AUTOMATIC PROCEDURE FOR LARYNGOGRAPHIC (Lx) ANALYSIS OF PHONATION CONTRASTS

J. H. Esling**/*, B. C. Dickson*/** and J. R. Woolsey*

**University of Victoria, Canada
*Speech Technology Research, Ltd., Victoria, Canada

ABSTRACT
This report describes an automated, microcomputer-based procedure for comparing laryngographic (Lx) waveforms. The program enables researchers to analyze the cycle-by-cycle changes in vocal fold vibratory characteristics that may signal linguistic contrasts or differences in long-term voice quality. The first differential of the original Lx signal, captured digitally, is marked to indicate beginning, ending, upper and lower limits. A set of ratios is then obtained relating increasing voltage to decreasing voltage for each period of the signal. Considerations such as the inherent variability of the Lx signal, techniques of Lx recording, and applications of the algorithm are discussed.

1. LARYNGOGRAPHIC ANALYSIS
Electrical impedance laryngography has been used in phonetic research to quantify differences between contrasting types of phonation [4]. Such contrasts appear linguistically in languages like Korean at the syllable level in conjunction with phonologically distinct manners of consonantal articulation [1] [6] [7] [8] [9]. Phonatory contrasts also appear as long-term postures in voice quality with largely indexical significance [3, 10]. One problem in the analysis of the larynx waveform (Lx) has been the highly variable data that it yields. The signal is obtained by means of superficial throat electrodes which measure decreasing impedance as the vocal fold mass comes together, and increasing impedance as these structures separate [5].

Different models of laryngograph and differing recording procedures result in Lx signals with varying phase characteristics. This makes it difficult to analyze characteristics of individual waveform periods to distinguish, for example, a breathy voice from a harsh voice. Another problem is the DC float that characterizes many Lx signals and which makes establishing a baseline for reliable measurement of individual period characteristics particularly difficult. Aspects of obtaining an initial, workable Lx signal are dealt with in section 2. A solution to the baseline problem is presented in section 3. The method of segmenting the waveform to obtain a ratio is presented in section 4.

2. RECORDING Lx
Recordings of Lx signals made on a standard AM tape recorder tend to be distorted by phase shift. For this reason, it is valuable to develop procedures for direct digitization of the Lx signal, using an adequate (16-bit) data acquisition system. However, if this is not possible, the signal may be recorded using a system that does not introduce phase distortion, such as a Sony PCM digital audio processor and recorder, as has been used for these experiments.

Attempts at controlling DC float in the Lx signal include assuring that proper coupling with the input preamplifier or similar analog conditioner is maintained. However, low-frequency oscillations can be expected as a result of laryngeal movement around the axis of the electrodes. This is more apparent during continuous speech than during examples taken from sustained vowels, owing to the natural raising and lowering of the larynx in the less controlled situations.

The polarity of the Lx signal must also be

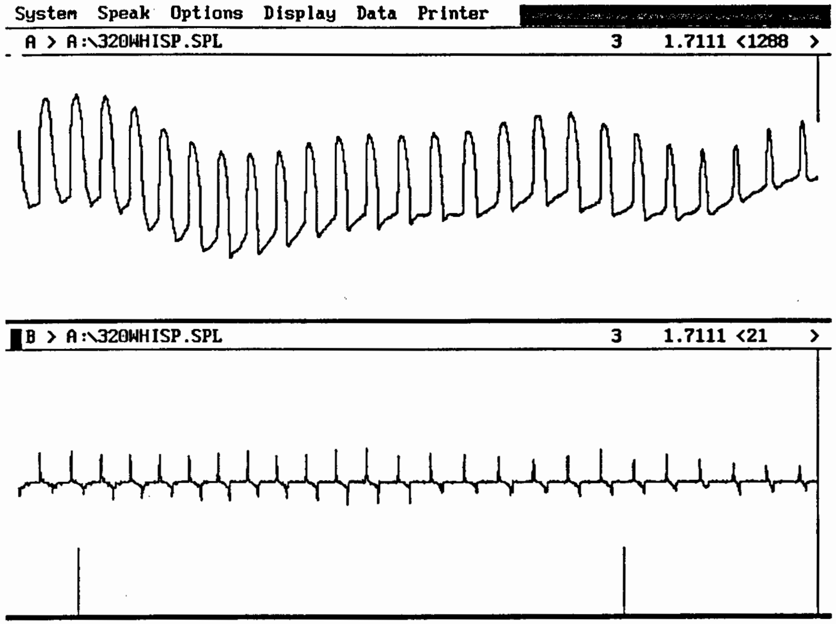

FIGURE 1. Original Lx waveform (top) and difference waveform (bottom) of a sample of whispery voice. Data marked (cursors), ready for ratio analysis.

considered. When the signal is taken directly from the laryngograph, the high impedance component of the signal is converted to the maximum positive voltage in the waveform, while the low impedance component results in a negative voltage. As the low impedance component is a result of maximum current flow across the electrodes, this will occur when the glottis is in its maximally closed phase of the voicing cycle. Because it is more representative of laryngeal behaviour to display the closed phase on the positive side of the waveform, we prefer to invert the polarity of a signal that has been taken directly from the laryngograph. However, if the signal is passed through a preamplifier at any stage, this will result in the polarity being reversed.

3. DATA ACQUISITION
Data acquisition is carried out using the CSL digital signal processing system [2], operating on an IBM-AT workalike. Data acquisition is performed at a rate of between 10K and 40K samples/second and the resulting sampled data files are passed to the EDIT320 software package. In that package, the waveform is displayed graphically and manipulated to enhance the laryngeal characteristics of interest (see FIG. 1).

The first differential of the waveform emphasizes the change in voltage over time, thus providing a representation of the Lx signal that closely models significant changes in current (as the impedance changes from, e.g., high impedance during the open phase of the laryngeal cycle to low impedance during the closed phase) as in equation (1).

(1) $d_i = y_i - y_{i-1}$

where $i = \{1, 2, 3 \ldots n\}$ sampled data points

A side effect of taking the first differential is that low frequency oscillations attributed to larynx movement, as well as DC float, are eliminated.

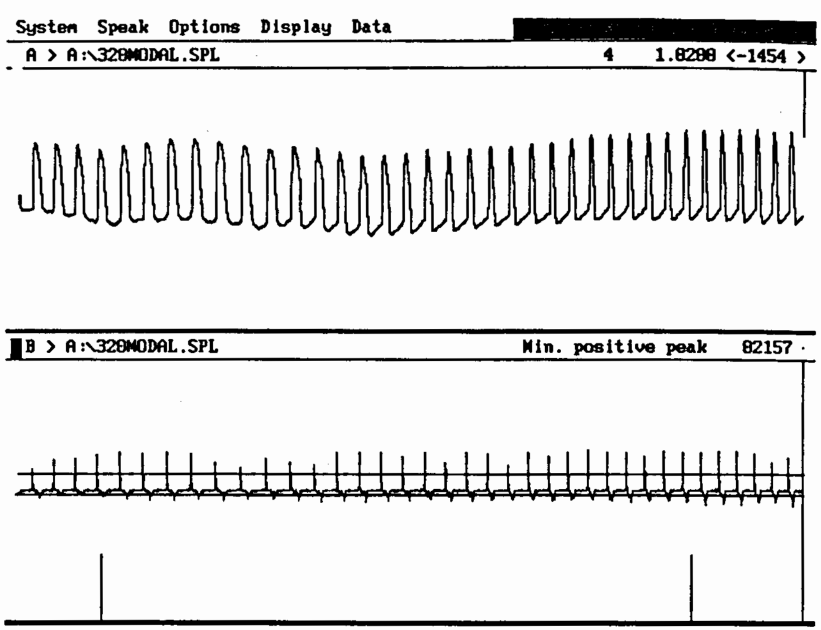

FIGURE 2. Lx waveforms of a sample of modal voice. Cursors marking minimum positive value and minimum negative value.

4. PROCESSING ALGORITHM

To derive the ratios for each Lx period of a range of voiced speech, the original Lx signal is loaded using EDIT320 and flipped if necessary (the sign is changed on each amplitude value) depending on recording conditions, and the first differential is computed. A minimum positive threshold and a minimum negative threshold are then selected, using horizontal cursors as illustrated in FIG. 2, to eliminate the effect of arbitrary zero crossings. For each period in the marked range, the greatest negative excursion (i_1) that is less than the negative threshold and the greatest positive excursion of the period (i_2) that is greater than the positive threshold are identified. A third value (i_3) is defined as the subsequent greatest negative excursion below threshold, beginning the following period. A ratio is then calculated for each Lx period as shown in equation (2).

$$(i_2 - i_1) / (i_3 - i_2) \tag{2}$$

where i = the selected sampled data point

For each succeeding period, the previous i_3 becomes the new i_1, until the end of the range is reached. The resulting ratios are stored in a file; as shown in FIG. 3, computed for a portion of the differenced signal for harsh voice.

5. APPLICATIONS

Applications of this analysis algorithm focus on the identification of phonatory differences at the segmental, CV, or long-term level. The hypothesis that a distinctive 'breathy' phonatory quality is associated prosodically with the lenis (vs. aspirated or fortis) consonant series in Korean as a principal cue in identifying meaning in CV sequences, for example, can now be tested. Lx rise-time to fall-time ratios of sets of controlled phonetic models can also be compared with specific language data or with examples of phonation in pathological speech.

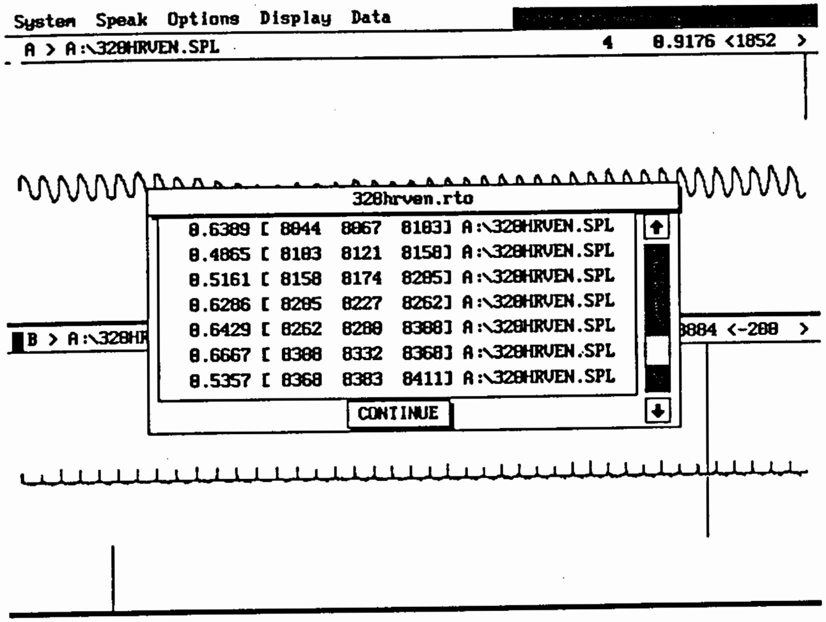

FIGURE 3. EDIT320 display showing part of the result file (ratio; start, peak and end point times for each period) for harsh (ventricular) voice.

Initial examination of phonation types using this procedure illustrates that harshness and creakiness, which have low Lx ratios, differ from modal voice, and from whisperiness and breathiness, which increase progressively in Lx ratio range, as predicted in prior research [3].

6. REFERENCES

[1] ABBERTON, E. (1972), "Some laryngographic data for Korean stops", *JIPA, 2*, 67-78.

[2] DICKSON, B. C., et al. (1990), *CSL User's manual*", Pine Brook, NJ: Kay Elemetrics Corporation.

[3] ESLING, J. H. (1984), "Laryngographic study of phonation type and laryngeal configuration", *JIPA, 14*, 56-73.

[4] FOURCIN, A. J. (1974), "Laryngographic examination of vocal fold vibration", *Ventilatory and phonatory control systems* (B. Wyke, ed.), 315-333. London: Oxford University Press.

[5] FOURCIN. A. J., & ABBERTON, E. (1976), "The laryngograph and the voiscope in speech therapy", *XVIth Int. Congr. Logop. Phoniatr.*, 116-122. Basel: Karger.

[6] FUJIMURA, O. (1977), "Control of the larynx in speech", *Phonetica, 34*, 280-288.

[7] IVERSON, G. K. (1983), "Korean *s*", *Journal of Phonetics, 11*, 191-200.

[8] KIM, C-W. (1965), "On the autonomy of the tensity feature in stop classification (with special reference to Korean stops)", *Word, 21*, 339-359.

[9] LADEFOGED, P. (1973), "The features of the larynx", *Journal of Phonetics, 1*, 73-83.

[10] LAVER, J. (1980), *"The phonetic description of voice quality"*, Cambridge: Cambridge University Press.