GLOTTAL DETERMINANTS OF DEAF VOICE QUALITY

BEN MAASSEN

Interdisc.Inst.of Child Neurology
University of Nijmegen
P.O.Box 9101
6500 HB Nijmegen, The Netherlands

DIRK-JAN POVEL

Dept.of Experimental Psychology
University of Nijmegen
P.O.Box 9104
6500 HE Nijmegen, The Netherlands

ABSTRACT

Vocal fold vibration of nine deaf children was recorded with help of an electro-laryngograph. Analysis of the laryngographic waveforms yielded several parameters that can be used as an objective measure of instability of voice and deviating voice quality characteristics like breathiness, hoarseness and cul-de-sac. The analysis algorithms will be implemented in a visual speech training aid for the deaf.

INTRODUCTION

Even highly trained, experienced phoneticians exhibit great variability in their evaluations of deaf speech characteristics. This is especially true for suprasegmental aspects and voice quality [8,4,13]. As indicated by the labels typically used in descriptions of deaf voices, such as "too high, monotonous, breathy, nasal, cul-de-sac", voice quality is conceived of as the overall auditory colouring of an individual speaker's voice, to which both laryngeal and supralaryngeal features contribute. The latter refer to long-term muscular adjustments or "settings" of the articulatory organs [9]. For instance, cul-de-sac or pharyngeal focus of resonance [2] is caused by a tendency to retract or "back" the lingual body. Differentiating aspects of voice quality and articulation is complicated by the differential susceptibility [9] of individual speech segments to the biasing effect of a given supralaryngeal setting. That is, a nasalized voice has a different effect on nasal sounds (/m,n,ŋ/) than it has on vowels, plosives, fricatives and affricates. Also the decomposition of voice quality i glottal pulse shape and supralaryngeal effects is problematic, especially when the description is based on perceptual judgment.

To analyze the purely laryngeal aspect of voice quality we have used the electro-laryngograph (ELG) [5]. The ELG is an instrument that measures the electrical impedance of the vocal cords, thereby providing information about opening and closure durations. The laryngographic signal has been validated as a measure of vocal fold contact area by comparing the signal with registrations of subglottal air pressure, and measurements of glottal opening by means of photoglottography and high-speed filming [1,3]. The most stable characteristic of the Lx signal is the steep slope corresponding to the beginning of the closure phase [10], which provides an easy reference point for determining glottal pitch period (see Figure 1). Apart from the fact that a simple algorithm suffices to extract fundamental frequency, a major advantage of using the Lx signal instead of the acoustic speech signal is that period-to-period fluctuations in the waveform can be detected. Thus, measures of jitter (period-to-period frequency fluctuations) and shimmer (period-to-period amplitude fluctuations) are easily obtained, thereby providing information on regularity of voice.

In this paper we present a study of nine deaf children that were selected by a speech therapist to represent a broad range of vocal abnormalities. Laryngographic recordings of these children were analyzed in an attempt to extract perceptually and articulatory relevant parameters. The analysis algorithms will be implemented in a visual speech training aid to be used in therapy [15].

VOICE QUALITY ANALYSIS

Recording Procedure

Nine congenitally deaf children were selected by their speech therapists to represent a broad range of voice abnormalities. These children, 7 boys and 2 girls, ranging in age from 9 to 15 years, with a hearing loss of more than 100 dB (Fletcher Index) in the better ear, read aloud a series of phonetically balanced sentences and words. The acoustic speech signal together with the Lx signal were recorded on two tracks of a Revox tape recorder. Since recording on tape introduces phase distortions, especially in the low frequencies, the Lx signal was re-recorded while running the tape in reverse. (The acoustic speech signal was also re-recorded to preserve temporal alignment on a single tape.) After low-pass filtering (cutoff frequency 4 kHz, slope 24 dB/octave) both signals were fed into a stereo A/D convertor (sampling rate 10 kHz for both channels) and stored in computer memory. In the present study only Lx signals were analyzed, but the figures also show the corresponding acoustic speech signals.

Apart from the nine deaf speakers, two adult, hearing speakers, one male and one female, were recorded and analyzed. These speakers differed only with respect to fundamental frequency. An excerpt of the male voice is presented in Figure 2a for comparison purposes.

Determining Pitch Periods

All analyses of the Lx signal were performed in the time domain. Isolating individual pitch periods started by calculating the first derivative (Lx'). If Lx' exceeded a criterium value, which was about one tenth of the Lx amplitude, that point was taken to correspond to a steep slope at the beginning of a new pitch period (see Figure 1) and marked accordingly. When no steep slope was found, a positive zero-crossing was taken instead. By setting a minimal spacing between period markings of 16 sample points (which corresponds to a maximum frequency of approximately 600 Hz) and a maximum spacing of 120 sample points (corresponding to a minimal frequency of about 80 Hz), and at the same time have steep slopes take precedence over positive zero-crossings, a reliable pitch detection algorithm was obtained. In a comparison of the outcome of the algorithm (the period markings) and the original Lx signal no errors could be detected. After isolation of single pitch periods, very low frequency-components were removed by subtracting from each sample point the mean value of the period it belongs to. The thus adjusted waveforms were further analyzed to obtain an objective description of deviations in voice quality.

Types of Voice Quality Deviations

We will now present the different types of deviating glottal pulse waveforms that occur in our speech samples, together with the measures to describe them.

1. Instability of voice. The accessibility of individual pitch periods permit detection of stability of voicing and laryngeal articulation [14] within a very short time window. A grossly instable voice is displayed in Figure 2b. Here, within a single period, fundamental frequency drops from 280 Hz to 135 Hz. The example is taken from a 14 year old boy, who typically produced such patterns at the beginning of voiced segments following silence.

In Figure 2c an articulation error is displayed. Another 14 year old boy attempted to say the Dutch word /bana:n/ ("banana"), but produced /banda:n/ instead. During the /d/ vocal fold vibration drops to zero, which may be caused by supraglottal pressure build-up during the erroneous complete closure of the vocal tract, or by an incorrect abduction-adduction gesture of the vocal folds.

2. Jitter, shimmer, low-frequency components. On the microlevel, irregularity of successive pitch periods is expressed by high jitter and/or shimmer values. Jitter is calculated by dividing two successive period durations, shimmer by taking the logarithm of amplitude ratios. Whereas for normal voices under sustained phonation jitter values of 0.5% - 1.0% and shimmer values below 0.20 dB are obtained [7], in the example of Figure 2d, produced by an unintelligible 9 year old boy, jitter and shimmer rise to 25% and 11 dB respectively, giving the voice a hoarse quality. In addition, Figure 2d shows a low-frequency component, indicating vertical displacement of the whole larynx. This may be caused by retraction of the tongue during cul-de-sac voicing.

3. Deviations of isolated waveforms. Figure 3a displays a breathy voice. Calculated were the relative closure duration (duty cycle), defined as the relative position within the pitch period of the negative zero crossing (i.e. C/L, see Figure 1), and the relative area below the positive curve (A/(L*P), see Figure 1). Like Hasegawa et al. [6], we took the cosine of the duty cycle, to magnify the important range around 0.5. In normal voicing the cosine of the duty cycle centers around 0, the relative positive area around 0.25; in this breathy voice we found values of 0.7 and 0.15 respectively.

Figure 3b displays a different type of breathiness. Here, the breathiness is caused by insufficient steepness of the positive slope. Relative closure durations were calculated by dividing the number of samples between the onset of the period (in most cases not far from the onset of closure) and the waveform peak by the pitch duration. Whereas a value of 0.10 is typical (see also [10]), in Figure 3b a relative closure duration of 0.25 was found.

In Figure 3c a terribly hoarse voice is displayed. To capture the irregular character of these waveforms, the number of times the second derivative (Lx'') exceeded a criterium value, was counted. During normal voicing, very few pitch periods contained direction changes exceeding the criterium; in the example of Figure 3c a mean number of 2 per period was obtained.

Finally, in Figure 3d a falsetto voice is presented. Note the sinusoidal character of the waveform. A falsetto voice is not only of very high pitch (540 Hz in this example) but also breathy according to the relative closure duration criterium (mean value 0.20).

DISCUSSION

Analyses of glottographic signals obtained from nine deaf children yielded several parameters that are related to voice quality. Fourcin (personal communication) displays "raw" Lx waveforms to deaf children for training purposes. Since we believe that interpretation of the Lx waveform is problematic, especially for the youngest age groups, we are currently implementing the analysis algorithms described above in a speech training device [15], such that voice quality can be displayed in a simplified visual trace. For instance, jitter and shimmer, which indicate hoarseness, might be transformed to the visual dimension texture.

The speech training curriculum proposed by Ling [11], in which teaching of respiration and phonation precede articulation, stresses the importance of simplifying voice quality and displays for young children. Moreover, in previous experiments [12] we showed that voice quality and articulation - rather than temporal structure and intonation contour - are the most important determiners of deaf speech intelligibility.

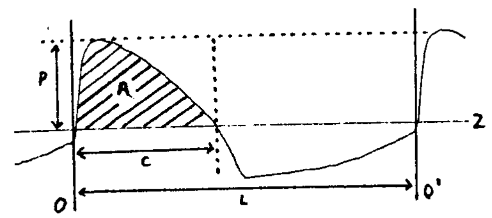

Figure 1. One period of a normal Lx waveform. Indicated are: O,O': start of closure at the onset of pitch periods; L: length (number of samples) of one period; Z: zero-line; P: maximal positive value; C: closure duration (until negative zero-crossing); A: area below the positive part of the curve.

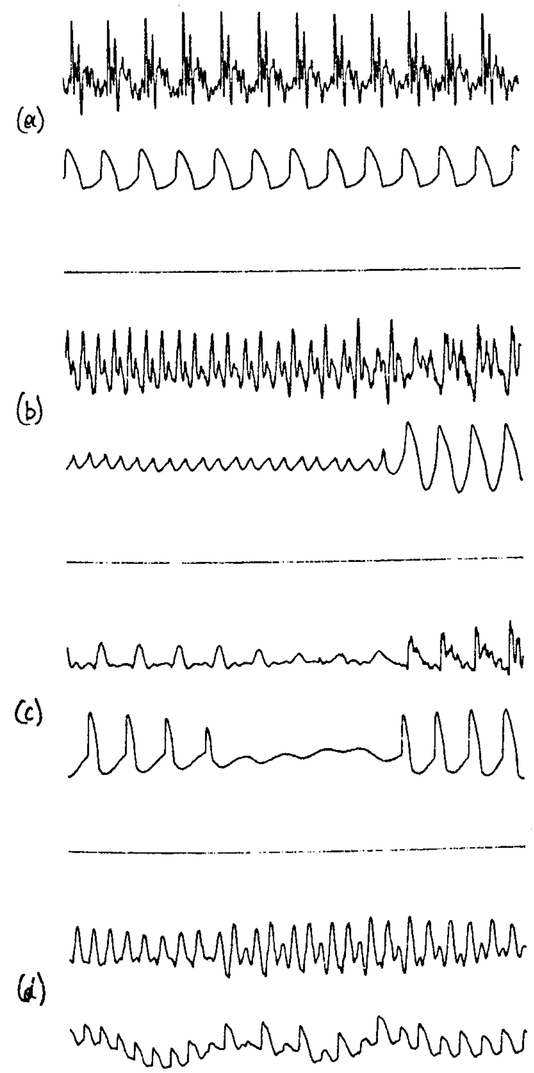

Figure 2. Sample registrations of Lx signals (lower traces) together with the acoustic speech signals (upper traces). Registration (a) id from a normal, male voice; (b), (c) and (d) represent incorrect period-to period fluctuations.

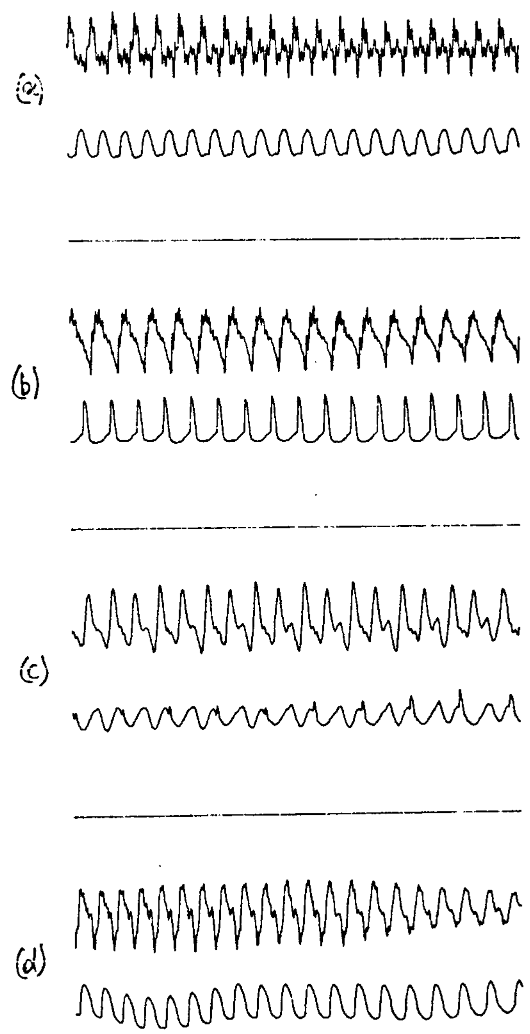

Figure 3. Sample registrations of Lx signals (lower traces) together with the acoustic speech signals (upper traces). In these examples the most noticeable deviation resides in the waveform of individual Lx periods.

Acknowledgement

This research was made possible by a grant from the Institute for the Deaf, St. Michielsgestel, The Netherlands.

REFERENCES

[1] Baer, Th., Löfquist, A. & McGarr, N.S. (1983). Laryngeal vibrations. A comparison between high speed filming and glottographic techniques. J.Acoust.Soc.Am., 73, 1304-1308.

[2] Boone, D. (1966). Modifications of the voices of deaf children. Volta Review, 68, 686 - 692.

[3] Boves, L. (1984). The phonetic basis of perceptual ratings of running speech. Doctoral dissertation, University of Nijmegen.

[4] Calvert, D.R. (1961). Some Acoustic Characteristics of the Speech of Profoundly Deaf. Unpublished doctoral dissertation, Stanford University.

[5] Fourcin, A.J. & Abberton, E. (1976). The Laryngograph and the Voiscope in Speech Therapy. In: E. Loebell (ed.) XVth International Congress of Logopedics and Phoniatrics, Basel: Karger.

[6] Hasegawa, A., Mashie, J., Herbert, E., Brandt, F., Metzler, M. M., Pickett, J. (1984). Real-time extraction of vocal quality parameters from electroglottographic signal. Paper presented at the 107th meeting of the Acoustical Society of America, Norfolk, Virginia, May 784.

[7] Heiberger, V. L. & Horii, Y. (1982). Jitter and shimmer in sustained phonation. In: N.J. Lass (Ed.) Speech and Language: Advances in basic research and practice (vol.7). New York, Academic Press.

[8] Hudgins, C. V., & Numbers, F. C. (1942). An investigation of the intelligibility of the speech of the deaf. Genetic Psychology Monographs, 25, 289 - 392.

[9] Laver, J. (1980). The phonetic description of voice quality. Cambridge University Press, London.

[10] Lecluse, F. L. E. (1977). Elektroglottografie. Doctoral Dissertation, University of Rotterdam.

[11] Ling, D. L. (1976). Speech and the Hearing-Impaired Child: Theory and Practice. Washington, D.C.: The Alexander Graham Bell Association for the Deaf, Inc.

[12] Maassen, B. & Povel, D.J. (1985). The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. J.Acoust.Soc.Am., 78, 877 - 886.

[13] Markides, A. (1970). The speech of deaf and partially-hearing children with special reference to factors affecting intelligibility. British Journal of Disorders of Communication, 5, 126 - 140.

[14] Metz, D. E., Whitehead, R. L. & Whitehead, B. H. (1984). Mechanics of vocal fold vibration and laryngeal articulatory gestures produced by hearing-impaired speakers. Journal of Speech and Hearing Research, 27, 62 - 69.

[15] Povel, D.J. & Maassen, B. (1987). Visual information and speech acquisition of the deaf. (This volume).