RECORDING AND INTERPRETING ARTICULATORY DATA-- MICROBEAM AND OTHER METHODS

Osamu Fujimura

The Ohio State University
Division of Speech and Hearing Science
Columbus, OH 43210-1002, U. S. A.

ABSTRACT

Keating's paper "Phonetics in the Next Ten Years" is discussed with two foci: (1) instrumentation in speech production research and (2) the phonetics-phonology interface. These comments supplement two recent review articles by the author. A contributed paper in this conference by the author and his colleagues provides an alternative theory of phonetic implementation.

1. DEVICES FOR SPEECH PRODUCTION RESEARCH

Upon Keating's invitation, I shall discuss instrumentation issues in speech production research in the coming years in the context of a new emphasis on speech dynamics. Obviously I am biased toward the use of the x-ray microbeam for speech studies, because I have spent about 30 years proposing, designing, implementing and using the system. At that time, it seemed to be the only feasible method for observing the dynamics of speech production with little disturbance to the subject, in conformity with appropriate safety considerations. It took us a long time to make the system widely available; the U. S. Government (National Institute of Health) implemented a special support for the nationally shared research facility at the University of Wisconsin (Principal Investigator: J. H. Abbs) ten years ago. The system has now started producing systematically controlled articulatory data. A wide use of the system in full operational capacity is now foreseeable. In the mean time, there are some discussions of the possibilities of implementing similar facilities in Europe and in Japan.

Recently, two additional devices for observing tongue movements in speech have emerged. Both are still in a developmental stage, but are indeed promising, and may well be close to productive research use. I do not believe, however, that new methods will entirely replace the x-ray method in the next ten years. The advantages and disadvantages of all these methods shall be discussed below in some detail. My general view is that speech production research is so complex and difficult that no single method fulfills the purpose. Different approaches, including instrumental, theoretical, and computational, must be exploited as much as possible. For example, despite remarkable progress in technology, the traditional film analysis of cineradiographic images of the tongue contour still provides very useful information, particularly as a supplemental means for interpreting articulatory data obtained otherwise. The palatographic methods, now computerized for dynamic observations, will continue to provide three-dimensional information which most other methods cannot provide, at least as easily. There is much to be done in devising and combining different methods of articulatory observation and measurements. For instance, direct evaluation of muscle contraction states is one largely unexplored area, electromyography being severely limited in its applicability.

On the other hand, we have already seen much progress in our understanding of speech production. As Keating mentions, research has been progressing quickly with very important changes of direction.

While research results obtained by the use of ultrasonic and magnetic devices are only preliminary at this stage, so too are those from the microbeam studies, despite a much longer history. I hope this will be no longer true in a few years. Examples of good speech production research using new observational methods convince us of the critical importance of this type of research for understanding what speech is and how it is organized.

I have discussed this issue fairly extensively in a recently published review paper [5]. Therefore, I shall only add some remarks which have come to my attention since the time I wrote that paper. It is my understanding at this time that the magnetic method is particularly promising as a competitor to the x-ray microbeam. However, I still remain to be convinced with regard to its reliability and facility of use. Schönle's method [15] is even commercialized, and I think the device could have wide application in clinical and pedagogical areas. Perkell's method [13] seems to be producing some early research data. I would not be surprised at all if the magnetic device, quite possibly along with the palatographic device which already has proved remarkably successful in practical situations [15], becomes widely used in training deaf children to speak. However, my information is that, as a rigorous research tool for exact measurements of flesh point positions on the tongue, the system needs substantial improvement. At the time of the conference, perhaps, more convincing demonstrations may be given.

In principle, magnetic methods are superior to x-ray methods in that they do not involve any ionizing disturbances to the human body at all, and that the interference of metal objects in the mouth should be less severe. The current version of Perkell's system seems to be using very small coils, equivalent to the pellets for sampling flesh points on the tongue, probably at a severe cost of signal to noise ratio in the position determination. With regard to the microbeam, the pellet can be made smaller, also at some cost of detection reliability. The gold pellet being used at the University of Wisconsin facility is typically 2.5 mm in diameter. We should be able to reduce the size and weight considerably, possibly deviating from the spherical shape which we use now. The size issue may sound trivial, but it has major consequences. For example, if we reduce the diameter from 2.5mm to 2.0mm, keeping the same shape, the weight of the pellet is just about halved. Such a reduction in weight can bring about a qualitative difference in the usefulness of the system. In particular, the force required to retain the pellet in position will be reduced accordingly, and the choice of adhesive materials will be much easier, since the surface area is not reduced in the same proportion as the volume or weight. Also, it is quite possible that we can dispense with the string attached to each pellet for safety precautions. The chance of detachment becomes smaller partly because of less protrusion of the object embedded on the tongue surface (also causing less articulatory disturbance), and also partly because of the reduced inertia reacting to large acceleration of the tongue surface. Probably more importantly, the.chance of inspiration of a pellet, once detached, will be reduced substantially when the weight is reduced, because of the reduced gravity force relative to surface tension.

Currently, from a subject's viewpoint, the most disturbing aspect of the use of pellets is the string. If we can dispense with the string entirely, or if we can limit the use of the string to a short segment of string attached to the immediate surrounding area of the tongue surface, as opposed to the attachment outside the mouth on the facial skin as we do now, most of the disturbance will disappear except when we try to fix the pellet at the very tip of the tongue. This possibility is a distinct advantage of the x-ray method over the magnetic method. The magnetic method requires not just a retaining string, but electrically conductive lead wires due to its very principle (unless a small radio transmitter with a small battery together with the coil can be made comparable in size).

Ultrasonic methods have the distinct advantage of wide use in general medical diagnosis with consequent availability of commercially developed apparatus. The spatial and temporal resolutions available by this method, however, seem not quite satisfactory from our research point of view. The strongest concern I have, assuming some devoted effort to optimize details of the apparatus for the specific purposes of speech research, is the wave reflection at the boundary between flesh/water and air. After writing the review paper mentioned above, I have communicated with Dr. Stone, who recently published a new article on tongue movement [16], discussing some technical details of her method; I still maintain the same opinion about this point. It is inherently difficult for the ultrasonic method to avoid interference with the movement of the lower surface of the mandible, which easily deforms according to the force applied by the contacting device, even when a highly compliant material is used as an impedance matching medium placed between the solid and the skin. When the skin moves, the internal tissue becomes displaced, resulting in possibly quite significant distortion of the three dimensional shape of the tongue surface. This is particularly difficult in a dynamic situation, where the transmitter can not move according to the movement of the mandible (or even muscle contraction inside). As I mentioned in my paper, it is technically possible to avoid this problem by using a servomechanism in order for the device to follow exactly the moving skin with minimal force. Such a method apparently has been studied by Müller and his colleagues many years ago [12] in lip movement measurement, but I have no followup information about this interesting attempt. Servomechanism can be extremely effective in such devices, but requires very advanced engineering involving highly mathematical analyses. For ultrasonic methods to be reliably useful for general articulation research, however, I think it is necessary to resort to such advanced techniques.

The current method of computer-controlled x-ray microbeam for pellet tracking uses an extremely small radiation dose, thus making it possible to use the same subject for rather extensive speech material. As research makes progress, and the scope of study expands, from the main focus on robust segmental characteristics to more general issues of speech organization and principles of speech organization beyond the minimal scope of segmental concatenation, we need more and more data, with a great number of the factors of speech utterance under control. The data, to a large extent, must be collected from the same subject; yet subject-to-subject variation requires still larger amounts of data obtained from many subjects. Given more and more powerful data processing computational tools in combination with efficient and hazardless acquisition methods, the amounts of data we should use are expanding rapidly. This is probably the most remarkable change we have seen in speech production research. Until just recently, an articulatory study typically involved one to three speakers' data for about 100 seconds worth net total of speech materials each. Now, with the microbeam facility, we count on having several to even twenty speakers, each a total net (with convenient breaks between utterances) of typically 1,000 seconds per speaker in one session. As the data processing/interpreting methods advance, using for example neuronetwork and other AI type processing techniques, backed up by inexpensive computer memory, substantially more data can be effectively used for our studies without much difficulty. In this context, it is worthwhile to reexamine the radiation dose problems, which once appeared almost completely solved. One approach is to look at non-ionizing measurement methods, perhaps even at the cost of high accuracy of position assessment. There is, however, still some room for further improvement in this respect with regard to the x-ray microbeam method also.

One such innovation, which I have proposed in connection with the Wisconsin microbeam project, is the use of scattered photons within the scheme of x-ray microbeam. The current method attempts to capture all the photons being transmitted through the subject's tissues in the area immediately surrounding each pellet, thus optimizing the use of the radiation dose by not wasting any useful photon that comes through to the detector. The scintillation counter as the photon detector is almost ideal in capturing penetrating photons. What about the photons that do not penetrate, then? Those are not "useful" photons, from this point of view, but still cause ionization within the body. In fact, it is the non-penetrating photons, rather than the penetrating photons, that form potential harm to the body. Fortunately, a large portion of those photons are absorbed by the metal pellet. A significant portion of the photons that hit the pellet, however, are scattered. Some of these scattered photons, particularly those with characteristic energy of the metal used as the pellet, can be detected and identified as coming from the pellet. This would constitute a positive identification of the location of the pellet, rather than the current negative method which identifies the shadow of a pellet. In order to make use of these photons, we need a supplemental detector which covers a large solid angle other than the directly penetrating direction of the microbeam. I expect this new method to enhance the accuracy and reliability of the pellet identification, further reducing the necessary dose. An additional advantage is that with an appropriate choice of pellet material, interference from metal elements in the mouth due to dental work will be eliminated. Particularly if one side of the mouth is free of gold and other heavy metals in the useful area of the head profile, this method may prove critically helpful for the microbeam method, retaining its high accuracy and reliability.

2. MODELS OF PHONETIC IMPLEMENTATION

Keating, in Section 2.2 of her paper, discusses theory-related issues, emphasizing strongly the emerging importance of speech production studies in this context. There have been many discussions and advocations of new ways of relating phonology to phonetics, and some insights have been acquired indicating the future direction of research in this area. From my point of view, however, we still are short of any explicitly formal phonetic theory in general phonetics. The only complete (though still vague in many ways) theory of phonetic implementation is segment concatenated-smoothing (coarticulation theory basically since Lindblom, [9,10,11],see also, Fujimura [6]). On the other hand, the theory of nonlinear phonology has declared a basic departure from the segmentalism of this approach, but it is still not clear what this departure means. I suspect this issue will not be clear until we know exactly what the alternative, or supplement, to the coarticulation theory is going to be, or even could be. Browman and Goldstein [1,2] have proposed a very interesting idea, which seems radical enough to achieve the needed change; unfortunately, the picture is only conceptual and remains very vague, and many difficulties including some apparent internal inconsistencies must be resolved [3,7].

I have discussed some of these issues in my recent articles [4,5]. With my colleagues at OSU, I am currently engaged in designing a new model of the phonetic implementation process as a comprehensive (but rough and tentative) quantitative model, relating phonological and other specifications to articulatory control organization and then to acoustic signals. A contributed paper in this conference provides a sketch of this theory. This model is being used for an application of the abduction method [8], to automatically interpret microbeam pellet data of prosodic control [17,18]. Hopefully, some specific details of the model can be determined by this abduction data processing; at the same time, an empirical validation of the theory will be provided. Such studies crucially require a large database of articulatory recordings from natural utterances from systematically controlled speech materials covering a large number of factors, both phonological and extraphonological. Hopefully, as Keating suggests, theory-driven experimental work with extensive articulatory data, including this approach, will pave the way to understanding the linguistic and paralinguistic organization of natural speech.

3. REFERENCES

[1] Browman, C.P. & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V.A. Fromkin (ed.)*Phonetic Linguistics--Essays in Honor of Peter Ladefoged* (pp. 35-53), New York: Academic Press.

[2] Browman, C.P. & Goldstein, L.

(1989). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M.E. Beckman (*eds.*) *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 341-376). Cambridge: Cambridge University Press.

[3] Fujimura, O. (1989). Towards a model of articulatory control: Comments on Browman and Goldstein's paper. In J. Kingston and M.E. Beckman (*eds.*) *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 377-381). Cambridge: Cambridge University Press.

[4] Fujimura, O. (1990). Articulatory perspectives of speech organization. In W.J. Hardcastle & A. Marchal (*eds.*) *Speech Production and Speech Modelling* (pp. 323-342). Dordrecht: Kluwer Academic Publishers.

[5] Fujimura, O. (1990). Methods and goals of speech production research, *Lang.& Speech* **33**, 195-258.

[6] Fujimura, O. (1991).Comment: Beyond the segment. In I.G. Mattingly & M. Studdert-Kennedy (*eds.*) *Modularity and the Motor Theory of Speech* (pp.25-31). Hillsdale, N.J.: Lawrence Erlbaum Assoc. Publishers.

[7] Fujimura, O. (in press). Comment on Beckman-Edwards- Fletcher's paper. In Docherty & D.R. Ladd (*eds.*) *Laboratory Phonology II*. Cambridge: Cambridge University Press.

[8] Josephson, J.R. (1987). A framework for situation assessment: using best-explanation reasoning to infer plans from behavior. In *Proceedings of Expert Systems Workshop* (pp. 76-85). San Diego, CA: Science Applications International Corporation (SAIC-87/1069).

[9] Lindblom, B. (1963). Spectrographic study of vowel reduction, *J. Acoust. Soc. Am.* **35**, 1773-1781.

[10] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. J. Hardcastle and A. Marchal (*eds.*), *Speech Production and Speech Modelling* (pp. 403-439). Dordrecht: Kluwer Academic Publishers.

[11] Lindblom, B. (1991). The status of phonetic gestures. In I.G. Mattingly and M. Studdert-Kennedy (*eds.*) *Modularity and the Motor Theory of Speech Perception* (pp. 7-24). Hillsdale, N.J.: Lawrence Erlbaum Assoc. Publishers.

[12] Müller, E., Abbs, J.H., Kennedy, J., &Larson, C. (1977). Significance of perioral biomechanics to lip movements during speech. *Paper presented to the American Speech and Hearing Association, Chicago.*

[13] Perkell, J.S. & Cohen, M.H. (1986). An alternating magnetic field system for tracking multiple speech articulatory movement in the midsaggital plane. *Technical Report 512. Research Laboratory of Electronics, MIT.*

[14]Schönle, P.W. (1988). *Electromagnetische Artikulographie. Ein neues Verfahren zur klinischen Untersuchen der Sprechmotorik*, Berlin: Springer-Verlag.

[15]Shibata, S., Ino, A., & Yamashita, S. (1979). *Teaching Articulation by Use of Electro-Palatography.* (English Translation by K. Kakita, H. Kawasaki, & J. Wright, *ed.* by S. Hiki, 1982), Tokyo: Rion Co., Ltd.

[16] Stone, M. (1990). A three-dimensional model of tongue movement base on ultrasound and x-ray microbeam data. *J.Acoust.Soc.Am.* **87**, 2207-2217.

[17] Westbury, J., & Fujimura, O. (1989). An articulatory characterization of contrastive emphasis in correcting answers. *J Acoust.Soc. Am.* **85**, Suppl. 1, S98.

[18] Westbury, J. and Fujimura, O. (1990).Articulatory correlates of contrastive emphasis in correcting answers in English. *ATR Workshop in Speech Production and Perception, November 1990, Kyoto, Japan.*