105.5
238	94.7	307	99.1	376	102.7	445	105.6
239	94.8	308	99.2	377	102.7	446	105.6
240	94.9	309	99.3	378	102.7	447	105.6
241	95.0	310	99.3	379	102.8	448	105.7
242	95.0	311	99.4	380	102.8	449	105.7
243	95.1	312	99.4	381	102.9	450	105.8
244	95.2	313	99.5	382	102.9	451	105.8
245	95.2	314	99.5	383	103.0	452	105.8
246	95.3	315	99.6	384	103.0	453	105.9
247	95.4	316	99.6	385	103.1	454	105.9
248	95.5	317	99.7	386	103.1	455	106.0
249	95.5	318	99.8	387	103.2	456	106.0
250	95.6	319	99.8	388	103.2	457	106.0
251	95.7	320	99.9	389	103.2	458	106.1
252	95.7	321	99.9	390	103.3	459	106.1
253	95.8	322	100.0	391	103.3	460	106.1
254	95.9	323	100.0	392	103.4	461	106.2
255	95.9	324	100.1	393	103.4	462	106.2
256	96.0	325	100.1	394	103.5	463	106.3
257	96.1	326	100.2	395	103.5	464	106.3
258	96.1	327	100.2	396	103.6	465	106.3
259	96.2	328	100.3	397	103.6	466	106.4
260	96.3	329	100.3	398	103.6	467	106.4
261	96.3	330	100.4	399	103.7	468	106.4
262	96.4	331	100.4	400	103.7	469	106.5
263	96.5	332	100.5	401	103.8	470	106.5
264	96.5	333	100.6	402	103.8	471	106.6
265	96.6	334	100.6	403	103.9	472	106.6
266	96.7	335	100.7	404	103.9	473	106.6
267	96.7	336	100.7	405	103.9	474	106.7
268	96.8	337	100.8	406	104.0	475	106.7
269	96.9	338	100.8	407	104.0	476	106.7
270	96.9	339	100.9	408	104.1	477	106.8
271	97.0	340	100.9	409	104.1	478	106.8
272	97.0	341	101.0	410	104.2	479	106.8
273	97.1	342	101.0	411	104.2	480	106.9
274	97.2	343	101.1	412	104.2	481	106.9
275	97.2	344	101.1	413	104.3	482	107.0
276	97.3	345	101.2	414	104.3	483	107.0
277	97.4	346	101.2	415	104.4	484	107.0
278	97.4	347	101.3	416	104.4	485	107.1
279	97.5	348	101.3	417	104.4	486	107.1
280	97.6	349	101.4	418	104.5	487	107.1
281	97.6	350	101.4	419	104.5	488	107.2
282	97.7	351	101.5	420	104.6	489	107.2
283	97.7	352	101.5	421	104.6	490	107.2
284	97.8	353	101.6	422	104.7	491	107.3
285	97.9	354	101.6	423	104.7	492	107.3
286	97.9	355	101.7	424	104.7	493	107.3
287	98.0	356	101.7	425	104.8	494	107.4
288	98.0	357	101.8	426	104.8	495	107.4
289	98.1	358	101.8	427	104.9	496	107.5
290	98.2	359	101.9	428	104.9	497	107.5
291	98.2	360	101.9	429	104.9	498	107.5
292	98.3	361	102.0	430	105.0	499	107.6
293	98.3	362	102.0	431	105.0	500	107.6
294	98.4	363	102.0	432	105.1	501	107.6
295	98.5	364	102.1	433	105.1	502	107.7
296	98.5	365	102.1	434	105.1	503	107.7
297	98.6	366	102.2	435	105.2	504	107.7
298	98.6	367	102.2	436	105.2	505	107.8
299	98.7	368	102.3	437	105.3	506	107.8
300	98.7	369	102.3	438	105.3	507	107.8
301	98.8	370	102.4	439	105.3	508	107.9
302	98.9	371	102.4	440	105.4	509	107.9
303	98.9	372	102.5	441	105.4	510	107.9
304	99.0	373	102.5	442	105.5	511	108.0
305	99.0	374	102.6	443	105.5	512	108.0

APPENDIX

МОРФОЛОГИЧЕСКАЯ ГРАНИЦА КАК СРЕДСТВО СЛОГОВОЙ СЕГМЕНТАЦИИ В ЗАПАДНО-ИРАНСКИХ ЯЗЫКАХ

Талбак Хаскашев

Таджикский госуниверситет
г. Душанбе, СССР

Существование такого феномена, как слог, в настоящее время не вызывает сомнения. Оно доказывается многочисленными лингвистическими и нелингвистическими фактами:

1) носители различных языков членят речевой поток на слоги и пользуются ими в своей речи при определенных условиях;
2) лингвисты в различных языках выявляют и описывают структурные элементы слога, говорят о доминирующем и сопутствующих средствах слогообразования и о различных функциях слога;
3) представители различных неязыковедческих наук (литературоведы, музыковеды, медики, инженеры связи и др.) пользуются понятием слога.

Однако проблема слога как лингвистического явления, по справедливому признанию подавляющего большинства исследователей, в современном языковедении считается еще "неясной и спорной в самих своих основаниях" (10, 251). Это прежде всего касается вопроса о том, как и на основе каких критериев выделяется слог из потока речи как объект наблюдения и исследования - то, что Е.Курилович удачно назвал основной проблемой теории слога (14, 267).

Причин для споров вокруг проблемы слоговой сегментации и определении ее критериев много, на что указывают все исследователи, имевшие когда-либо дело с проблемой членения потока речи на слоги. Нам бы хотелось остановиться на одной из них, которая, на наш взгляд, имеет существенное значение для слоговой сегментации потока речи.

Главная причина разногласий заключается в том, что результаты научного слогоделения (т.е. сегментации потока речи на слоги на основе той или иной теории) в различных языках по-разному соотносятся с результатами слогоделения, осуществляемыми носителями языка; будучи с успехом применимы в одних языках, эти результаты не находят успешного применения в других. Причина несовпадения результатов научного слогоделения с результатами слоговой сегментации, осуществляемой носителями языка, кроется, **как** нам кажется, в отсутствии четкой границы между типологическими (общими) и специфическими (национальными) аспектами проблемы слога.

Слог, как и другие языковые явления (фонема, ударения, интонация и т.д.), с одной стороны, имеет такие черты, которые присущи всем или большой группе языков одинакового (фонемного или слогового) строя, с другой стороны, характеризуется такими особенностями, которые присущи только одному конкретному языку или группе близкородственных языков.

К первому аспекту проблемы, т.е. к типологическим особенностям относятся вопросы: а) о фонетической природе слога, т.е. о средствах (доминирующем и сопутствующих) слогообразования без учета их количества, сочетаемости и способов их распределения в пределах слога и связанные с ними явлений; б) о функциях слога.

К специфическому аспекту проблемы, очевидно можно отнести также два круга вопросов: о количестве структурных элементов слога, сочетаемости и связанных с ними явлений; о слоговой сегментации, ее средствах и критериях.

О специфическом характере последних двух групп вопросов говорит прежде всего само определение слога как наименьшей произносительной единицы (4; 10; 12; 22; 23), получившее широкое распространение в современном языкознании. Если слог является единицей произношения, то слогоделение и происходящие в слоге явления у носителей различных языков не могут быть одинаковыми.

Известно, что "звуки человеческой речи - продукт действия произносительных органов и формы резонансных полостей произносительного аппарата человека" (11, 11). Произносительные органы, как физиологические категории, одинаковы у всех рас и народов, однако навыки их работы при речепроизводстве (произношении) не могут быть одинаковыми у всех народов - носителей языков, ибо разные языки имеют свойственные им фонологические системы, которые реализуются по своим закономерностям, что и определяет в конечном счете способы работы произносительных органов и навыки произношения. Навыки произношения, будучи исторической категорией, развиваясь и совершенствуясь вместе с развитием и совершенствованием фонетической системы языка, приобретают национальную форму, общую для всех носителей национального языка, нарушение которой расценивается как отклонение от нормы или как акцент. Новейшие исследования показали, что навыки национальной артикуляционной базы (произношения) обнаруживаются уже в младенческом возрасте. Об этом говорит тот факт, что звуки (плач) новорожденных младенцев различных национальностей не одинаковы даже в возрасте одного дня (7, 58-64).

Таким образом, различение двух аспектов - типологического и специфического при рассмотрении слога как лингвистического явления необходимо, ибо исследования этих аспектов ведут к нетождественным результатам: результаты исследования специфических вопросов, характерных для того или иного языка, могут не совпадать с результатами рассмотрения аналогичных вопросов в другом (или других) языке. Поэтому вопрос о слоговой сегментации, будучи связанным с реализацией конкретной фонологической системы, закономерностями функционирования ее элементов должен рассматриваться для каждого конкретного языка отдельно.

Очевидно, исследование специфических особенностей слога должна предшествовать рассмотрению его типологических свойств, так как о слоге, об образующих его элементах и функциях мы можем судить лишь после того, как слог вычленен из потока речи и определено количество его структурных элементов. "Выделение слога внутри слова, - писал Е.Курилович, - это в сущности предварительная и необходимая операция, предшествующая любому описанию структуры или свойств слога" (14, 267).

Вопрос о слогоделении, как известно, во весь рост встает тогда, когда перед нами оказывается консонантная группа, состоящая из двух и более согласных. "В слогоделении решающую роль играют, - писал Е.Курилович, - промежуточные (интервокальные - Т.Х.) группы согласных и задача слогоделения состоит в основном в разделении групп согласных на имплозивную часть (= принадлежащую к предшествующему слогу) и эксплозивную часть (= принадлежащую последующему слогу)" (14, 268).

Известно, что любая научная теория (или теоретическая модель) имеет большую ценность тогда, когда она соотносима с фактами действительности и как-то объясняет их.

Исходя из этого положения и имея в виду существующие основные теории слоговой сегментации, для выяснения вопроса о слоговой сегментации в современном таджикском языке нами был специально составлен текст, состоящий из 300 словесных знаков, включающих в себя все возможные в таджикском языке двухфонемные и трехфонемные сочетания согласных, описанные в литературе (1; 3). Этот текст был сегментирован специалистом-фонетистом и носителями языка: специалист сегментировал текст на слоги дважды: один раз согласно сонорной теории в том ее варианте, который был предложен Р.И. Аванесовым, и второй раз - по теории мускульного напряжения Л.В. Щербы. (Мы назовем этот вид слогоделения теоретическим слогоделением).

Две группы испытуемых по 10 человек в каждой (студенты второго курса гуманитарного и естественного профилей, для которых таджикский язык является родным) этот текст сегментировали на слоги устно перед микрофоном. Их сегментация была записана на магнитной ленте, а затем перенесена на бумагу. (Эту сегментацию назовем интуитивной, а полученные после нее слоги - интуитивными).

Сравнение результатов теоретического и интуитивного слогоделения показало:

а) полное совпадение результатов обоих случаев интуитивного слогоделения без расхождений;

б) значительное совпадение результатов обоих случаев теоретического слогоделения при определенных расхождениях;

в) значительное несовпадение результатов теоретической и интуитивной сегментации при небольшом их сходстве.

Так, согласно сонорной теории, таджикские слова типа *мактаб* 'школа', *чашма* 'родник', *бодринг* 'огурец' сегментируются на слоги: *ма-ктаб, ча-шма, бо-дринг*, а слова типа *марказ* 'центр', *манзил* 'жилье', *пумба* 'хлопок' - на слоги: *мар-каз, ман-зил, пум-ба*, так как в словах первой группы первые слоги, завершаясь на гласный, и слова второй группы - на сонорный, соблюдают принцип возрастания звучности (2, 42). Однако только сегментация второй группы слов совпадает с делением их на слоги, которое наблюдается у носителей языка. Сегментация первого типа слов, очень характерного для таджикского языка, не совпадает с их членением носителями языка, которые делят эти слова на слоги следующим образом: *мак-таб, чаш-ма, бод-ринг*.

Сравнение результатов теоретических и интуитивных слогоделений выявило еще большее расхождение между теорией мускульного напряжения и интуитивного слогоделения. Так, теория мускульного напряжения, исходя из места ударения по отношению к консонантной группе, следующие таджикские слова делит на слоги: *ҳу-шнуд* 'довольный', *у-стод* 'учитель', *со-хтмон* 'стройка', *ҳа-вли* 'дом', *на-здик* 'близкий', т.к. конечное ударение в этих словах, притягивая консонантную группу к себе, делает их сильноконечными (22, §87 - 90). Однако такое слогоделение не совпадает слогоделением носителей таджикского языка: *ус-тод, сохт-мон, ҳав-лӣ, наз-дик*.

Таким образом, сонорная теория и теория мускульного напряжения, успешно работающие в ряде языков, в том числе и русском, не находят своего успешного применения в таджикском языке.

Исходя из известного положения о невозможности членения речевого потока на лингвистические единицы исключительно на основе акустических и артикуляторных признаков его составляющих (9, 10, 36 - 42) и указания на то, что "распознавание лингвистических единиц основано на знании контекстуальных, грамматических и семантических закономерностей данного языка" (20, 211), можно предположить, что закономерности слогоделения должны объяснить те теории слоговой

сегментации, которые основаны на лингвистических, семантических критериях. И действительно, результаты интуитивного слогоделения лучше других объясняет теория слоговой сегментации Е.Куриловича, основанная на лингвистических критериях. Е. Курилович, как известно, опираясь на выдвинутые им же принципы изоморфизма языковых единиц, пытался сформулировать ряд правил, исходя из сходства начальных и конечных консонантных элементов слога и слова. Он писал: "При первой попытке появляется искушение, чтобы применить здесь (при членении консонантной группы - Т.Х.) критерии начала и конца слова. Действительно начальные группы слова, являясь одновременно началом слога, всегда имплозивна, точно также конечная группа, являясь концом последнего слога слова, всегда эксплозивна. Таким образом, начало и конец слова могли дать нам надежные критерии, позволяющие разграничить внутри группы имплозивную и эксплозивную части. Например, др.-инд.: *mantra* 'гимн' = *man-tra* так, как в языке, с одной стороны, существуют слова, оканчивающиеся на *n*, например, *brahan* 'несущий', с другой стороны, *tr* возможно в начале слова (*trayah* 'третий'). Расчленение на *ma-ntra* или *mant-ra* было бы невозможным, так как слов с конечным *ntr* или с конечным *nt* не существует. Исходя из этого, разделение на слоги *man-tra* является единственно возможным решением" (14, 268).

Если обратиться к таджикскому языку, то действительно мы в нем обнаружим такое же сходство начальных и конечных консонантных элементов слога и слова: слоги и слова начинаются одним согласным и имеют в исходе один или два согласных. Учитывая это совпадение, слова *чармгар* 'сапожник', *расмкаш* 'художник', *гармрӯ* 'приятный' можно членить на слоги: *чарм-гар, расм-каш, гарм-рӯ*, так как другое их членение (*чар-мгар* или *ча-рмгар, рас-мкаш* или *ра-смкаш, гар-мрӯ* или *га-рмрӯ*) противоречило бы изоморфизму начальных элементов слога и слова. Такое членение вполне соответствует результатам интуитивного членения на слоги. Однако, как показывает материал, действенность этого критерия ограничена. Это правило хорошо работает во всех случаях, когда мы имеем дело с группой, состоящей из трех согласных.

Трудности возникают тогда, когда перед нами консонантная группа, состоящая из двух согласных, например, в словах *мактаб, маҳфил* 'общество', *чорбоғ* 'парк', *чорсӯ* 'платок', *мардум* 'люди', *наздик* 'близкий', *осмон* 'небо' и др. Дело в том, что слова этого типа могут быть расчленены на слоги трояко. Например, слово *чорбоғ* может быть сегментировано как *чо-рбоғ, чор-боғ* и *чорб-оғ*, из которых, согласно теории Е. Куриловича, только первое членение (*чо-рбоғ*) не приемлимо, так как при нем второй слог начинается двумя согласными, что противоречит изоморфизму начала слога и слова. Два других членения, согласно этой теории, допустимы, так как при них начальные и конечные элементы соответствуют начальным и конечным элементам слова. Однако с точки зрения носителей языка, как показывают наши данные, неприемлемым является и членение *чорб-оғ*, несмотря на полное совпадение начальных и конечных элементов этих слогов с соответствующими элементами слова: в таджикском языке есть слова, имеющие в начале один согласный и заканчивающиеся на два согласных, например, *дард* 'боль', и слова, начинающиеся гласным - *орд* 'мука'. С точки зрения носителей таджикского языка единственно правильным членением является *чор-боғ*, т.е. такое членение, при котором слоговое и морфемное членения совпадают.

Таким образом, критерии начала и конца слова, хорошо объясняющие сегментацию консонантных групп из трех согласных, оказываются непригодными при объяснении расчленения консонантной группы, состоящей из двух согласных, а это ставит исследователя перед необходимостью искать более надежные критерии слоговой сегментации.

Ю.С. Маслов, обсуждая вопрос о членении речевого потока на отдельные звуки (фонемы), и говоря о недостаточности акустико-артикуляторных признаков звуков для его осуществления, писал: "Можно думать, что для таких языков, как русский, истоки выделимости фонемы нужно искать в факте принадлежности в ряде случаев соседящих и нередко тесно "сплетающихся" в потоке речи фонов к разным значащим единицам - к разным морфемам и даже словам. Особенно наглядны в этом отношении примеры однофонемных морфем и слов: в них данная фонема оказывается сразу с двух сторон выделенной в речевом потоке или в составе той или иной формы определенными "смысловыми" границами" (17, 52).

Эта методика членения потока речи, т.е. членение через границу значимых единиц, применяемая для выделения отдельного звука речи (фонемы) в традициях Щербовской фонологической школы (6; 9; 10; 17; 22), нам представляется приемлемой для слоговой сегментации в таджикском языке с той разницей, что при слоговой сегментации мы исходим не из однофонемных морфем и слов как при сегментации на звуки речи (фонемы), а из фонетических характеристик односложных морфем и слов. Преимуществом этого критерия по сравнению с критерием Е.Куриловича является то, что он учитывает не только начальные и конечные элементы слога (слова), но и весь его звуковой состав, что очень важно, как увидим ниже для принятия решения об определении места слогораздела.

Основы этой методики были заложены еще в начале нашего столетия Л.В.Щербой (22, 4 - 7) и получили дальнейшее развитие в трудах его учеников и последователей (4; 6; 9; 10; 17).

Суть этой методики заключается в том, что при членении потока речи на звуки речи через однофонемную морфему и слово берется в основу двоякое сходство фонемы и морфемы:

а) по функционированию, т.е. возможностью функционирования звука речи в качестве экспонента морфемы;

б) материально, т.е. совпадение акустико-артикуляторных характеристик звука речи и морфемы. Можно думать, что не будь этих сходств звука речи и морфемы, членение на отдельные звуки речи было бы невозможным.

Возвращаясь к слогоделению в таджикском языке и возможности выделения слога через морфологическую границу (через морфемы), мы исходим из следующих известных положений.

Во-первых, из того, что "фонетические комплексы (например, слоги) и семантические комплексы (например, предложения) независиво от объединяющих их функциональных отношений обнаруживают глубокий параллелизм структуры. Можно установить между ними удивительное сходство формы (изоморфизм)" (15; 14) и того, что "явления изоморфизма или однотипности структуры конститутивных языковых единиц и отношений между ними образуют особенности системы языка и теснейшим образом связаны с его различными уровнями..." (15, 56).

Во-вторых, из утверждения о том, что в "таксономической шкале языковых единиц слог занимает промежуточное положение между фонемой и словом, параллельно морфеме. Отношение между фонемой, слогом и словом в языках неслогового (фонемного) строя аналогичны отношениям фонемой, морфемой и словом: фонема входит составной частью в слог и морфему, слог и морфема формируют слова..." (18, 122 - 123).

В-третьих, из положения о том, что "для носителей языка членение потока речи на фонемы или слоги не является исходной, отправной точкой. Ведь в процессе речи носитель языка оперирует значащими единицами, по меньшей мере морфемами..." (17, 51).

В четвертых, из изоморфизма слога и морфемы как по составу структурных элементов и их характеру, так и по глубине в некотором отношении сходным с изоморфизмом слога и слова (19, 41).

Эти общие сходные черты слога и морфемы, будучи основой для осуществления слоговой сегментации через односложные морфемы и слова, сами по себе недостаточны, а необходимо, как отмечалось выше, материальное совпадение и функционирование слога в качестве экспонента морфемы.

Обращаясь с этой точки зрения к материалам таджикского языка, мы обнаруживаем удивительное совпадение интуитивных слогов, выделенных испытуемыми, и односложных морфем и слов как материально, так и по функционированию.

Во-первых, обнаруживается интересное материальное сходство слова и односложной морфемы:

а) интуитивные слоги и односложные морфемы (и слова) могут состоять от одного до четырех звуков: *ӯ* 'он', *бо* (предлог), *бом* 'крыша', *барг* 'листья', один из которых обязательно является гласным;

б) если интуитивные слоги и односложные морфемы (и слова) состоят из одного звука, то этот звук обязательно является гласным;

в) интуитивные слоги и односложные морфемы (и слова) начинаются только одним согласным (примеры выше);

г) интуитивные слоги и односложные морфемы в исходе имеют один или два согласных звука (примеры выше).

Во-вторых, слоги и односложные морфемы (и слова) в своем функционировании могут совпадать друг с другом, т.е. слоги могут выступить в качестве экспонентов морфем и слов: *ӯ рафт* 'он ушел', *соябон* 'зонтик', *боғ-бон* 'садовник' и др. (21, 17-22).

Именно благодаря этим последним сходствам слога и односложной морфемы (слова) в таджикском языке и оказывается возможным членение потока речи на слоги через морфему и односложные слова.

Таким образом, в таджикском языке при членении на слоги мы будем исходить из звукового состава односложной морфемы (и слова) как более в семантическом отношении самостоятельного и непосредственно данного явления и соответственно ему будем принимать решение о месте слогораздела в том или ином слове. Допустим, что перед нами таджикское слово *марзбон* 'пограничник', подлежащее сегментации на слоги. Оно может быть сегментировано так: *ма-рзбон*, *мар-збон*, *марз-бон*, *марзб-он*. Исходя из звукового состава морфемы, мы должны первую, вторую и четвертую сегментацию считать неприемлемыми потому, что:

а) в первом членении (*ма-рзбон*) второй слог начинается тремя согласными и в своем составе имеет пять звуков, - в таджикском языке нет односложных морфем (и слов), которые имеют в своем составе пять звуков и начинаются тремя согласными;

б) второе членение (*мар-збон*) неприемлемо, так как при нем второй слог начинается двумя согласными, ибо нет в таджикском языке односложных морфем и слов, начинающихся с двумя согласными;

в) четвертое членение (*марзб-он*) также не подходит потому, что при нем первый слог из пяти звуков и в исходе имеет три согласных, а таких односложных морфем в таджикском языке нет.

Единственно правильным и приемлемым является третье членение (*марз-бон*), т.е. такое членение, при котором слоговая и морфемная границы совпадают. Результаты этой сегментации полностью совпадают с результатами интуитивного слогоделения.

Таким образом, мы можем безошибочно сегментировать на слоги все исконно таджикские слова и заимствованные из других (арабского, тюркского, индийского и др.) языков слова, прошедшие в таджикском языке стадию освоения, которые имеют в интервокальном положении консонантные группы. Эта методика действенна во всех случаях, кроме границ синтагм, где членение на слоги подчиняется более сильному, синтагматическому членению.

Результаты этой слоговой сегментации хорошо согласуются с результатами слогоделения, осуществляемого носителями языка. Это показало сравнение слоговой сегментации, осуществленной по предлагаемой методике, с результатами слогоделения, осуществленного нашими информантами.

Предлагаемая методика слоговой сегментации может быть применена в близкородственных таджикскому языку персидском, дари, так как они характеризуются таким же слоговым строем (26), как и таджикский.

Предлагая осуществить в таджикском языке слоговую сегментацию на основе односложных морфем (и слов), считаем необходимым отметить, что мы далеки от мысли, что в таджикском языке и других западно-иранских языках обнаруживается абсолютное совпадение слоговых и морфемных границ, как в слоговых (вьетнамском, китайском и др.) языках, и что слоговые границы определяются исключительно морфемными. В таджикском языке, как и в персидском и дари, действительно обнаруживается совпадение слоговых и морфемных границ.

Но это не самое главное для предлагаемой методики (и поэтому название настоящего доклада носит несколько условный характер). Для нас важно материальное совпадение слога и морфемы. Благодаря этому совпадению носители таджикского языка сегментируют на слоги не только такие слова, в которых слоговые и морфемные границы совпадают (например, *сардор* 'руководитель'), но и такие, в которых эти границы не совпадают (*мак-таб* 'школа'). Очевидно, если в первом случае слоговая сегментация осуществляется по морфеме непосредственно, то во втором носители таджикского языка используют аналогию, осуществляя слогоделение под сильным влиянием морфемы и фоно-морфологического строя слова.

ЛИТЕРАТУРА

1. Авазбаев Н. Структура слога в языках различных типов. - Ташкент: Фан, 1986.
2. Аванесов Р.И. Фонетика современного русского литературного языка. - М: МГУ, 1956.
3. Бобомуродов Ш. Фонетико-фонологическая структура односложных слов современного литературного таджикского языка. Автореф. дис. канд. филол. наук. Тбилиси, 1975.
4. Бондарко Л.В., Павлова Л.П. О фонетических критериях при определении слоговой границы. // Рус. яз. в школе, 1967, № 4, с. 10-19.
5. Бондарко Л.В. Слоговая структура речи и дифференциальные признаки фонем. Автореф. дис. док. филол. наук. Л., 1969.
6. Гордина М.В. О различных функциональных единицах. // Исследования по фонологии. М., 1966, с. 172-183.
7. Джапаридзе З.Н., Стрельников Ю.А. О различиях в плаче новорожденных разной национальности и пола. / К проблеме уровней фонетической организации членораздельной речи. // Эксперимертально-фонетический анализ речи. - Л.: ЛГУ, 1984. Вып. I. С. 58-64.
8. Зиндер Л.Р. Материальная сторона языка и фонема. // Ленинизм и теоретические проблемы языкознания. - М., 1970. С. 372 - 381.
9. Зиндер Л.Р. Проблема сегментации речевого потока на кратчайшие звуковые единицы. // Вопросы фонетики и фонологии. Доклады советских ученых на VII международном конгрессе фонетических наук. - М., 1971. Ч. I, с. 138-152.
10. Зиндер Л.Р. Общая фонетика. 2-е изд. - М., 1979.
11. Зиндер Л.Р. К вопросу об артикуляционной базе. // Экспериментально-фонетический анализ речи. - Л., 1984. Вып. I. С. 8-12.
12. Касевич В.Б. Слог в общем и дальневосточном языкознании.// Теория языка, методы его исследования и преподавания. - Л., 1981. С. 141-146.
13. Кукольщикова Л.Е. Об одном спорном случае слогоделения в английском языке. // Экспериментально-фонетический анализ речи. - Л., 1984. Вып. I. С. 29-40.
14. Курилович Е. Вопросы теории слога. // Очерки по лингвистике. - М., 1962. С. 267-306.
15. Курилович Е. Понятие изоморфизма. // Очерки по лингвистике. - М., 1962. С. 307-319.
16. Макаев Э.А. К вопросу об изоморфизме.// Вопр. языкознания, 1961, № 5, с. 50-56.
17. Маслов Ю.С. Введение в языкознание. - М., 1975.
18. Метлюк А.А. Слог как языковая единица. // Романское и германское языкознание. Минск, 1984. Вып. 14. С. 121-125.
19. Моисеев А.И. Изоморфизм фонетического строения слога и слова.// Вестн. высш. шк. Филол. науки. - М., 1985, № 1, с. 41-44.
20. Фланган Дж. Анализ, синтез и восприятие речи. - М., 1968.
21. Хаскашев Т.Н. Фонетикаи забони адабии ҳозираи тоҷик (Фонетика современного таджикского литературного языка). Просодика. - Душанбе, 1984. Қ. 2.
22. Щерба Л.В. Русские гласные в качественном и количественном отношении. - Л., 1983.
23. Щерба Л.В. Фонетика французского языка. 4-е изд. - М., 1963.
24. Kramsky J.A. Phonological Analysis of Persian Monosyllables. // Archiv orientalni, 1947, v. 16, p. 103-134.
25. Palgram E. Syllable, word, nexus, cursus. - The Hague-Paris, 1970.

DELIMITATION AND FOCUSSING FUNCTIONS OF INTONATION IN BULGARIAN

ANASTASIA MISHEVA

Phonetic Laboratory, Bulgarian Language Institute
Bulgarian Academy of Sciences, Sofia, Bulgaria

ABSTRACT

The paper deals with the results from an experimental study of the phonetic segmentation of Bulgarian speech.
On the basis of experimental data an attempt is made to reduce the variety of the observed melodic configurations of segmentation units (syntagmas) describing them in terms of two kinds of accents, conditioned by the phonetic manifestation of the delimitation and focussing functions of intonation, and their possible combinations. The proposed classification may be used for comparative intonological analysis of languages with different phonetic structures.

INTRODUCTION

Theoretical ground of the present experimental study is the generally acknowledged significance of intonation within the framework of language communications as a phonetic source of information transferred from speaker to hearer. The different approaches to the versatility of intonational phenomena and their interaction with extralinguistic and linguistic factors from different levels result in a great variety of lists describing the functions of intonation. Regardless of their different number, definitions and terminology, almost all of the proposed classifications include the delimitation and focussing functions under one or other lable [1,2,3,4 5,6 and many others]. Here the term "delimitation function" is used for describing the structural role of intonation - the segmentation of the speech flow into phonetic units, the so called syntagmas, and the organizing role of intonation in their formation. The syntagma is defined in the sense of the Leningrad Phonetic School as a semantic-syntactic entity consisting of one or more words organized by intonation into phonetic integrity. The main phonetic condition for a string of words to be grouped into a syntagma is the absence of perceived pauses between them, the pauses being of two kinds: real (a sound gap) or psychological (due to interruptions of the continuity of the prosodic parameters of speech signal) [4].
The focussing function of intonation is understood as the role of intonation in localizing the communicatively important elements in the utterance.
The aim of the study is to determine the phonetic means used in Bulgarian for the realization of the delimitation and focussing functions of intonation and on this ground to specify the basic phonetic regularities in the suprasegmental organization of speech allowing adequate comparison with other languages.

FOCUSSING FUNCTION

The manifestation of the focussing function of intonation in Bulgarian follows the universal tendency to phonetic prominence by complex enhancing of the acoustic prosodic parameters (F_0,I,T) of the stressed vowel in the phonetic word containing important information.
A result of a previous experimental study [7] of the F_0-patterns of short simple sentences uttered as one-syntagmatic statements is that their most typical melodic contour is of rise-fall configuration with its peak on the stressed syllable of the word bearing the phrase accent. The application of the "theme - rheme" conception to our experimental data determines the melodic contrast between the syllables in the word as a main acoustic feature of the rheme and the lack of it, i.e. the unidirectional course of F_0-contour as an acoustic characteristic of the theme the direction depending on its position in the phrase.

DELIMITATION FUNCTION

The phonetic breaking up of the speech flow into shorter or longer stretches by means of pauses shows a great variety in the size and intonational shape of the divided units due to individual variations in speech. The intervention of the subjective factor however is not a reason to consider this process an accidental one without linguistic significance since individual variations are possible only within certain limits conditioned by the particularities of intonation as a structural device in a given language and the universal restrictions in the speech production process. The phonetic manifestation of the delimitation function of intonation may be estimated by means of the statistical parameters of syntagmas representing the basic tendencies in the temporal organization of speech.
The statistical data given below concern the segmentation of speech in the particular case of text reading as the first stage of a larger expermental investigation of the suprasegmental organization of Bulgarian speech. The data are obtained by means of an auditory and acoustic analysis of three texts in journalistic style (40 simple and 21 compound and complex sentences, the total number of clauses being 91) read by 12 persons.

Sintagma size

Fig.1 shows the occurrence frequencies of the syntagmas of different size the measure of the size being the number of phonetic words in a syntagma. The minimal size is 1 word or 1 syllable, the maximum size found in our corpus is 6 phonetic words or 28 syllables, the average size being 2.35 phonetic words or 8 syllables. The comparison of these values with the corresponding data for other languages (Russian [8],[9]; English [10]; Spanish [11]) shows a great similarity which makes it possible to assume that the main factors determining the size of these units are the universal restrictions in the speech production.

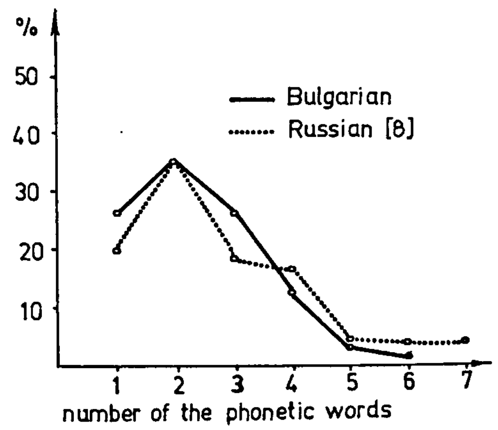
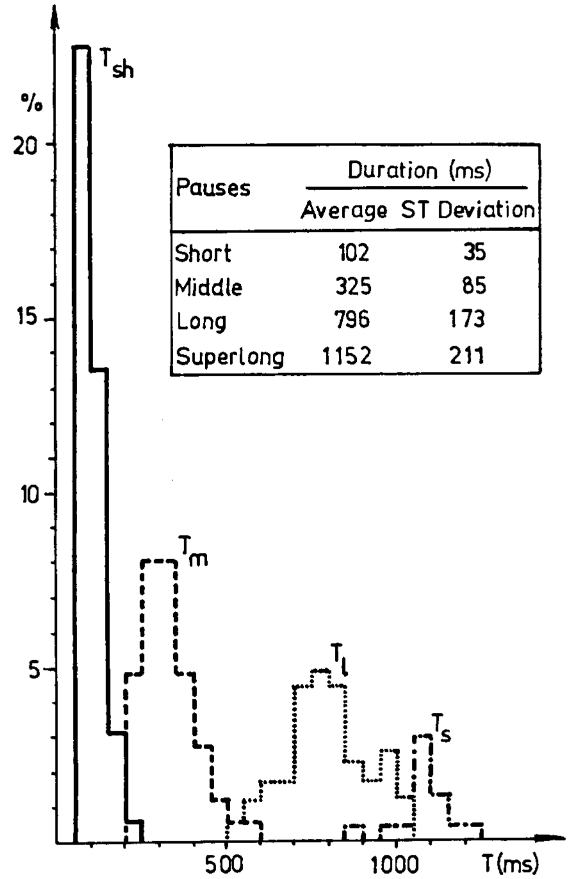


Fig.1. Occurrence frequencies of syntagmas (%) according to their size.

The statistic treatment of the syntagmas with respect to their size and position in the sentence show that the short syntagmas (1-2 phonetic words) prefer the initial position, the longest ones (5-6 phonetic words) usually occupy the middle of the sentence and for syntagmas of the middle size (3 phonetic words) the final position is typical.

Pauses

Under the experimental conditions, i.e. reading of grammatically well formed units, it has been found that real pauses are much more frequently used (87%) than psychological ones (13%).



Pauses	Duration (ms)	
	Average	ST Deviation
Short	102	35
Middle	325	85
Long	796	173
Superlong	1152	211

Fig.2. Occurrence frequencies of pauses according to their duration.

Fig.2 represents the occurrence frequencies of real pauses according to their duration. The configuration of the curve (4 peaks) and the dependence of the pause durations on the syntactic relations between the adjacent syntagmas give ground to distinguish 4 kinds of pauses: short, middle, long and superlong. The short pauses usually separate groups of words in the simple sentence, the middle ones - clauses in the compound and complex sentences, the long ones occur between the separate sentences and the superlong ones mark the end of the phono-paragraph. It must be pointed out that although closely related the phonetic and syntactic units do not necessarily coincide. For example the syntactic entity of the noun group may be phonetically divided by a short pause between the attribute agreed with the noun and the noun itself only because of the larger number of syllables in the two words.

Number of accents in a syntagma
The number of accents in the syntagmas in our material is defined by auditory tests the instruction being to underline phonetically prominent words without any restriction in their number. The results show that in Bulgarian speech a syntagma may have more than one accents with an equally perceived weight. The maximum number of accents found in our recordings is 3 (in 3% of all units). Most of the syntagmas have one accent (72%) and 25% of all cases - 2 accents. The possibility of a syntagma to have more than one accent has been noted in other languages as well (e.g. Russian [5],[12]).

Types of accents
The auditory and acoustic analysis of our recordings show that the accentuation in a syntagma may be realized in two different ways perceived as phonetic prominences of equal strength but differing in their quality. This suggests that in Bulgarian there are also two kinds of accents as described in [12] which differ in their functions,i.e. in the nature of the information intended by the speaker. The first one - the logic accent - is a result from the phonetic manifestation of the focussing function of intonation and its acoustic correlate is the melodic contrast between the syllables of the prominent word (rise-fall F_o-configuration). The second type, called structural, is connected with the whole intended structure of the utterance and signals the continuation or the finality of speech,i.e. it is a result of the delimitation function of intonation. The basic acoustic parameter used in this case is the segmental duration, the course of F_o is unidirectional - rising with maximum steepnes on the last syllable irrespective of the word stress for a continuation accent and lightly falling without melodic contrasts for a final accent.
It is well known fact that almost any word in the syntagma can bear a logic accent and its placement depends on the context, both linguistic and situational. It must be pointed out that in Bulgarian the structural accent also depends on situation (the term including the relationship between speaker and hearer, speaker and social framework and the momentary mood of the speaker [2])

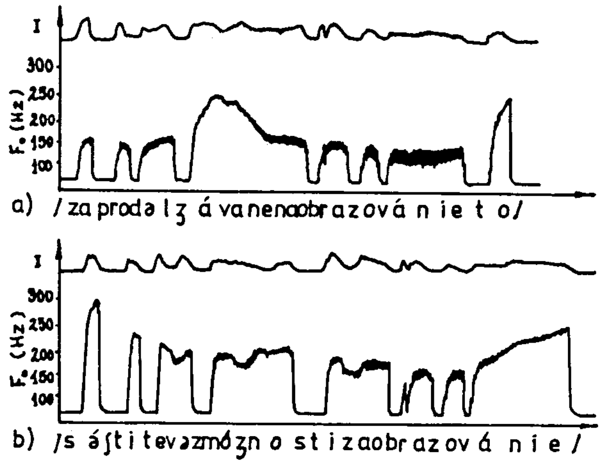


Fig.3. F_o-contours of syntagmas showing the acoustic difference between the logic and continuation accents and their combination on the last word: a) /za prodəlʒávane na obrazovánieto/ 'for the continuation of education', b) /səʃtite vəzmóʒnosti za obrazovánie/ 'the same possibilities for education'.

although its place is fixed on the last word(s), which means that the end of the syntagma is not necessarily marked by an accent.
When the logic accent is on the last word bearing at the same time a continuational accent the most typical F_o-contour is a rising convex curve with maximum steepness on the stressed vowel in contrast to concave configuration corresponding to a continuation accent alone. Figure 3 illustrares acoustic differences in F_o-configurations of syntagmas whose last words are marked a) by a continuation accent alone and b) by a combination of logic and continuation accents. The first word in both syntagmas are emphasized by a logic accent forming a melodic peak on their stressed syllables. The accents in the syntagma may also be differentiated in respect of the number of words they underline. In our experimental material we have found both types of accents defined in [10] - centralized (on a single word) and decentralized (on two words emphasized as one accent unit). On acoustic level the difference between the two cases is in the shape of the melodic peak - a sharp one on the stressed syllable of the individually accentuated word and a high relatively flat plateau situated between the stressed syllables of the two words (fig.4).

Place of accents
In one third of all examined syntagmas each phonetic word is accentuated by one

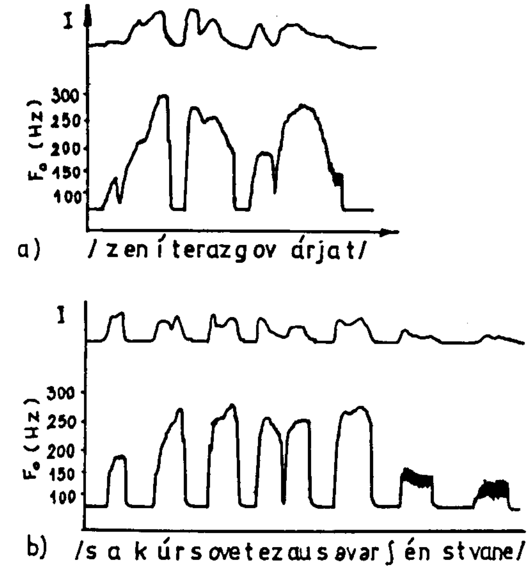


Fig.4. F_o-contours of two-words syntagmas in which both words bear a logic accent: a) of centralized type /zeníte razgovárjat/ 'the women are talking'; b) of decentralized type /sa kúrsovete za usəvərʃénstvane/ 'are the qualification courses'.

or another type of accent. In the rest of the units the number of accents is smaller than that of phonetic words.
An estimation of the global tendencies in accent organization of syntagmas may be derived from the statistical distribution of the observed accent places which is as follows: on the initial phonetic word - 36%, on the final one - 38%, on the initial and final - 12%, on the medial word - 14%. The small number of syntagmas in which the medial word bears the accent shows that in Bulgarian syntagmatic accents prefer boundary positions. This means that on a syntagmatic level the accent function is rather a delimitative than a culminative one as it has been shown in [13] dealing with the word stress in Bulgarian and Russian. An essential particularity of the syntagmatic accent organization in Bulgarian distinguishing it e.g. from Russian is that the initial position attracts the accent in the same degree as the final one, i.e. the classical rule about the syntagmatic accent being always on the last word isn't valid in Bulgarian.

CONCLUSION
The experimental data obtained in this stage of our investigation show that the registrated variety of F_o-configuration of syntagmas may be described by the number and nature of their accents which are directly connected with the formal configurational features such as the number of peaks and the character of the terminal tone. Thus, all different F_o-patterns found in our corpus (14 types) may by represented by the number (from 0 to 3) and kind (centralized or decentralized) of logic accents, the presence or absence of a structural accent (continuational or final) and their possible combination. The statistical aspect of this representation makes it possible to evaluate in a more general way the activity of the delimiation and focussing functions of intonation in Bulgarian and to ascribe a higher weight to the latter one as the main factor determining the sound form of Bulgarian speech. In this way the experimental F_o-data may by used as a source of infomation for the purposes of an intonation typology of languages revealing in a more explicit manner the functional particularities of intonation.

References

[1] N.S.Trubetzkoy,‹Grundzüge der Phonologie›, Vandenoeck & Ruprech in Göttingen, 1958.
[2] W.J.Barry, Prosodic Functions Revisited Again! Phonetica 38: 320-340, 1981.
[3] Т.М.Николаева, Фразовая интонация славянских языков, Москва, 1977.
[4] Л.Р.Зиндер, Общая фонетика, Москва 1979.
[5] Н.Д.Светозарова, Интонационная система русского языка, Ленинград, 1982.
[6] Д.Тилков, Интонацията в българския език, София, 1981.
[7] A.Misheva, Changes of the acoustical features of a phonetic word conditioned by its place in the utterance structure, Abs. of 8-th Colloquium on Acoustics, Budapest, C27, 1982.
[8] Л.В.Златоустова, Сопоставительное изучение интонации русского и болгарского языков, Сборник доклади на КФФ при МКС, София, 1984.
[9] Б.В.Братусь, Л.А.Вербицкая, Пособие по фонетике для иностранных студентов-филологов, Москва, 1983.
[10] А.М.Антипова, Ритмическая система английской речи, Москва, 1984.
[11] Т.Д.Змеева, Основные интонационные контури испанского языка. Сборник Лаборатории устной речи Филологического факультета, Москва, 1986, 60-79.
[12] Л.В.Златоустова, Фонетические единицы русской речи, Москва, 1981
[13] Л.В.Златоустова, А.Мишева, Акцентно-ритмические единицы в русском и болгарском языках, Актуальные вопросы интонации: 47-54, Москва, 1984.

ВЕРОЯТНОСТНАЯ МОДЕЛЬ СТИХА

М.Л. Гаспаров

Институт мировой литературы
АН СССР, Москва, I2I069

Описывается метод сопоставления стихотворного ритма и естественного языкового ритма в рамках заданной метрической схемы,- с помощью модели, вычисляемой на основании частоты слов разного ритмического строения и списка словосочетаний, укладывающихся в данную метрическую схему. Этот метод, первоначально разработанный на материале русского стиха, здесь применяется к материалу английскому, французскому, итальянскому, испанскому и средневековому латинскому. Он позволяет отделить в стихе чисто-языковые явления от специфически-художественных и этим помочь согласованию лингвистического и литературоведческого подхода к стиху. Особенно он плодотворен, как кажется, при определении форм стиха, колеблющихся между разными системами стихосложения (например, между силлабикой и силлабо-тоникой).

I. Когда литературовед исследует звуковое строение стиха, то перед ним встает вопрос первостепенной важности: как выделить в стихе те особенности, которые возникают стихийно, в силу естественного ритма языка, и какие на их фоне порождаются творческой волей поэта. Только последние представляют собой предмет исследования для литературоведа. Чем лучше мы сможем отделить в стихе общеязыковые явления от явлений специфически-стиховых, тем легче будет работать над стихом и лингвисту и литературоведу.

Решение этого вопроса впервые было найдено Б.В.Томашевским /I/ в I9I7 г. и уточнено А.Н.Колмогоровым в начале I960-х гг. Был найден способ построить вероятностную модель стиха - определить тот ритм расположения ударений и словоразделов в стихотворной строке, какой явился бы, если бы стих строился только по естественным данным языка и по формально заданной схеме данного размера. Эта вероятностная модель сравнивается с реальным ритмом такого-то стихотворения, поэта или эпохи. Там, где их показатели совпадают,- например, частота ударности на таком-то слоге или частота словораздела после такого-то слога,- там мы можем сказать, что для поэта данный элемент ритма был безразличен, и он пассивно следовал здесь естественному ритму языка. Если же оказывается, что на таком-то месте поэт ставит ударение значительно чаще или значительно реже языковой вероятности,- то мы должны признать, что здесь вмешивается творческая тенденция: поэт предпочитает такие-то ритмические формы и избегает таких-то ритмических форм. Выявить эти предпочтения и избегания, свести их в систему и по возможности объяснить,- в этом и состоит задача стиховеда.

И для Томашевского, и для Колмогорова, и для его учеников материалом для обследования служил только русский классический стих. Мы попытались обследовать этим же методом некоторые размеры иноязычного стиха - английского, французского, итальянского, испанского, средневекового латинского. Результаты этого обследования излагаются ниже.

2. Построение модели. Вероятностная модель стиха строится следующим образом.

а) Подсчитывается частота различных типов фонетических слов в языке (по прозаическим текстам): какую часть словаря составляют I-сложные слова, какую - 2-сложные с ударением на I слоге, какую - 2-сложные с ударением на 2 слоге и т.д. Здесь самое важное - найти правильную трактовку двойственных (преимущественно служебных) слов, которые могут звучать и как ударные и как безударные (т.е. проклитики и энклитики) /2/. Очень важно, чтобы акцентуационные принципы разных исследователей были согласованы: только тогда их результаты будут сопоставимы.

б) Составляется список словосочетаний, укладывающихся в метрическую схему исследуемого размера ("ритмических вариантов", различающихся положением ударений, и "словораздельных вариантов", различающихся положением словоразделов при той же расстановке ударений). Здесь самое важное -полностью учесть все практически употребительные комбинации и сознательно оставить в стороне ритмические раритеты, которые только загромоздили бы картину. Иногда одни и те же ритмы для одной эпохи приходится считать дозволенными, а для другой запретными (такова, например, разница между раскатанным и строгим ритмом английского ямба чередующихся эпох, исследованная М.Тарлинской /3/).

в) В предположении, что в естественном языке слова ритмически независимы друг от друга, мы считаем, что вероятность встретить в тексте такое-то сочетание слов равна произведению вероятностей этих слов, взятых отдельно. Мы перемножаем вероятности ритмических типов слов, входящих в каждый словораздельный вариант и получаем вероятность встречаемости этого словораздельного варианта в целом.

г-д) Полученные вероятности всех вариаций данного размера складываются. От этой суммы вычисляется, какой процент составляет каждое слагаемое. Предположим, это будет I,7%. С этим показателем мы и сопоставляем частоту данной вариации у исследуемых поэтов. Если расхождение невелико, то можно считать, что поэт при пользовании данной формой руководствуется только естественными данными языка, т.е. эстетически она ему безразлична. Если расхождение будет значительно, то приходится признать, что поэт систематически (пусть бессознательно) предпочитает или избегает данную вариацию, т.е. ощущает ее как особенно благозвучную или особенно неблагозвучную; это и есть "специфически-художественная" тенденция стиха.

Оперировать при сравнении такими маленькими объектами, как отдельный словораздельный вариант, может быть рискованно; поэтому обычно сравниваются целые группы таких минимальных вариаций, объединенных а) по расположению схемных ударений (кривая ударности), б) по расположению сверхсхемных ударений, в) по расположению словоразделов.

3. Исходный ритмический словарь. Для расчета ритмического словаря английского языка было взято приблизительно по I000 фонетических слов из Дефо, Свифта, Стерна, В.Скотта, Диккенса, Теккерея и Г.Джеймса - всего 68I7 слов. Для французского языка - по 900-I300 слов из Расина (предисловия к трагедиям), Мольера, Вольтера и Бальзака, - всего 4278 слов. Для итальянского языка - по 200 слов из Боккаччо, Саккетти, Банделло, Гращини, Гоцци, Альфьери, Пеллико, Кардуччи, Верга, д'Аннунцио - всего 2000 слов. Для испанского языка - I000 слов из Сервантеса. Для латинского языка - по I000 слов из "Сатирикона" Петрония и "Исповеди" Августина, всего 2000 слов. Для русского языка XVII-XVIII вв. - прозаические тексты Симеона Полоцкого, Прокоповича, Кантемира, "Повесть о Савве Грудцыне" и трагедия "Иудифь" - всего 5000 слов. Цифры по отдельным авторам обнаруживают достаточную однородность, поэтому для предварительной разведки они могут считаться надежными; но, конечно, эти подсчеты необходимо продолжать далее.

4. Английский 4-ст.ямб и 4-ст.хорей. Для разбора ямба было взято I400 строк из поэмы Теннисона "Ин мемориам" и I000 строк из лирики и поэмы "Пасха" Браунинга; для разбора хорея - первая I000 строк из "Песни о Гайавате" Лонгфелло. Из сравнения с моделью видно: а) реальный стих полноударнее, чем модель: 4-ударные вариации употребляются чаще вероятности, 2-ударные - реже вероятности. (В русском стихе тенденция та же). б) Ритм схемных ударений в модели ямба - ровный, в модели хорея - альтернирующий (I и III стопы слабоударны, II и IV стопы сильноударны); в реальном хорее эта альтернация еще более усиливается, в реальном ямбе - по крайней мере, намечается. (В русском стихе альтернирующий ритм развился гораздо сильнее, но тоже раньше и больше в 4-ст.хорее, чем в 4-ст.ямбе /4/). в) Ритм сверхсхемных ударений в ямбе идет по убывающей от первого до последнего слабого места (отмечено еще Тарлинской и Бейли /3;5/); у Браунинга сверхсхемных ударений значительно больше, чем у Теннисона. г) В ритме словоразделов те формы, которые преобладали в модели, еще больше преобладают в реальном стихе. (В русском стихе тенденция - та же /6/).

4. Средневековый латинский 4+3-ст. хорей ("I5-сложник 8ж+7д). Материал - 27 латинских стихотворений VIII-IX вв. (I989 стихов), исследованных Б.И.Ярхо в неопубликованной работе I9I5-I920 гг. /7/. Определение размера спорно: то ли это "силлабо-тонический хорей с отступлениями к силлабике", то ли "силлабический стих с тенденцией к силлабо-тонике"? Мы рассчитали 3 вероятностные модели: А) чисто-силлабическую, по схеме "8ж+7д"; Б) силлабическую с дополнительными цезурами внутри каждого полустишия, по схеме (4+4ж)+(4+3д), что по условиям латинской акцентологии порождает некоторое усиление хореического ритма; В) силлабо-тоническую, по схеме 4+3-ст.хорея. Сравнение с реальным стихом показывает: реальный стих явно лежит ближе к силлабо-тонической модели, чем к обеим силлабическим моделям. "Показатель хореичности" (процент ударений, падающих на нечетные слоги, от общего числа ударений) для модели А равен (для I и II полустиший) 76% и 82%; для модели Б - 84% и 86%; для модели В - I00%; для реального же стиха - 95% и 91%. Это позволяет утверждать, что "открытие силлабо-тоники" в европейской поэзии состоялось не в конце XII в. у миннезингеров, а лет на 300 раньше у латинских книжников.

6. Итальянский II-сложник. Как известно, французские, итальянские, испанские писатели, читатели и исследователи считают свой стих силлабическим, а немецкие ученые и их последователи (даже Курылович, /8/) - силлабо-тоническим с перебоями ритма. Наиболее адэкватное описание II-сложника: это (А) стих из II слогов с обязательным ударением на I0 слоге, (Б) а также на 4 и/или 6 слоге, (В) и - по крайней мере, с XVI в. - с запретом самостоятель-

ного ударения ("стресс-максимума") на 7 слоге. Соответственно с этими нарастающими ограничениями были рассчитаны вероятностные модели А, Б и В, а также модель Г для чисто-силлаботонического ямба. Материал – по 500 строк из сицилийской школы ХШ в., тосканской школы ХШ в., Данте, Петрарки,Полициано,Ариосто,Тассо,Марино, Метастазио,Парини,Альфьери,Мандзони,Леопарди,Кардуччи,Пасколи,д'Аннунцио,Гоццано; "Показатель ямбичности" (процент ударений на четных слогах от общего числа ударений) для модели А – 68%, Б – 78%, В – 91%, Г – I00%. Показатели ямбичности всех рассматриваемых поэтов колеблются между Б и В, постепенно эволюционируя от Б к В, т.е. к усилению ямбичности. Эволюция эта проходит две волны: от Данте (Б) до Метастазио и Парини (В) и от Мандзони (Б) к д'Аннунцио и Гоццано (В); рубежом является Альфьери с его резко индивидуальным ритмом. Таким образом, итальянский стих, действительно, стоит на полпути от силлабики к силлабо-тонике /9/.

7. Испанский 8-сложник и французский 7-сложник (это – один и тот же размер, в 8 слогов при женском и в 7 при мужском окончании). Материал – подборки испанских народных романсов (по изд.Дюкамена), литературных романсов ХУП в. (Лопе де Вега, Гонгора, Кеведо) и ХУШ в. (Н.Моратин, Мелендес Вальдес, Ховельянос) и поэм Соррильи для XIX в.; всего 3480 строк. Для сравнения взята французская имитация – "Романсеро Сида" Гюго (из "Легенды веков) Любопытно,что вероятностные модели испанского и французского стиха, несмотря на разницу языков, оказываются близки друг к другу. Сопоставление показывает: показатели реального стиха совпадают с показателями моделей почти идеально; данный размер – действительно чисто-силлабический. (Традиционная передача его в немецких,русских и английских имитациях 4-ст. хореем – условность.) Таким образом, из двух ведущих размеров испанской поэзии, более "книжного" II-сложника (итальянского типа: с моделью сверены тексты Гарсиласо, Лопе де Вега, Моратина, подборки лириков ХУШ в. и Р.Дарио) и более "народного" 8-сложника первый тяготеет к силлабо-тонике, а второй чисто-силлабичен; в общей семантической структуре испанской поэзии это существенно.

8. Французский I2-сложник "6+6" (александрийский стих). Модель рассчитывалась для каждого 6-сложного полустишия с обязательным ударением на 6 слоге. Материал – по 600-I000 строк дю Белле, Расина, А. Шенье, Бодлера. Совпадение показателей модели с показателями реального стиха – почти полное; данный стих – силлабический, без каких бы то ни было силлабо-тонических тенденций. (Традиционная передача его немецким и русским 6-ст.ямбом – условность). Можно отметить,что (а) в первых полустишиях ударения расположены гуще и стыки их чаще, стих как бы становится легче к концу; (б) в первых полустишиях несколько чаще встречается ямбический ритм (ударения на 2 и 4 слогах), а во вторых – анапестический (ударение на 9 слоге) (в) никакой тенденции к сочетанию однородных полустиший (ямб+ямб, анапест+анапест) не обнаруживается; (г) все эти особенности наиболее заметны у Расина и Шенье,ритм же раннего стиха у дю Белле и позднего стиха у Бодлера ближе к модели – к естественному языковому ритму.

9. Русский силлабический I3-сложник (7+6ж) ХУП-ХУШ вв. Этот стих был заимствован из польской силлабики Симеоном Полоцким (I664) и преобразован в 4+3-ст.хорей В.Тредиаковским (I735). В I920-I950-х гг. велись споры, был ли этот переход от силлабики к силлабо-тонике эволюционно-плавным или революционно-резким. Для проверки была построена модель чисто-силлабического стиха "7+6ж", и с нею сравнивались памятники I670-I700 гг. (Симеон, Медведев, Дм.Ростовский), I700-I734 (Прокопович,Журавский,Хмарный,Буслаев,анонимная пьеса "Дафнис"), I729-I735 (первая редакция сатир Кантемира) и после I735 (Тредиаковский и вторая редакция сатир Кантемира). Сравнение показывает: на протяжении двух 30-летий все показатели (процент женской цезуры, сочетания слов в полустишии,полустиший в стихе,стихов в двустишии) близко держится естественной языковой вероятности, а затем в течение 5 лет происходит перелом: у Кантемира снижается женская цезура и усиливается хореический ритм I полустишия, а у Тредиаковского женская цезура исчезает совсем и хореический ритм становится строго выдержанным. Переход русского стиха от силлабики к силлабо-тонике был не эволюционным, а революционным /I0/.

I0. Таковы предварительные результаты применения вероятностной модели к исследованию европейского стиха. Можно видеть, что наиболее показательные результаты обнаруживаются там, где приходится решать самый общий (и самый трудный для аргументации) вопрос: силлабический стих перед нами или силлабо-тонический? Это важно: в истории едва ли не каждой поэзии (особенно в пору ее становления) многие памятники колеблются между разными системами стихосложения. Но и для более частных вопросов – о ритме ударений,словоразделов и т. п. – предлагаемый метод может быть так же полезен, как был он полезен на материале русского стиха. Он требует дальнейшего совершенствования в том,что касается учета проклитик и энклитик, в том,что касается построения моделей с различным охватом вариаций в зависимости от версификационного стиля разных эпох. Тогда доступный этому методу анализ сможет стать еще точнее. Это – один из тех участков, где лингвистика и литературоведение, смыкаясь, могут по праву притязать на статус точных наук.

Литература:

/I/ Б.В.Томашевский. О стихе. Ленинград, I929.

/2/ В.М.Жирмунский. Введение в метрику. Ленинград,I925 (особенно гл.3).

/3/ M. Tarlinskaja. English verse: theory and history. The Hague,I976.

/4/ К.Тарановски. Руски дводелни ритмови. Београд,I953.

/5/ J.Bailey.Towards a statistical analysis of English verse. Lisse,I975.

/6/ М.Л.Гаспаров. Современный русский стих: метрика и ритмика. Москва, I974.

/7/ М.Л.Гаспаров. Каролингские тетраметры и теоретическая модель латинского ритмического стиха. "Античная культура и современная наука".Москва,I985.

/8/ J.Kuryłowicz. Metrik und Sprachgeschichte. Wrocław (etc.),I975.

/9/ М.Л.Гаспаров. Итальянский стих: силлабика или силлабо-тоника? "Проблемы структурной лингвистики № I978",Москва,I98I.

/I0/ М.Гаспаров. Русский силлабический I3-сложник. "Metryka słowiańska", Wrocław (etc.), I97I.

THE INTERRELATIONSHIP BETWEEN PHONOLOGICAL AND PHONETIC SOUND CHANGES: A GREAT RHYTHM SHIFT OF OLD ESTONIAN

Arvo Eek, Toomas Help

Institute of Language and Literature,
Estonian Academy of Sciences,
Tallinn 200106, USSR

1. PURELY PHONETIC SOUND CHANGES

In the literature on historical linguistics, phonological and phonetic sound changes are often contrasted. A *phonological change* alters the phoneme composition of certain speech flow appearances, a *phonetic change* alters the way in which different expiratory-phonatory-articulatory gestures and their acoustic-auditory correlates are presented in certain speech flow appearances. For instance, in Old Litomyšl Czech, there was a sound change **p, > t* , manifested e.g. in **p,i:vo > ti:vo* 'beer'. It was a phonological change in the sense that the phoneme /p,/ was exchanged for the phoneme /t/ by replacing (in terms of Andersen 1973) the distinctive feature 'heightened low consonant tonality' of /p,/, by the feature 'high consonant tonality' of /t/. It was a phonetic change in the sense that the actual labial articulation of [p,] characterized (in terms of Ladefoged 1980) by certain values of the articulatory parameters like 'lip height', 'lip width', and 'lip protrusion' and its corresponding 'heightened low tonality' in acoustics-audition were exchanged for the actual dental articulation of [t] characterized by certain values of the articulatory parameters like '(tongue) tip raising' and '(tongue) tip advancing' and its corresponding 'high tonality' in acoustics-audition.

In historical linguistics all relevant sound changes are traditionally reduced to phonological changes. The prevalent strategy of historical linguists has long been to conceive the main course of a sound history as a *chain* of phonological changes and to view phonetic changes as mere detailed specifications of the stated phonological changes. Any single phonetic change, relevant to the main course of the sound history, has been viewed as obligatorily subordinated to a concrete phonological change. The traditional strict parallelism between phonological and phonetic changes is revealed in the theoretical framework of Andersen 1973, the original source of the above Old Litomyšl Czech example. Here the speciality of phonetic changes is emphasized to the extent that 'abductive' phonological and 'deductive' phonetic changes are dealt with as two different categories (about the equation of 'abduction' to phonology and 'deduction' to phonetics cf. p. 774). The framework articulates that a concrete historical sound change is always a complex of a phonological change and a phonetic change, i.e. phonological changes are understood as abductive innovations in the mental coding of sounds and phonetic changes as subsequent deductive innovations in their physico-physiological manifestation.

The traditional focus on phonological changes only rather than on phonetic changes as well, seems to be due to two methodological aspects of historical linguistics. On the one hand, most methods of historical sound reconstruction imply the view that sounds are *discrete* units: they rely on comparisons and differentiation between lexicon units (words, morphemes). On the other hand, only phonological changes are easy to be conceived as changes of discrete units (phonemes, distinctive features), whereas phonetic changes are better to be conceived as *continuous* changes of parameter values (determiners of the actual articulatory gestures and their acoustic-auditory correlates).

We argue that some phonetic changes, relevant to the main course of a sound history, have no concrete phonological changes as an one-to-one support. Let us call them *purely phonetic changes*. We conceive the main course of a sound history as a chain of changes some of which allow both phonological and phonetic specifications but some phonetic specifications only. Consequently, we do not reduce all relevant sound history to its phonological changes. The proposed equal focus on phonological and phonetic changes makes the recognition of historical sound changes more sophisticated. In order to reconstruct purely phonetic changes, the discrete philosophy of the method has to be adapted to the continuous nature of changes. However, these methodological complications do not hinder us in attributing the phonetic changes the essential ontological value. Instead, we view them as an inevitable forfeit to be paid for the proposed theoretical adjustment.

The essence of purely phonetic sound changes is exemplified by introducing a Great Rhythm Shift of Old Estonian. However, before going to the change itself, we outline a theoretical framework in which the

essentials of speech rhythm are treated in purely phonetic terms so that two optional rhythm strategies - striving vs. switching foot control - are extracted.

2. STRIVING VS. SWITCHING CONTROL OF FOOT

2.1. *Phonological vs. phonetic perspective on language*

According to Ladefoged 1984, the role of phonemes for individual speakers and listeners has been grossly exaggerated in the tradition of language research. Phonemes, distinctive features and other units of this size are empirically valid devices of speech description only if language is viewed as a *social* norm of a community. They are too abstract devices while describing language as a *psychological* act or state of an individual, e.g. while describing the actual realization or the mental coding of concrete speech samples. To draw on his original parallel, phonemes are like moral prescriptions or economical laws: they are manifested in the behaviour of a human group rather than a single member of this group.

In concrete linguistic analysis, the social rather than psychological nature of phonemes is revealed best by the characteristics of a boundary between two subsequent segmental phonemes. In the psychological perspective, the boundary is a fiction: if to observe one single speech flow sample alone, no invariant cues of a boundary between phoneme-size segments are available (cf. Hammarberg 1976). The boundaries become a reality in the social perspective. Here, the observed speech flow sample is confronted with a set of potential resembling samples and only its recurrent details are extracted as relevant features. Similarly, a suprasegmental phoneme is evident in a set of possible speech flow samples rather than in a single observed sample.

The relationship between phonology and phonetics leads to a crucial revaluation. Phonology and phonetics do not deal with different empirical data, they are not complementary portions of a unitary description system. In particular, the mental coding of speech is reflected not only by phonology and the articulatory-acoustic-auditory realization of speech is reflected not only by phonetics. Rather, phonology and phonetics deal with the same empirical data but represent two essentially incompatible strategies that are proper for different purposes.

Phonology describes both the mental coding and the articulatory-acoustic-auditory realization of speech flow in the *functional* perspective. Here, speech flow is viewed as a combination of details that are selected from the total set of observation results as these features that motivate the existence of the particular phenomenon 'speech flow'. Speech flow exists in order to convey linguistic meaning, i.e. to indicate which linguistic units (morphemes, lexical words etc.) are actually used one after another in social communication.[1] Consequently, the functional perspective entails the view that speech flow is a string of phonemes, i.e. abstract meaning-differentiating capacities that are evident when speech flow equivalents of different linguistic units contrast with each other.

Phonetics describes both the mental coding and the articulatory-acoustic-auditory realization of speech flow in the *formal* perspective. Here, speech flow is viewed in all its details available in the observation of a speech sample as a psycho-physiological and acoustic act. According to Ladefoged 1984, phonemes, distinctive features, and other units of this size do not belong to such empirical realities. Instead, the formal perspective entails the view that speech flow is a continuously produced energy wave in the sense that it does not split into segmented units of phoneme size. In this context we may follow Plomp 1984 in identifying the acoustic appearance of speech flow as an air flow supported by continuous activity of respiratory mechanism and radiated from a human being as a wide-band signal that is modulated continuously in time by manipulating vocal cords (fundamental frequency), by narrowing and widening the vocal tract (temporal intensity envelope), and by modifying the vocal tract cavities (frequency spectrum), and received by the peripheral hearing apparatus. We may change the perspective and claim that speech flow appears as a continuous *speech energy* that passes subsequently through three media. First, it is produced in the physiological medium as the mentioned four-fold fluctuations in speech organs (speaker's respiratory, phonatory, and articulatory activities); second, it is transmitted in the physical medium as a modulated wide-band signal (acoustics); and third, it is received in the physiological medium as a fluctuation in the peripheral hearing apparatus (listener's auditory activation).[2]

Speech rhythm is a phenomenon that is directly manifested in speech flow: speech flow is actually a continuous alternation of the minimal and maximal levels of speech energy. We have posed phonetics and phonology as two mutually exclusive research strategies. In this context, we try to fix the essentials of speech rhythm in purely phonetic, energetical terms without invoking on phonological consideration.

2.2. *Stress, foot, syllable and demisyllable: a model of speech rhythm*

Explaining historical sound changes and describing typological differences in contemporary languages we proceed on the following crude model of speech rhythm.

Stress is the total energy amount spent by the speaker's complex expiratory, phonatory and articulatory activities (gestures) while producing a stretch of speech flow (for a review of literature on stress production and acoustics: Eek 1982; MacNeilage 1972) rather than some special reinforcing energy added to a certain independently defined unit of speech flow. Stress itself is intrinsically segmented into units rather than appears something like an energetical increment of stressed syllables relative to unstressed syllables.

Foot is the minimal integral unit of stress. Foot organizes speech energy into a stress impulse; its *general* shape, i.e. a hypothetical temporal energy envelope, is physiologically determined by the tension/relaxation phases, inevitably needed in the activities of speech organs, and its *detailed* shape is specified by language-specific commands on speech organs (Fig. 1). Foot cannot be understood as a chest pulse, a unit of laryngeal fluctuation or a motor unit like the articulatory syllable of Чистович, Кожевников et al. 1965, it is temporally organized amount of speech energy produced by all the activities.

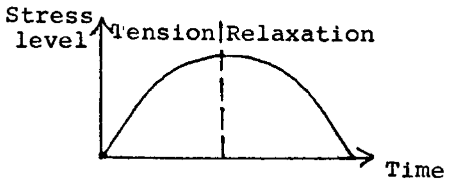


Figure 1. A model of foot.

In particular, it is foot that is best to be conceived as the domain of *accentuation*, i.e. the purposeful variation in the amount of stress. Psycholinguistic experiments (e.g. Terken 1984, cf. also Scott 1982) indicate that the amount of stress displayed in speech production is in a direct correlation with the importance of the signal in the interpretation of the corresponding message. A speech signal is maximally accentuated, if it alone serves as the basis for computing the meaning of the corresponding message; a speech signal is minimally accentuated, if the preceding linguistic context and the extralinguistic state of affairs rather than the speech signal itself serve as the basis for computing the meaning.[3] At the maximum end of the accentuation scale, there are *emphasized feet* that show an integral portion of stress of the largest size. In this case the needed amount of stress is warranted by special reinforcing activities including registerable pulmonic activities (cf. Ohala, Riordan, Kawasaki 1979). At the minimum end of the accentuation scale, there are *tonic feet*. In this case the needed amount of stress is warranted by keeping to the speech organs activities that are necessary and sufficient for the speech signal transmission. Between these two extremes there remain simply *accentuated feet*.

The intrinsic mode of stress appearance is a foot. In this physiologically predetermined limit the detailed shape of energy distribution is controlled by both universal and language-specific segmental and suprasegmental commands.

Speech is decomposed into continuous alternations of the narrowing (i.e. consonantal) and the widening (i.e. vocalic) gestures in the vocal tract. As the first approximation we may suppose that the detailed time-ordered segmental specifications for the basic CV- and -VC *demisyllable-gestures* are controlled by *demisyllabic commands* (cf. Fujimura 1983). Demisyllabic command scans all muscle channels exploited in generation of the respective minimal articulatory sequence and turns on simultaneously these channels which activity is not contradicted to the concrete state of the movement (cf. coarticulatory phenomena e.g. in the sequence of *plu-*: during the articulation of *p* lips are already rounded, immediately before *p* release the tongue tip is raised to the alveolar region and the postdorsum is moved towards the velum). Demisyllabic commands regulate articulatory movements in the speech flow. Demisyllable as a unitary articulatory sequence generates an indispensable (inherent) amount of speech energy that is sufficient for the complex gesture; its duration is determined by inherent durations of the combined movements. Consequently, demisyllables themselves do not specify the shape of a temporal energy envelope (i.e. the shape of a foot), they simply divide a foot into minimal inherent energy blocks.

Detailed modifications of a stress impulse may be controlled by two kinds of hypothetical energetic commands. These commands - switching and striving - determine the turning point between the tension and relaxation phases (cf. Fig. 1) of the temporal energy envelope.

If the *switching*-type foot control is used, there are usually two switching commands in a foot. The first of them turns on a stress impulse and predetermines the quickest way to the maximal energy level (thus controlling mainly the tract-widening segment, i.e. the vowel quality), but does not define the temporal characteristics of the maximal energy level itself; duration of the latter follows mainly from the inherent temporal properties of the corresponding demisyllable(s). Physiologically conditioned relaxation is turned on by the second switching command, thus showing the location of the turning point of a foot and simultaneously guaranteeing relatively exact sound quality. Thus, any switching command entails a rapid rise to an energy level. The switching-type foot control predicts a relatively tight coarticulation before the initial point of the maximal energy level of a foot and relatively loose connections between underlying segments after that point. As a rule, a foot tends to be a isobaric (in energetic terms) or isochronic unit (in terms of timing). In the case of switching the isobaric/isochronic tendency to an average strength or length value of all feet would be revealed relatively weakly.

If the *striving*-type foot control is used, there is one striving command in a foot. A striving command turns on a stress impulse smoothly and its main force is directed to the maximal energy level of a foot. The relaxation, i.e. the decline towards the minimal energy level, begins automatically after the final point of the tension phase is reached. In this case one would expect a relatively loose connections between underlying segments at the beginning of a foot (e.g. a diphthongization-like vocalic gesture may appear there) and close contact immediately before the exactly defined final point. A quality reduction of the underlying vocalic gesture is predictable in energetically uncontrolled relaxation phase. As the controllable stretch of a foot is essentially longer than in the switching case, a stronger isochronic tendency would appear.

If we want to delimit intra-foot segments that are influenced by energetic commands more properly we must define a syllable-size unit. Articulatorily, *syllable* is a relatively homogeneous sequence of demisyllables delimited by opening gestures of the vocal tract. This definition relies on two findings. On the one hand, the essential energetic commands - striving and switching - are inclined to apply to these demisyllabic commands at which a transition to the vocal tract opening gesture begins. On the other hand, these demisyllabic commands that are directed to a closing gesture are usually under the influence of preceding switching or striving commands. This regularity presents a crucial support to speech perception. Namely, in the corresponding acoustic sound wave used in the transmission of the message from speaker to listener, syllables are designated by the temporal amplitude envelope, i.e. the beginnings of syllables are definitely marked by transitions to the intensity rise of the sound wave. Thus, in acoustics a syllable is the foot stretch containing of one easily identifiable (cf. Mermelstein 1975) temporal amplitude envelope. The complex relationship between articulation and the produced sound wave acts as a temporally selective energy filter. For instance, the acoustic onset of a syllable does not immediately coincide with the relative time-point marked by the onset of the speech organs activity, i.e. by the onset of a foot (for the literature on perceptual centers see e.g. Fowler 1979, Marcus 1981, Howell 1984). However, on the basis of suggested close physiological interrelationship between speech production and perception (cf. the motor theory of speech perception of Liberman et al. 1962, the auditory-motor theory of speech production of Ladefoged et al. 1972; cf. also the observed parallels of syllable production and perception by Tuller and Fowler 1980) the information about an articulatory syllables may still be relevant in speech perception and vice versa.

Within the accentuation range that is determined by the functional characteristics of an utterance, the speech flow is produced by stress segments represented by foot-size alternations of the minimal-maximal-minimal physiological energy levels, i.e. by a sequence of temporal energy envelopes. In acoustics and perception, there are clearly identifiable intra-foot units characterized by separate temporal intensity envelopes we have defined as syllables. A *foot* may consist of *one*, *two* or *three syllables*. However, some accentuated or emphasized stress impulses may be conceived as displaying more than three syllables. In such cases we suppose that the accentuated or emphasized foot is followed by one or more intervening weak tonic feet (cf. Lea 1974: 41 for the phonetic justification of the 'one-two-three-syllables' principle).

2.3. *Types of rhythm organization*

Languages differ as regards the method by which energetic stress commands apply to their feet. Generally, the most "natural" foot types seem to be those in which every single syllable is affected by one command only. There are two "natural" types of languages that exploit this principle as a rule, i.e. display foot control systems subordinated to syllable-switching or syllable-striving.

1) *Syllable-switching* foot control is characteristic of languages with prototypical polysyllabic feet (e.g. in Finnish and Italian); in this type of languages monosyllabic feet are rarely used.

Syllable-switching languages give much attention to the beginnings of syllables. As a result of that the target of the following vowel is clearly defined, there are no perceivable diphthongization of short vowels; in the regulation of stress degrees also foot-initial consonant may take a part (cf. e.g. data for Tamil: Balasubramanian 1979). In foot-final syllables, syllable-switching languages have an open set of vowels to choose from (cf. e.g. the vowel harmony in Finnish) and do not, regularly, show reduced vowels, or more correctly, do show less reduced vowels than e.g. foot-striving languages (for Finnish: Wiik 1965, for Italian: Bertinetto 1981). Syllable-switching languages may have an opposition between short and long segmentals. This language type displays a relatively weak foot-level and syllable-level isochrony (for Italian: Vayra, Avesani, Fowler 1984, for Finnish: Lehtonen 1970 and for Tamil: Balasubramanian 1979). Here the inherent endeavour to foot isochrony may be strongly reduced, because the duration of a foot is determined roughly by their inherent temporal properties.

2) *Syllable-striving* foot control is characteristic of languages with prototypical monosyllabic feet (e.g. Vietnamese).

In syllable-striving languages, polysyllabic words cannot be analyzed into a continuous alternation of stressed and unstressed syllables (cf. Касевич 1983) due to foot monosyllabicity: here relaxation occupies a final part of a syllable rather than a separate syllable. Syllable-striving languages are usually tone languages, since the control over the vowel matter and the whole maximal energy segment develops conditions for tonogenesis.

The principle according to which every single syllable is affected by one energetic command may be violated in two additional, less "natural" types of languages. Here the "natural" foot types are accompanied by less "natural" types. In languages that display foot control systems subordinated to foot-striving, there are feet in which some syllables are left without any energetical control altogether. In languages that display demisyllable-switching, there are feet in which one syllable is affected by two separate energetic commands.

3) *Foot-striving* control is characteristic of languages with mono- and polysyllabic feet (e.g. English, German, Russian).

All that has been said about syllable-striving languages may also be valid for foot-striving languages and vice versa (the only exception being phenomena related to non-foot-initial syllables of foot-striving languages).

Striving command does not pay special attention to the beginning part of a syllable (a reason of possible diphthongization), it is directed to the place of the maximal energy point either on a vowel, a consonant matter or on some consonant in a consonant cluster (a reason for allowing relaxed syllable affixes, cf. Fujimura, Lovins 1978). At the same time, eligibility of the maximal energy point for different syllable segments in different words usually changes the character of short/long segmental opposition known in syllable-switching languages for a complementary distribution type of short/long opposition between vowels and consonants, thus guaranteeing fairly convenient conditions for foot isochrony (a strong negative correlation between neighbouring vowel and consonant, sometimes defined as loose vs. close contact). Foot-striving languages have a strong tendency towards foot isochrony. Data on prominent intra-foot temporal compression are available from a number of foot-striving languages (for English: Klatt 1973, Fowler 1981; for Swedish: Lindblom, Rapp 1973; for Dutch: Nooteboom 1972). A characteristic feature of foot-striving languages is the foot-final vowel reduction (for English: Wiik 1965, for Russian: Бондарко et al. 1966, for Swedish: Lindblom 1963). As for Russian, experimental data corroborate the model-predicted uncontrollability of the foot-final syllable quality. Here the full complexity of reduction cannot be established through the study of only the vowels themselves; the analysis of stressed and unstressed syllables has revealed that with the intensification of reduction the degree of coarticulation changes (Бондарко et al. 1966). Experiments with varied speech tempo have shown that at fast speaking rates the quality of the stressed vowel is not susceptible to qualitative reduction. The tendency to reduce the formant frequencies only shows up in unstressed vowels, and it does so even when the unstressed vowel has the same duration as its fast-rate stressed counterpart (Зиндер 1964). It is in accordance with the viewpoint that duration and articulatory effort may be controlled separately.

4) *Demisyllable-switching* foot control may emerge in languages with mono- and polysyllabic feet (e.g. Japanese; we consider e.g. the word *Sapporo* as consisting of a monosyllabic foot *sap-* and a polysyllabic foot *-poro*)

In the case of monosyllabic feet the turning point between tension and relaxation phases can be controlled in two principally different ways, i.e. by striving or switching commands. In Japanese, the striving possibility is refuted by the peculiarities of word rhythm (cf. Homma 1981). However, the domain of a syllable-switching command is the syllable beginning only and it cannot apply to the end of a monosyllabic foot. In this context, switching has to be shown on demisyllables both in and outside syllable beginnings, converting syllable-switching into demisyllable-switching. In short, a Japanese monosyllabic foot has to be controlled by two demisyllable-switching commands. As much as the temporal organization of a demisyllable is dependent upon inherent timing properties of underlying segments, we may expect that demisyllables represent temporally more or less equal units. In Japanese, where a demisyllable may be interpreted as a mora-size unit, Sawashima et al. 1982 have reported that the relative timing of articulatory and vocal pitch control is organized so as to compensate for timing variations in the internal mechanisms and maintain constant temporal relations in the acoustic output. The equality of the demisyllabic units is supported by the perceptual data (cf. Fujisaki, Horiguchi 1979). Modern Japanese, a language with the demisyllable-switching foot control, has neither isochronic feet (cf. the parallel existence of feet of two and three moras) nor syllables (cf. data in Homma 1981).

5) *Compound* foot control may emerge, inter alia, in languages with mono- and polysyllabic feet (e.g. Estonian).

We have presented the above four strategies as mutually exclusive options a language may follow. Note that our argumentation relied on the assumption that the chosen strategy is a mere inevitable answer to a physiological constraint, the alternation of tension and relaxation in feet, not directly used in meaning differentiation.

However, the strict boundary between the types may vanish in languages with contrastive meaning-differentiating accents, in case of which a special attention is paid to the energetic behaviour itself (cf. the psycholinguistic parallel of sound repetitions across morpheme boundaries that reveal the same twofold treatment: Menn, MacWhinney 1984).

In Estonian, in a quantity and accent language, syllable-switching and foot-striving control strategies together constitute an integral whole.

On the one hand, some essential features of Standard Estonian refer to the syllable-switching foot control: a) foot-initial short vowels are not diphthongized; b) non-foot-initial vowels do not show quality reduction; c) all 9 vowels and 17 consonants may occur as short or long phonemes (traditionally treated in terms of three quantity degrees, cf. Ariste 1938); a short/long opposition does not display any rules of complementary distribution (for the distribution of short and long phonemes see Eek 1986).

On the other hand, there are some substantial characteristics of foot-striving in Standard Estonian: a) all feet occur in either an even or in a sharp accent (see Table 1; for productional and acoustic data on accents see Eek 1986); b) a strong temporal compression is supported by the fact that mono- and disyllabic feet, irrespective of the accent type and segmental duration, differ durationally less than do intra-feet segments; among all segments that constitute a foot there exist a significant temporal relationship (cf. Lehiste 1972, Eek 1974); the foot tends to be as an isochronic unit of a temporal program (cf. Eek, Remmel 1974); c) the occurrence of vowels in non-foot-initial position is restricted, etc.

Note: in the discussion below, we use the following designations: → - a demisyllabic command, ↑ - a switching command, ↱ - a striving command.

Table 1
The appearance of the opposition 'even vs. sharp accent' in modern Standard Estonian

	even accent ´ (foot consists of 2 - 3 syllables)	sharp accent ` (foot consists of 1-3 syllables)
polysyllabic foot with a long first syllable	*k´auna* 'pod. GENITIVE'	*k`auna* 'pod. PARTITIVE'
polysyllabic foot with a short first syllable	*k´ana* 'hen'	∅
monosyllabic foot (obligatorily long syllables)	∅	*k`aun* 'pod'

2.4. *Stress-timing vs. syllable-timing*

Pike 1946 and Abercrombie 1967 are among the principal works that introduced the distinction between stress-timing and syllable-timing, two mutually exclusive, essentially different types of speech rhythm that a language may show. In a stress-timed language like English, stress beats were assumed to recur at approximately equal time intervals, in a syllable-timed language like French, syllables were assumed to recur at approximately equal time intervals. The two categories are viewed as mutually exclusive on the assumption that intervals between subsequent stress beats may be filled invariantly with a varying number of syllables.

The distinction between stress-timed and syllable-timed languages reflects undoubtedly some optional fundamental qualities of speech rhythm: note that it has been made use of in a large number of papers dealing with very different languages (cf. e.g. Lehiste 1977 and Dauer 1983 for a bibliography). However, the whole issue has been labelled as a linguistic controversy (Roach 1982). First, phoneticians do not agree with the view on speech production that the distinction implies: all syllables cannot be associated with separate chest pulses and stressed syllables cannot be extracted as special reinforced chest pulses (Ladefoged 1968). Second experimental psycholinguists deny the reality of any measurable timing difference between the rhythm types even in the prototypical opposition of English and French (Scott, Isard, Boysson-Bardies 1985). Third, both naive and expert listeners in experimental situations and comparative linguists in their theoretical treatments often disagree in attaching a particular language to either of the categories or, instead, claim that it belongs to neither (cf. Miller 1984).

To abandon these contradictions, we propose the model of speech rhythm in which 'stress' and 'foot' are considered as basic notions rather than 'stress' and 'syllable'. First, as for special short-term pulmonic activities in speech production, a foot rather than a syllable could be regarded the minimal integral domain to which they may apply (emphasized foot) but by no means need apply (tonic foot). Second, as for the basic non-timing nature of the distinction, the actual temporal rhythm pattern of a language could be viewed as deriving from two essentially non-temporal appearances of feet. Whether a language is stress-timed or not, depends on the interrelationship between feet and stress beats, accurately, on the frequency in which feet appear in continuous speech flow in the accentuated (reinforced) form displaying thus stress beats. Whether a language is syllable-timed or not, depends on the interrelationship between feet and syllables, accurately, on the manner in which syllables associate with the internal structure of feet. Third, as for the actual fuzzy boundary between the rhythm types in cross-linguistic research, the foot perspective entails a much more complicated picture of factor interplay than a strict twofold opposition between stress-timing and syllable-timing languages.

2.4.1. *Stress-timing: general motor rhythm vs. temporal speech rhythm*

Allen 1975 emphasizes the distinction between *general motor rhythm*, i.e. a pattern of sequence, and *temporal rhythm*, i.e. a pattern of temporal sequence. We have already introduced speech rhythm in the sense of a general motor rhythm, a successive sequence of feet. However, speech flow is subject to a universal bias towards a temporal rhythm as well. The portion of accentuated feet, displaying the highest energy levels, are conceived as *stress beats*, i.e. speech energy concentration sites in a continuum of speech flow that usually display more or less equal energy distribution. Like other sequences of rapid movements in human behaviour (e.g. finger tapping), subsequent stress beats tend to recur often regular intervals according to the properties of a universal physiological temporal rhythm pattern (cf. Allen 1975). In particular, stress beats are inclined to cluster around a mean interval of 0.4 to 0.5 s with an overall range of interval durations limited to between 0.2 to 1 s.

Languages are apparently *stress-timed* if they show clearly the temporal pattern of stress beats superimposed on speech flow by the physiological temporal rhythm, otherwise they are *non-stress-timed* displaying rather an intrinsic timing of their feet. In other words, the distinction 'stress-timing' vs. 'non-stress-timing' reflects properties of the real inter-foot timing phenomenon that is revealed between feet in continuous speech flow.

In all languages, there are obligatorily some feet that show accentuation to the extent that exceeds the so-called *stress beat threshold*. A language shows stress-timing, if such beats in continuous speech flow are *frequent* enough to converge into a pattern of temporal rhythm, and does not show stress-timing if they are too *rare* for that, cf. Fig. 2.

In typological research, it is often easy to decide whether a particular language is stress-timing or not. On the one hand, English (Dauer 1983) and Polish (Biedrzycki 1980) are obvious stress-timing languages because they display salient stress beats recurring frequently around 0.5 s (as a rule, indicating functionally that a new lexical entry is present in message) and are thus subject to a clear temporal rhythm. However, it does not mean that the so-called syllable-timed languages, on the basis of mean interbeat intervals, cannot belong to the same group with English (cf. e.g. Spanish: Navarro 1932). If lexical stress is saliently marked and lexical entries are not too long (not exceeding 4-5 syllables or 2 tonic feet) there is high probability to perceive such syllable-timing language as a stress-timing language.

On the other hand, French and Japanese (Dauer 1983) are obvious non-stress-timing languages because they display salient stress beats rarely after longer intervals than 0.5 s (as a rule, indicating functionally that especially important lexical entries of a phrase, clause, or sentence are present in message) and thus are not subject to a temporal rhythm. Rather long distance between stress beats in these languages cannot be accounted for in pure functional terms. If a language is characterized by non-prominent word stress and at the same time, accentuation does refer to every lexical entry, we can perceive a non-stress-timing phenomenon.

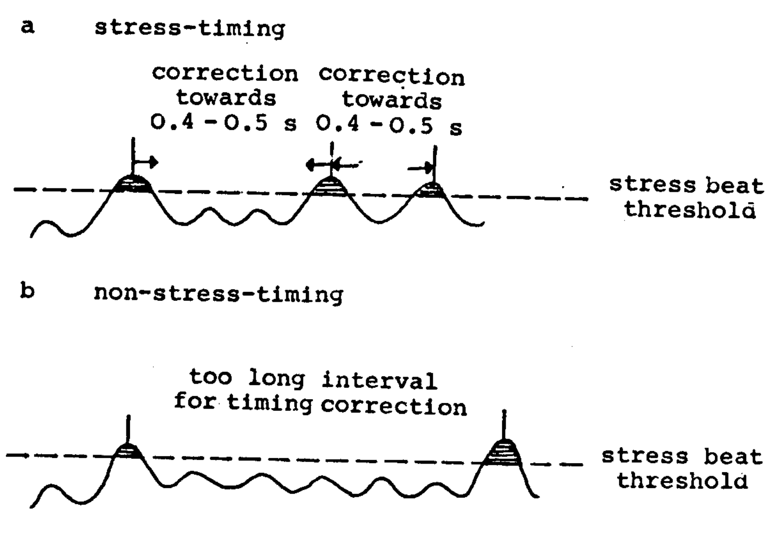


Fig. 2. Temporal rhythm superimposed (a) or not superimposed (b) on a sequence of feet.

We assume that it is because of the polarization between the non-stress-timing French and the stress-timing English that, in executing tasks of temporal rhythm manipulation, subjects who have French as their mother tongue show vagueness that is alien to subjects speaking English as first language (cf. data and discussion on the issue in Scott, Isard, Boysson-Bardies 1985). Nevertheless, we follow Dauer 1983 in claiming that whether a langue is stress-timed or non-stress-timed is a matter of degree rather than of mutual exclusion: many particular languages show neither obvious stress-timing nor obvious non-stress-timing but something between the extrema. For example, Finnish displays salient 'primary' stress beats (as a rule, indicating functionally that a new lexical entry is beginning in message) that occur too rarely in speech flow to converge into a pattern of temporal rhythm (in Finnish, an agglutinative language, the distance between the beginnings of subsequent lexical entries may be rather long). Consequently, Finnish is not subject to a clear temporal rhythm. However, a slight 'secondary' stress beat is provided by all other Finnish stress impulses as well so 'primary' and 'secondary' stress beats together are frequent enough to converge into the pattern of temporal rhythm. Accordingly, Finnish is still subject to a dim temporal rhythm (cf. the treatment of the Finnish problematics in O'Connor 1973).4

Table 2 summarizes our argumentation.

Table 2

The scale of stress-timing

less stress-timing ⟵		more ⟶ stress-timing
French Japanese	Finnish	English Italian

2.4.2. *"Syllable-timing"*

In polysyllabic feet of many languages, the average duration of a non-final syllable is generally under 0.2 s. In this context, syllables cannot be fundamentally subject to the general temporal rhythm pattern like that of stress beats (Dauer 1983) and as it is supposed by the term "syllable-timing".

We suggest that the term "syllable-timing" stands simultaneously for two different properties of the internal structure of feet the main common feature of which is the fact that they are alien to English, the prototype of "non-syllable-timing" languages. On the one hand, a language is "syllable-timed" if it displays a switching foot control. On the other hand, a language is "syllable-timed" if it has always one syllable for one striving or switching command (Table 3).

Table 3

The combinations of "syllable-timing"

	1) switching rather than striving	2) one switching/striving command for one syllable rather than one demisyllable or many syllables	
Finnish	+	+	↑ more "syllable-timed"
Japanese	+	-	
Vietnamese	-	+	
English	-	-	↓ less "syllable-timed"

3. THE GREAT RHYTHM SHIFT: OLD ESTONIAN

3.1. *Conservative Finnish vs. innovative Estonian*

In a number of general works on language (e.g. Anttila 1972, Comrie 1981), the comparison of the two main Balto-Finnic languages, Finnish and Estonian, serves to illustrate the point that genetically close-related languages may differ remarkably in respect of their typological characteristics. In outline, Finnish has preserved the original fairly clear-cut agglutinating morphology but Estonian has exchanged it for a morphology that is much more strongly characterized by fusion. The morphological differences are accompanied by crucial differences in the sound architecture of the languages. Finnish has preserved firmly the original Balto-Finnic prototype of long polysyllabic words that consist of simple syllables of the structure CV or CVC, display an extensive vowel harmony, and begin at an accentuated foot of an invariable quality. On the contrary, Estonian has introduced many short mono- and disyllabic words, complex syllables like CCVCC, and word-level restrictions on vowel distribution that have abandoned the original vowel harmony altogether. Estonian accentuated foot is mobile (a word need not begin at an accentuated foot), and variable (it displays either the even or sharp version of the contrastive accent).

These essential differences between modern Finnish and modern Estonian originated with a row of phonological changes that, on the one hand, occurred in the history of Estonian during the first centuries of the second millenium A.D. (roughly, 1100 - 1500) but, on the other hand, were absent in the history of Finnish. This claim has a high degree of confidence as it is supported by historical and comparative linguistic evidence and early textual data on Estonian. We concentrate on some central component changes of the row, cf. Table 4.

The general pattern of the below presented phonological changes includes an essential

Table 4
Some crucial phonological changes of Old Estonian in 1100-1500

(1) RISE OF CONTRASTIVE ACCENTS
(´ = EVEN ACCENT, ` = SHARP ACCENT)

**kaunan* 'chaff/pod.GENITIVE'	>	**k´aunan*	
**kaunaan* 'chaff/pod.ILLATIVE'	>	**k`aunan*	
Finnish *kaunan*		Estonian	*k´auna*
kaunaan			*k`auna*

(2) SHORTENING OF LONG VOWELS
IN NON-INITIAL SYLLABLES

**mustaa* 'black.PARTITIVE'	>	**musta*	
Finnish *mustaa*		Estonian	*m`usta*

(3) LOSS OF SHORT VOWELS
IN CERTAIN NON-INITIAL SYLLABLES

**kakkara* 'chamomile'	>	**kakkar*	
**kakkarasta* 'chamomile.ELATIVE'	>	**kakrast*	
Finnish *kakkara*		Estonian	*k´akar*
kakkarasta			*k`akrast*

(4) OBSTRUENT GEMINATION
BETWEEN A SHORT VOWEL OF AN INITIAL SYLLABLE
AND A LONG VOWEL OF A NON-INITIAL SYLLABLE

**käte(h)en* 'hand.ILLATIVE'	>	**kätteen*	
Finnish *käteen*		Estonian	*k`ätte*

incongruity. On the one hand, the changes are rather diverse as regards their typological characteristics. On the other hand, they still occurred during a relatively short time-span in an interconnected way. We treat the incongruity by claiming that the general pattern of these phonological changes is due to two subsequent purely phonetic changes, we call *the Great Rhythm Shift of Old Estonian*.

The predecessor of the modern Standard Finnish has been displaying a *syllable-switching* control of foot from time immemorial up to nowadays. However, the predecessor of the modern Standard Estonian was subject to a twofold reorganization of stress processing method before and at the time of the shift to the historic era in the Northern Europe. At the first stage of the Great Rhythm Shift, the original Balto-Finnic *syllable-switching* control of foot was abandoned in favour of an innovative *demisyllable-switching* control. At its second stage, the demisyllable-switching control was exchanged for a combination of *foot-striving and syllable-switching* control.

3.2. *From syllable-switching to demisyllable-switching.* At a time-point in the prehistory of most modern Balto-Finnic dialects, the loss of certain intervocalic obstruents, e.g. **mustaťa* 'black.PARTITIVE' > **mustaa*, introduced long vowels into non-word-initial syllables.

Before the change, the sound architecture of the dialects was constrained by two general principles. First, the opposition between short and long segmental vowels was possible only in a foot-initial syllable, e.g. there was a real opposition **tuli* 'fire' vs. **tuuli* 'wind' but any opposition of the type **tuli* vs. **tulii* was excluded. Second, the prototypical foot of the language was polysyllabic, e.g. the real feet **tuli* 'fire' and **tuuli* 'wind' of two syllables could not be accompanied by monosyllabic feet like **tul* or **tuul*.[5] The absolute prevalence of polysyllabic feet together with the presence of the opposition 'short vs. long segmental vowels' indicates that the Balto-Finnic dialects showed originally a syllable-switching control of foot, cf. Table 5:

Table 5
The syllable-switching control of foot in original Balto-Finnic

**ka na* ↳ ↳	'hen'
**kau na* ↳→ ↳	'chaff/pod'
**kau na ťa* ↳→ ↳ ↳	'chaff/pod.PARTITIVE'

After the change, to enable the pronunciation of the large number of words like **mustaa*, one of the original general principles had to be violated in order to retain at least the other. In Old Finnish, the prototypical foot remained polysyllabic but the opposition between short and long segmental vowels spread from foot-initial syllables to the rest of the syllables as well: **mustaa* was interpreted as a disyllabic foot with a long segmental vowel in its non-initial syllable. In Old Estonian, the opposition between short and long segmental vowels continued to be restricted to foot-initial syllables but many polysyllabic feet were replaced by monosyllabic feet so that the prototype of a polysyllabic foot vanished altogether: **mustaa* was interpreted as a string of two monosyllabic feet the latter of which contains a long vowel in its single 'foot-initial' syllable.

The option of Old Finnish, i.e. the elimination of a restriction on long vowel distribution, did not affect the conditions that determine the type of foot control and the original syllable-switching persisted, cf. Table 6:

Table 6
The syllable-switching control of foot in Old Finnish

**ka na* ↳ ↳	'hen'
**kau na* ↳→↳	'chaff/pod/'
**kau naa* ↳→↳→	'chaff/pod.PARTITIVE'

The option of Old Estonian, i.e. the introduction of the principle that monosyllabic and polysyllabic feet are prototypical to an equal extent, created conditions in which syllable-switching is impossible. In the context where the opposition between short and long segmental vowels has to be preserved (it was still necessary to differentiate between e.g. **tuli* and **tuuli*), the switching control of foot could not be exchanged for a striving control, which eliminates this opposition for sure. Consequently, the original syllable-switching control was exchanged for an innovative demisyllable-switching control, cf. Table 7 (— stands for a boundary between subsequent feet):

Table 7
The demisyllable-switching control of foot in Old Estonian

**ka na* ↳ ↳	'hen'
**kau na* ↳↳ ↳	'chaff/pod'
**kau - naa* ↳↳ ↳↳	'chaff/pod.PARTITIVE'

The different treatment of the long vowels in words like **kaunaa* guided Old Finnish (Table 6) and Old Estonian (Table 7) to the different paths of further development as discussed above. The option of Old Finnish created a stable situation that has been persisting without greater phonetic or phonological changes for centuries. For instance, the Table 6 could as well illustrate the situation in modern Standard Finnish. However, the option of Old Estonian created an unstable situation with which a row of pervasive phonetic and phonological changes originated.

At the *beginning* of the demisyllable-switching period of Old Estonian, some phonological changes provided phonological correspondences to the phonetic adjustments that accompanied the exchange of the foot control type. In Table 4, two of such phonological changes are included.

First, the final short vowel of a trisyllabic foot was dropped, cf. (1):

(1) **kak ka ra* 'chamomile' > **kak kar*
↳↳ ↳ ↳ ↳↳ ↳↳

A foot-internal syllable boundary is a very salient heuristic that points to a syllable-switching control of foot. Consequently, it is in conflict with a demisyllable-switching. The phonological change (1) is a reflection of a general phonetic process with which all trisyllabic feet (two foot-internal syllable boundaries) were eliminated so that any foot could be either monosyllabic (no boundary) or, in the extreme, disyllabic (one boundary). Note that, in this connection, the final short vowel of a disyllabic foot was retained, cf. (2):

(2) **kak ka - ras ta* 'chamomile.ELATIVE' ≡
↳↳ ↳ ↳↳ ↳

Second, as a result of the change *h* > *ø*, the word-internal obstruent at the beginning of a foot was geminated after a monosyllabic foot of one short syllable, cf. (3):

(3) **käte - hen* 'hand.ILLATIVE' >
↳ ↳ ↳↳
**kät - teen* (rather than **käø - teen*)
↳↳ ↳↳↳

An obstruent is subject to an ambisyllabicity at a word-internal foot boundary between two sonorous sounds. The phonological change (3), "emergency gemination", is a reflection of the situation in which the obstruent ambisyllabicity was phonetically reinforced in order to provide monosyllabic feet of one short syllable with additional sound material. Note that the obstruent ambisyllabicity was not reinforced phonetically and did not yield gemination phonologically after monosyllabic feet ending in a long vowel, diphthong, or sonorant. In this case, the necessary sound material of a monosyllabic foot was provided by the original long syllable itself, cf. (4):

(4) **jalka - han* 'foot. ILLATIVE' >
↳↳↳ ↳↳
**jal - kaan* (rather than **jalk - kaan*)
↳↳ ↳↳↳

3.3. *From demisyllable-switching to foot-striving*

In a prolonged time-span perspective on a language, feet that show the same degree of accentuation tend towards an average value of their stress amount. Within the demisyllable-switching period of Old Estonian, this goal was reached by manipulating the first and the second demisyllables in certain feet. At that time, the average foot consisted of three demisyllables. To approximate to its average stress amount, feet consisting of two or four demisyllables were altered. In two-demisyllable feet, the second demisyllable was reinforced, we designate the reinforcement by ¯ (it was inconvenient to reinforce the first, foot-initial demisyllable that had a reinforced value already by itself); in four-demisyllable feet, the first demisyllable was reduced, we designate the reduction by ˘ (Table 8).

Table 8
The trend towards the average stress value of feet in Old Estonian

*ka na	'hen'	2 ds	+	*[kana̅]
*ka nan	'id.GENITIVE'	3 ds	OK!	*[kanan]
*kau na	'chaff/pod'	3 ds	OK!	*[kauna]
*kau nan	'id.GENITIVE'	4 ds	–	*[kăunan]
*kau – (naa)	'id.PARTITIVE'	2 ds	+	*[kau̅]
*kau – (naan)	'id.ILLATIVE'	2 ds	+	
*maan	'earth.GENITIVE'	3 ds	OK!	*[maan]

The trend towards the average stress value of all feet was a statistical tendency that manifested itself clearly in long speech stretches. However, language users reduced it to a single phonetic rule that was applicable to every concrete foot. Because of the trend, both in two-demisyllable and four-demisyllable feet, the energetical value of the first two demisyllables became roughly equal, they revealed more or less the same amount of stress. Relying on this correspondence, language users deduced a general phonetic principle, *demisyllable balancing*, according to which the first two demisyllables of all feet have to display the same amount of stress. As for two-demisyllable and four-demisyllable feet, the rule was applicable without complications. However, as for three-demisyllable feet, it caused additional phonetic adjustments: in order to raise the energetical value of the second demisyllable to the level of the first demisyllable, some of the stress of the third demisyllable had to be reattached to the second demisyllable instead (Table 9).

The phonetic adjustments of the three-demisyllable feet in the context of demisyllable balancing triggered a chain of phonetic and corresponding phonological changes that exerted a crucial influence on speech rhythm. As the final result, they created conditions in which the demisyllable-switching control of foot was exchanged for a combination of foot-striving and syllable-switching control. In other words, they introduced the *ending* of the demisyllable-switching period of Old Estonian. In Table 4, this chain is represented by the phonological performance of its three main changes.

First, the final short vowel of a disyllabic three-demisyllable foot was dropped, cf. (5):

(5) *kak ka – ras ta 'chamomile.ELATIVE' > *kakk – rast

Table 9
The application of demisyllable balancing (the first two demisyllables of a foot display the same amount of stress) in Old Estonian (⇑ stands for a manipulated command)

2-demisyllable feet

*ka na > *[kana̅], *kau > *[kau̅]

1st ≈ 2nd : OK!

4-demisyllable feet

*kau nan > *[kăunan]

1st ≈ 2nd : OK!

3-demisyllable feet

*ka nan > *[kanan], *kau na > *[kauna], *maan > *[maan]

1st ≠ 2nd :

additional phonetic adjustments

*[kanan] > *[kana̅n̆], *[kauna] > *[kau̅n̆a], *[maan] > *[ma̅a̅n̆]

The result: 1st ≈ 2nd

This phonological change reflects the phonetic leveling of a consonant-ending foot like *[ma̅a̅n̆] and a vowel-ending foot like *[kau̅n̆a] by the shortening process *[kau̅n̆a] > *[kau̅n̆ᵃ] > *[kau̅n̆].

The above phonetic leveling ushered in an *overall phonetic shortening* of vowels in non-word-initial syllables. The phonetic process *[kau̅n̆a] > *[kau̅n̆ᵃ] was extended, for instance, to the final short vowel of a disyllabic two-demisyllable foot, e.g. to the second [a] in *[kana̅]. However, here a real phonological loss appeared only in tonic non-word-initial feet; in accentuated word-initial feet the vowel persisted, cf. (6):

(6) *ka va – la ta 'sly.PARTITIVE' > *kava – lat

Second, long vowels in all non-word-initial syllables were shortened, cf. (7):

(7) *mus – taa 'black.PARTITIVE' > *mus ta

This phonological change reflects the same overall phonetic shortening of vowels in non-word-initial syllables as in the previous case: on the analogy of e.g. *[kau̅n̆a] > *[kau̅n̆ᵃ] > *[kau̅n̆], the phonetic process *[kau̅ – naa̅] > *[kau̅ – naa] > *[kau̅na] started. Note that the shortening of long vowels replaced the original word-internal sequences of monosyllabic feet by innovative single disyllabic feet. In the original *[kau̅-naa̅], the second syllable consisted of two demisyllables and, thus, could form a separate foot. In the innovative *[kau̅na], the second syllable consisted of one demisyllable only and, thus, had to cohere with the preceding foot.

Third, and that is the crucial point, the opposition of two contrastive accents, the even accent (ˊ) and the sharp accent (ˋ), arose, cf. (8):

(8) *kau nan 'chaff/pod.GENITIVE' *kˊau nan
*kau – naan 'chaff/pod. ILLATIVE' *kˋau nan

This phonological change was an immediate functional result of the fact that the demisyllable structure of the innovative disyllabic feet (*kaunan from *kau-naan) merged often with that of some original disyllabic feet (*kaunan from *kaunan). This merger was functionally inconvenient, as it could give rise to a large-scale homonymy, cf. the difference of the grammatical meaning in *kaunan 'chaff/pod.GENITIVE' vs. *kau-naan 'chaff/pod.ILLATIVE' and of the lexical meaning in *kiiren 'ray.GENITIVE' vs. *kii-reen 'quick.GENITIVE'. In this context, the original phonological contrast between short and long segmental vowels in non-word-initial syllables (*a vs. *aa) was reinterpreted as a phonological contrast between two different types of feet (*kˊaunan vs. *kˋaunan) relying on those of their phonetic properties that had appeared differently in connection with the short and long vowels before the change.

The two types of disyllabic feet differed in the behaviour of the energy level of their first, second, and third demisyllable taken as an integral pattern. In feet like *kaunan from *kaunan, the demisyllables displayed an *even* energy level of a low value throughout the pattern, as it had been characteristic of the original 4-demisyllable feet. In feet like *kaunan from *kau-naan, the first two demisyllables displayed an even energy level of a high value, as it had been characteristic of the original 2-demisyllable feet but the third, shortened demisyllable displayed an energy level of much lower value. Correspondingly, there was a pattern of a broken energy level with a *sharp* projection at the boundary of the second and third demisyllables. The two different energetical patterns that resulted from different combinations of demisyllables were reanalyzed as two different energetical patterns that could characterize feet with one and the same demisyllable composition. After the reanalysis, the even pattern stood for the *even accent* characterized by an even energy distribution in the foot-initial syllable and a smooth beginning of the subsequent foot-internal syllable. Note that in the process of total accent split, the even accent was found not only in feet like *kˊaunan 'chaff/pod.GENITIVE' but also in feet like *kˊana 'hen' displaying a foot-initial short syllable. On the contrary, the sharp pattern stood for the *sharp accent*, characterized by an localized energy distribution in the foot initial syllable, i.e. a sharp movement to the maximal energy level at its end, and an apart beginning of the subsequent syllable. Note that in the process of total accent split, the sharp accent was found not only in feet like *kˋaunan 'chaff/pod.ILLATIVE' but also in feet like *kˋaun 'chaff/pod' of one syllable (Table 10).

Table 10
The accent split of Old Estonian

1. even pattern		>	even accent
*kau nan	'chaff/pod.GENITIVE'	>	*kˊau nan
*ka na	'hen'	>	*kˊa na
2. broken pattern with a sharp projection		>	sharp accent
*kau nan	'chaff/pod.ILLATIVE'	>	*kˋau nan
*kaun	'chaff/pod'	>	*kˋaun

It was a rise of contrastive accents that moved Old Estonian from the type of demisyllable-switching languages to the complex type in which foot-striving and syllable-switching are interwoven. As for striving, contrastive accents entail the manipulation of the maximal energy region in a foot that is out of the question in pure switching languages. As for foot-striving, there were polysyllabic feet in Old Estonian that excluded the possibility of syllable-striving. As for switching, the opposition between short and long segmental vowels had still to be preserved (it was necessary to differentiate between e.g. *tˊulen 'fire.GENITIVE' and *tˊuulen 'wind.GENITIVE'), so striving commands that eliminate the distinction for sure did not exchange the original switching commands entirely. As for syllable-switching, the equal prototypicality of polysyllabic and monosyllabic feet was no more a hindrance: monosyllabic feet like *kˋaun could be handled by a combination of a switching and a striving command rather than needed two subsequent switching commands in one syllable.

We have shown that the pervasive changes in the phonological structure of Old Estonian during the first centuries of the second millenium A.D. were accompanied by an equally thorough-going revision in the speech rhythm appearance of the language. Note that the path from syllable-switching via demisyllable-switching to foot-striving had to be especially manifest in the history of the Insular dialect of modern Estonian. Here, the sharp accent appears in the form of a *pure* striving command, cf. (9):

(9) kˊɔuna 'pod. GENITIVE'

kˋɔunə 'pod.PARTITIVE'

4. SOUND CHANGE PRESCRIPTION

According to Lass 1980, no inherent explanation is available for any linguistic change: he differentiates, inter alia, between deductive-nomological explanations and teleological explanations and argues against the possibility of either in the context of the diachronic research. On the contrary, Itkonen 1986 argues for the possibility of *teleological explanations* of linguistic changes: according to him, any linguistic change may be seen as contributing somehow to the increase of the form-meaning isomorphism in linguistic units. Still, for him, deductive-nomological explanations of linguistic changes are out of the question. We try to show that, in the framework where phonological-phonetic and purely phonetic sound changes are kept apart, the possibility of *deductive-nomological explanations* of linguistic changes cannot be excluded categorically.

A sound change may allow two types of formalizations with regard to formalization strength. A *description* represents it as a transformation that *occurred* given certain conditions, a *prescription* represents it as a transformation that *has to occur* given certain conditions. Prescription is preferable to description. Standing on a higher level of abstraction, it fits in with all functions a formalization of a sound change may serve rather than is appropriate for single explicit aims only. As a matter of fact, a deductive-nomological, "law-like" explanation of a sound change is equal to its prescription.

In order to provide a sound change with a prescriptive formalization, it is inevitable to extract exhaustively and arrange in a mutually exclusive, categorical manner all the *factors* that trigger the sound change in question. We have presented phonology as a social phenomenon and phonetics as a psycho-physiological and acoustic phenomenon. In this context, the prescription of a *phonological* change is precluded in principle: the sociolinguistic constraints on sound changes are many and fuzzy to the extent that any exhaustive factor extraction/arrangement is an insoluble task for a human being. However, some *phonetic* changes could allow a prescriptive formalization: the physiological and environmental constraints on sound changes are few and distinct to the extent that the needed exhaustiveness in factor extraction and arrangement may be achievable at least in some cases.

We have proposed that two different kind of sound changes should be extracted in the course of the sound history of a concrete language. As for *phonological-phonetic* changes, the above reasoning excludes prescription in principle; as for *purely phonetic* changes, it leaves the possibility open.

The essence of this claim is exemplified by comparing the actual course of the Great Rhythm Shift, i.e. a row of purely phonetic changes, and corresponding phonological-phonetic changes in the history of two different Balto-Finnic dialects, i.e. Standard Estonian and Southwestern Finnish.

5. THE GREAT RHYTHM SHIFT: THE SOUTH-WEST OF THE BALTO-FINNIC AREA

Up to here, we have treated the Great Rhythm Shift as present in the history of Estonian but absent in the history of Finnish. However, the straightfoward distinction crumbles if to consider the sound history of Estonian and Finnish dialects in addition to the history of standard languages. On the one hand, the conservative development has to be extended from Finnish alone to North-eastern Estonian also; on the other hand, the Great Rhythm Shift has to be extended from Estonian alone to Southwestern Finnish also (Map 1).

Map 1. The Great Rhythm Shift in Balto-Finnic (to the southwest of the line ⤥).

The presence of the Great Rhythm Shift both in the history of Standard Estonian and Southwestern Finnish provides a rather unique case of a row of phonetic and phonological changes that applied to the essentially same sound material (the original similarity of Balto-Finnic dialects) in different communities (the opposite coasts of the Gulf of Finland) not affected by pervasive mutual contacts (cf. the opinion expressed in the articles of the representative collection Gallén 1984). As a matter of fact, around 1000 A.D. both Estonian and Southwestern Finnish were influenced rather by Old Norse, the language of the Vikings. We follow Wiik 1986 in claiming that the Scandinavian vowel balancing could affect the speakers of the Balto-Finnic dialects in question: when the critical words like **mustaa* appeared, they followed the demisyllabic-switching option that was already known from the speech of the foreigners.

In this context, the exact comparison of the Estonian and Southwestern Finnish sound changes may display far-reaching theoretical implications.

On the one hand, the possibility of a prescriptive formalization of some purely phonetic changes is supported by the fact that the purely phonetic course of the Great Rhythm Shift had to be rather identical in both dialects. On the other hand, the fundamental impossibility of a prescriptive formalization of phonological-phonetic changes is supported by the fact that the phonological extensions of the Great Rhythm Shift did not coincide in the Estonian and Southwestern Finnish norm.

Indeed, some phonologizations could be the same for the two dialects. For instance, the Estonian and Southwestern Finnish patterns of vowel shortenings in non-initial syllables are rather similar. However, and that is the main point, the same phonetic changes could lead to rather different phonologizations as well. For instance, the counterpart of the Estonian "emergency" gemination in **kätehen* 'hand.ILLATIVE' > **kät-teen* (cf. the lack of gemination in **jalkahan* 'foot.ILLATIVE' > **jal-kaan*) was the Southwestern Finnish pervasive "special" gemination both in **kätehen* > **kät-teen* and **jalkahan* > **jalk-kaan*. Similarly, the shortening of long vowels in non-word-initial syllables triggered in Estonian the opposition of contrastive accents (**kaunan* 'chaff/pod.GENITIVE' > **k´aunan* vs. **kau-naan* 'chaff/pod.ILLATIVE' > **k`aunan*) but yielded a total merger in Southwestern Finnish (**kaunan*, **kau-naan* > **kaunan*). Note that, because of the latter difference, Estonian has turned into a foot-striving language but Southwestern Finnish still continues as a demisyllable-switching language.

NOTES

1. In addition, speech flow has the function to convey information about the person who is speaking (cf. Ladefoged 1984: 84) and about the extralinguistic, pragmatic situation in which he is speaking. Note, however, that these are not the <u>special</u> functions of speech flow, they are rather the functions of whatever human sound, be it e.g. speech, cry or wheeze.

2. We are aware of the bad connotation of the term 'speech energy'. In general, energy is something to be measured. Indeed, the exact amount of acoustic energy may be computed easily from the temporal intensity envelope for different frequency bands. However, the fixation of the exact value of the physiological energy while speaking or listening is a too complex task to be solved by exact measurements nowadays. Here, the proper term to indicate the generalization degree we mean would be rather 'the presence of muscular vs. neural activity' rather than 'physiological energy'. Nevertheless, the cover term 'physiological energy' has to be used in order to point to the fact that speaker's physiological activities, acoustic energy, and listener's physiological activities form a unitary chain. As a matter of fact, acoustic energy cannot be dealt with as a physiological activity.

3. In psycholinguistic literature, the term pair 'accentuation vs. deaccentuation' stands for our 'maximal vs. minimal accentuation'. We emphasize the gradual rather than the directional nature of the phenomenon.

4. Our account on stress-timing differs from that of Dauer 1983 as being based on a purely phonetic argumentation. Relying on the postulate that phonetics and phonology are incompatible within one treatment, we deduce the scale nature of stress-timing from the diversity of the possible patterns in which stress beats may be revealed rather than from an interplay of phonological, phonetic, lexical, and syntactic facts about a particular language.

5. In original Balto-Finnic, there was a handful of monosyllabic feet in addition, e.g. **maa* 'earth'. However, they could not affect the general prototype of foot polysyllabicity.

REFERENCES

<u>D. Abercrombie</u> 1967. Elements of General Phonetics. Edinburgh: Edinburgh University Press.

<u>G.D. Allen</u> 1975. Speech rhythm: its relation to performance universals and articulatory timing. - Journal of Phonetics 3: 75 - 86.

<u>H. Andersen</u> 1973. Abductive and deductive change. - Language 49, 4: 765 - 793.

<u>R. Anttila</u> 1972. An Introduction to Historical and Comparative Linguistics. New York: Macmillan.

<u>P. Ariste</u> 1938. A quantitative language. - Proceedings of the Third International Congress of Phonetic Sciences: 276 - 280. Ghent.

<u>T. Balasubramanian</u> 1979. Timing in Tamil. - Edinburgh University, Department of Linguistics, Work in Progress 12: 57 - 74.

<u>P.M. Bertinetto</u> 1981. Strutture prosodiche dell'italiano. Firenze: Accademia della Crusca.

<u>L. Biedrzycki</u> 1980. Isochronous feet in a reading of Polish verse. - Edinburgh University, Department of Linguistics, Work in Progress 13: 50 - 59.

B. Comrie 1981. Language Universals and Linguistic Typology. Oxford: Basic Blackwell.
R.M. Dauer 1983. Stress-timing and syllable-timing reanalyzed. - Journal of Phonetics 11, 1: 51 - 62.
A. Eek 1974. Observations on the duration of some word structures: I. - Estonian Papers in Phonetics 1974: 16 - 30.
A. Eek 1982. Stress and associated phenomena: a survey with examples from Estonian. I. - Estonian Papers in Phonetics 1980-1981: 20 - 59.
A. Eek 1986. Problems of Estonian word prosody. - Estonian Papers in Phonetics 1984-1985: 13 - 66.
A. Eek, M. Remmel 1974. Context, contacts and duration: two results concerning temporal organization. - Speech Communication Seminar, Speech Production and Synthesis by Rules, 2: 187 - 192. Stockholm.
C.A. Fowler 1979. 'Perceptual centers' in speech production and perception. - Perception and Psychophysics 25: 375 - 388.
C.A. Fowler 1981. A relationship between coarticulation and compensatory shortening. - Phonetica 38: 35 - 40.
O. Fujimura 1983. A linear model of speech timing. - Paper to be published in: In Honor of Ilse Lehiste, R. Channon, L. Shockey, eds.
O. Fujimura, J.B. Lovins 1978. Syllables as concatenative phonetic units. - Syllables and Segments: 107 - 120. A. Bell, J.B. Hooper, eds. North-Holland Publishing Company.
H. Fujisaki, N. Higuchi 1979. Temporal organization of segmental features in Japanese disyllables. - University of Tokyo, Research Institute of Logopedics and Phoniatrics 13: 155 - 161.
J. Gallén (ed.) 1984. Suomen väestön esihistorialliset juuret. - Bidrag till kännedom av Finlands natur och folk. Finska Vetenskaps-Societeten, 131. Helsinki.
R.E. Hammarberg 1976. The metaphysis of coarticulation. - Journal of Phonetics 4: 353-363.
Y. Homma 1981. Durational relationship between Japanese stops and vowels. - Journal of Phonetics 9: 273 - 281.
P. Howell 1984. An acoustic determinant of perceived and produced anisochrony. - Proceedings of the Tenth International Congress of Phonetic Sciences, vol. IIB: 429 - 433.
E. Itkonen 1986. Form-meaning isomorphism, or iconicity, in diachronic linguistics (and elsewhere). - Symposium on Formalization in Historical Linguistics, Tallinn, November 1986, Summaries: 38 - 46.
D.H. Klatt 1973. Interaction between two factors that influence vowel duration. - Journal of the Acoustical Society of America 54: 1102 - 1104.
P. Ladefoged 1968. Linguistic aspects of respiratory phenomena. - Annals of the New York Academy of Sciences 155: 141 - 151.
P. Ladefoged 1980. What are linguistic sounds made of? - Language 56, 3: 485 - 502.
P. Ladefoged 1984. 'Out of Chaos Comes Order'; physical, biological, and structural patterns in phonetics. - Proceedings of the Tenth International Congress of Phonetic Sciences, IIB: 83 - 95.
P. Ladefoged, J.D. De Clerk, M. Lindau, G. Papcun 1972. An auditory-motor theory of speech production. - UCLA Working Papers in Phonetics 22: 48 - 75.
R. Lass 1980. On Explaining Language Change. Cambridge: University Press.
W.A. Lea 1974. Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding. - Univac Report No. PX10791. St. Paul, Minnesota: Sperry Univac, DSD.
I. Lehiste 1972. Temporal compensation in a quantity language. - Proceedings of the Seventh International Congress of Phonetic Sciences: 929 - 939. The Hague, Paris: Mouton.
I. Lehiste 1977. Isochrony reconsidered. - Journal of Phonetics 5: 253 - 263.
J. Lehtonen 1970. Aspects of quantity in Standard Finnish. - Studia Jyväskyläensia 6. Jyväskylä.
A.M. Liberman, F.S. Cooper, K.S. Harris, P.F. MacNeilage 1962. A motor theory of speech perception. - Proceedings of the Speech Communication Seminar, vol. 2: Paper D 3, Stockholm.
B. Lindblom 1963. Spectrographic study of vowel reduction. - Journal of the Acoustical Society of America 35: 1773 - 1781.
B. Lindblom, K. Rapp 1973. Some temporal regularities of spoken Swedish. - PILUS 21: 1-59.
P.F. MacNeilage 1972. Speech physiology. - Paper prepared for the conference on Speech Production and Perception; Their Relation to Cortical Function, Vancouver, April 12-14, 1972.
S.M. Marcus 1981. Acoustic determinants of perceptual center (P-center) location. - Perception and Psychophysics 30, 3: 247 - 256.
L. Menn, B. MacWhinney 1984. The repeated morph constraint: toward an explanation. - Language 60, 3: 519 - 541.

P. Mermelstein 1975. Automatic segmentation of speech into syllabic units. - Journal of the Acoustical Society of America 58: 880 - 883.
M. Miller 1984. On the perception of rhythm. - Journal of Phonetics 12: 75 - 83.
T. Navarro 1932. Manual de Pronunciación Española. Centro de Estudios Históricos, Madrid.
S.G. Nooteboom 1972. Production and Perception of Vowel Duration. A Study of Durational Properties of Vowels in Dutch. Utrecht.
J.D. O'Connor 1973. Phonetics. Harmondsworth, Middlesex: Penguin Books.
J.J. Ohala, C.J. Riordan, H. Kawasaki 1979. Investigation of pulmonic activity in speech. - Proceedings of the Ninth International Congress of Phonetic Sciences, vol. 1: 205.
K.L. Pike 1946. The Intonation of American English. Ann Arbor: University of Michigan Press.
R. Plomp 1984. Perception of speech as a modulated signal. - Proceedings of the Tenth International Congress of Phonetic Sciences, II B: 29 - 40.
P. Roach 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. - Linguistic Controversies: 73 - 79, D. Crystal, ed., London.
M. Sawashima, H. Hirose, H. Yoshioka 1982. Interaction between articulatory movements and vocal pitch control in Japanese word accent. - University of Tokyo, Research Institute of Logopedics and Phoniatrics, Annual Bulletin 16: 11 - 19.
D.R. Scott 1982. Duration as a cue to the perception of a phrase boundary. - Journal of the Acoustical Society of America 71: 996 - 1007.
D.R. Scott, S.D. Isard, B. Boysson-Bardies 1985. Perceptual isochrony in English and in French. - Journal of Phonetics 13: 155 - 162.
J.M.B. Terken 1984. The effect of accentuation on comprehension: an experiment. - Proceedings of the Tenth International Congress of Phonetic Sciences, II B: 558 - 561.
M. Vayra, C. Avesani, C.A. Fowler 1984. Patterns of temporal compression in spoken Italian. - Proceedings of the Tenth International Congress of Phonetic Scienes, II B: 541-546.
K. Wiik 1965. Finnish and English Vowels. A Comparison with Special Reference to the Learning Problems Met by Native Speakers of Finnish Learning English. Annales Universitatis Turkuensis B, 94. Turku.
K. Wiik 1986. Suomi ja sen lähimmät sukukielet kahden germaanisen äännelain puristuksessa. - Paper presented at Institute of Language and Literature, Tallinn, April 4, 1986.
Л.В. Бондарко, Л.А. Вербицкая, Л.Р. Зиндер 1966. Акустическая характеристика безударности (на материале русского языка). - Структурная типология языков. Москва.
Л.Р. Зиндер 1964. Влияние темпа речи на образование отдельных звуков. - Ученые записки Ленинградского университета 325, Серия филологических наук 69: 3 - 27.
В.Б. Касевич 1983. Фонологические проблемы общего и восточного языкознания. Москва: Наука.
Л.А. Чистович, В.А. Кожевников et al. 1965. Речь. Артикуляция и восприятие. Москва, Ленинград.

GENERAL INDEX OF AUTHORS