LARYNGOGRAPHIC ANALYSIS OF PHONATION IN KOREAN CONSONANT-VOWEL SEQUENCES

John H. Esling

Department of Linguistics
University of Victoria

1. LARYNGOGRAPHIC ANALYSIS

1.1. Manner of Articulation in Korean

Perceptual tests of Korean CV syllable sequences (Esling, 1988) indicate that information carried by vowel quality alone can be sufficient to identify whether the manner of articulation of a preceding consonant is 'lenis', 'aspirated' or 'fortis'. This finding suggests that it is important to reinvestigate the type of laryngeal activity that occurs during the vowels following lenis (medium-aspirated), (full) aspirated and fortis (so-called 'tense') consonants, in order to determine how characteristics of phonation type provide a cue to their identification. The method of investigation involves an experimental phonetic procedure developed at University College, London, in the early 1970s to isolate the direct larynx waveform (Lx) by means of an electrical impedance laryngograph. Laryngographic analysis has the advantage of separating the larynx (source) signal from the effects of supralaryngeal vocal tract modification, and provides information about the state of vocal fold vibration even when the glottis is closed (Fourcin, 1974). Research on Korean stop consonants indicates that the fortis series 'has a considerable degree of glottal stricture' (Ladefoged, 1973, p. 81), and exhibits 'laryngealized' or 'glottalized' laryngographic characteristics (creaky voice) immediately following voice onset (Abberton, 1972). In Abberton's laryngographic study (1972, p. 75), the lenis series, although purportedly followed by breathy vowels (Kim, 1965), occasionally demonstrates 'breathy voice onset', but no more consistently than the aspirated series. The present study introduces a new approach to analyzing the characteristics of the larynx waveform, to reexamine these relationships and to reassess these earlier findings.

1.2. Data Acquisition

A sample of Korean words was recorded, illustrating CV-initial syllables of the three manners of consonantal articulation in minimal triplets and pairs. Items are the same as those recorded in the Phonetic Database (PDB) of languages of the world in the Department of Linguistics (Esling, 1987). In this case, recordings were made of both the acoustic signal (with a Sennheiser microphone) and the output of an electrical impedance laryngograph by means of superficial electrodes placed on either side of the throat at the level of the thyroid cartilage. Both signals were captured digitally through a Sony Pulse Code Modulator to a Beta cassette recorder. Data include one example each of vowels /i/, /u/, /ɑ/ following each type of initial *p-*, *t-*, *k-*, *tʃ-*, and one example of a minimal triplet for each of the vowels /o/, /ə/ and /æ/ Minimal pairs for each of five vowels following the two types of Korean *s-* are also included. For comparison of the Korean data with the laryngographic characteristics of phonetically modelled phonation types, six contrasting phonation types were recorded by the author using the same equipment and procedure, and a balanced phonetic text. Laryngographic analysis of these phonetic models constitutes a repetition, using the revised microcomputerized analysis algorithm, of an earlier experiment to define and categorize phonation types laryngographically and laryngoscopically (Esling, 1984).

The laryngograph signal of each Korean item, and of each model phonation type, was transferred from Beta format to MS-DOS format using the *Computerized Speech Lab (CSL)* with a TMS320C25 digital-signal-processing board. Five seconds of each model phonation type were also transferred to disk files following this procedure. The resulting Lx signals present some differences from the types of waveforms obtained previously (Esling, 1984), due to differences in the type of amplification equipment used and the specifications of the laryngograph. Unmodified signals, viewed using *CSL* as if Lx were an acoustic signal, demonstrate considerable DC float and a nonlevel baseline, as in the top screen of Fig. 1. To quantify and compare these Lx waveforms, a preprocessing treatment was devised and an algorithm developed to allow automatic processing of the Lx signal.

1.3. Algorithm Development

Lx waveform analysis is carried out in the *Micro Speech Lab 320* environment, operating on an IBM-AT workalike, using a modified version of the speech/sound waveform analysis program *EDIT320* to accommodate unmodified Lx waveforms, account for DC float, and eliminate the effect of the nonlevel baseline in calculating a ratio of waveform rise time to fall time. Input files captured with the electrical impedance laryngograph (or 'raw' Lx) are loaded the same way as speech files, marked for detailed analysis, and viewed on a scale where the cursor moves one pixel at a time. The algorithm operates with two cursors, the first of which is used to mark the point corresponding to onset of closure of the vocal folds, the beginning of the rapid rise indicating decreasing impedance across the larynx. The second cursor is then moved to the next onset of closure, at the beginning of the subsequent Lx period. When the second point is marked, a third mark is automatically calculated and inserted at the highest value between the two, or peak of Lx, and a fourth mark is calculated and drawn where the line joining the two periods intersects the fall of the Lx trace. The ratio of rise time (from base to peak) to fall time (from peak to baseline intersection) is then calculated and displayed. In cases where marks are placed at locations which do not accurately reflect the observed rise or fall of a period, any mark can be repositioned using the cursor, and a new ratio obtained. Generally, repositioning is only required to reset the intersection of the fall, or to adjust marks where short-term plus-minus fluctuations of the signal interfere with automatic identification of the correct period peak. An output 'ratio' file can then be created for iterative storage of the ratios of an utterance and for subsequent statistical processing.

Tests of the contrastive model phonation type data using this initial procedure, however, show little of the expected differentiation of phonation types according to Lx rise-time-to-fall-time ratios based on the evidence in Esling (1984). A revised method of preprocessing Lx waveform data using an additional *CSL* analysis function appears to solve this problem. Raw Lx files are converted to *CSL* format and loaded for preprocessing, with normalization applied at each step. This procedure then entails flipping the data if amplification characteristics during recording had caused the signal to invert. Then, in order to flatten the baseline of the Lx signal and to enhance the changes in velocity of the signal, preemphasis is applied to the original Lx ($d_i = y_i - y_{i-1}$). The resulting differenced waveform, shown in the bottom screen of Fig. 1, is a reflection of the greatest change in rising velocity (analogous to vocal fold closure) and of the greatest change in falling velocity (roughly analogous to opening). The ratioing algorithm can be applied to the preemphasized Lx signal as before, by placing the marker cursors at successive negative peaks.

The resulting ratio is an expression of rise time (from the point of greatest velocity of increasing impedance to the point of greatest velocity of decreasing impedance) to fall time (from the point of greatest velocity of increasing impedance to the point of greatest velocity of decreasing impedance). Values thus obtained, instead of being between 0 and 1.0, are likely to exceed 1.0 and to range up to 4.0 or 5.0 or more, depending on the location of the Lx period in the syllable. Nevertheless, this procedure yields a reasonable distribution of measurements, judging by the results for the model phonation types.

2. RESULTS OF LX ANALYSIS

The distribution of the differenced Lx-waveform ratios for the set of model phonation types confirms the relationship between phonation types established in Esling (1984) (see Fig. 2). Pitch range, average ratio and ratio range data for each phonation type are based on calculations for all Lx periods in the first five seconds of a balanced phonetic text, in French in this case (Harmegnies & Landercy, 1986) for the sake of comparison with the results obtained previously for English (Esling, 1984). Between 200 and 500 pitch periods have been analyzed for each type. *Breathy voice, whispery voice, modal voice* and *harsh/ventricular voice* fall within the same pitch range, between about 110-150 Hz, but are distinguished hierarchically by decreasing Lx ratios. The three classic 'registers' are distinguished on the basis of both Lx ratio and frequency. *Creaky voice,* generally below 110 Hz, has low Lx ratios in the same range as *harsh/ventricular voice.* Mean and median Lx ratios for *modal voice,* in the mid-frequency range by definition, lie above the limits of the range of *creaky voice,* and below the limits of the range of *falsetto* at the 65% level of confidence. *Falsetto* has both a high frequency and high Lx-ratio range. Mean and median values and ranges at 65% and 95% confidence levels are presented in Table 1. These observations provide a benchmark for the interpretation of Lx data from Korean.

Findings for the Korean vowel data are grouped first according to vowels following *p-*, *t-*, *k-*, or *tʃ-*, and then according to vowels following *s-*. Initially, results are presented for all of the vowels in the environment of each manner of consonantal articulation, and then for each vocalic environment individually.

While vowels following fortis consonants are consistently lower in Lx ratio in stop and affricate environments, results for vowels following lenis and aspirated consonants are virtually identical. Thus, vowels in fortis stop, affricate and /s+/ environments are distinct from their lenis and aspirated counterparts in both mean and median Lx ratio, but vowels in lenis and aspirated stop, affricate and /s/ environments are undifferentiated in mean and median Lx ratio when all vowels are considered together. These global results are presented in Table 2.

Considering each individual vocalic environment separately gives a very different perspective on the behaviour of phonation in the contrasting consonantal environments. A vowel-by-vowel analysis of the data confirms that Lx ratios of vowels following aspirated consonants have higher ratios than vowels following lenis consonants, which in turn have higher ratios than vowels following fortis consonants (see Fig. 3). There is, however, a cross-over in the case of the /ɑ/ vowel, where the lenis environment shows the highest ratios. It should be noted that in the present comparison only the /u/, /i/ and /ɑ/ vowel categories contain four triplets of vowel tokens of n=175, 190 and 240 periods, respectively. Other categories contain fewer sample tokens and Lx periods for comparison. Considering only /u/, /i/ and /ɑ/ as significant, the vowel cross-over phenomenon persists as a salient characteristic of the present data. Thus, only the open vowel /ɑ/ following *p-*, *t-*, *k-*, *tʃ-* demonstrates the predicted higher Lx ratio in the lenis environment. Other vowels demonstrate a higher Lx ratio in the environment of the aspirated consonant series.

Similarly, results for **s**- suggest initially that the fortis sibilant is accompanied by lower Lx ratios during vocalic phonation for all vowels. Some vowels appear to contrast more sharply than others, but more data are required to establish significance. Absolute Lx ratios vary a great deal from vowel to vowel. Whether Korean /s/ demonstrates more aspirated or more lenis characteristics, based on the criterion of a following vowel, cannot be determined from these data because of the possibility that different vowels may exert a variable influence on the perception of /s/ as more aspirated or more lenis. Caution is advised in interpreting these preliminary data, since each vocalic environment for **s**- in Fig. 3 represents only one minimal pair of n=36 Lx periods on average, except for /ɨ/ with two pairs.

3. DISCUSSION AND CONCLUSIONS

The pattern of Lx ratios for the Korean vowel data confirms that the three manners of consonantal articulation are indeed distinct, as predicted, according to the phonatory quality of the following vowel. A major qualification of this conclusion is that the Lx characteristics associated with manner of consonantal articulation vary depending on the identity of the particular vowel. Comparing Lx ratios obtained for Korean vowels with the ratios of modelled phonation types, these findings suggest that lenis initial consonants may be identified on the basis of *breathiness* (as distinct from *creakiness* or *harshness* in the case of fortis consonants) only in the open vowel /ɑ/ The hypothesis that a higher ratio, consistent with *breathy voice,* will characterize the vowels associated with lenis consonants is therefore not supported in the case of /o/, /u/, /i/, /æ/ or /ə/. Since the phonatory quality of the vowel has been said to characterize the lenis series, without reference to the articulatory identity of the vowel, the most plausible explanation that can be offered in the light of these data is that lenis syllables most often have a *whispery creaky* or *harsh whispery* phonatory quality rather than a *breathy* quality. The conclusion is inescapable that, for most vowels, greater *breathiness* in the phonatory quality of the vowels, as a function of Lx ratio, is a property of aspirated and not of lenis syllables. The assertion that fortis stops are accompanied by a vowel of 'laryngealized' quality (Ladefoged, 1973, p. 76) is supported by the evidence presented here of low Lx ratios for that series, although it cannot be determined by Lx ratio whether the vowel is *creaky* or *harsh/(ventricular)* or both.

If the type of phonation observed varies with the quality of the particular vowel, a new interpretation of the role of laryngeal quality in the perception of the Korean consonant series would appear necessary. Lx ratios of the sort derived here are not known to differ according to supralaryngeal vowel quality, and have been assumed not to do so. If this assumption is incorrect, then that finding would be a significant result in its own right. If in fact supralaryngeal configurations can be demonstrated not to affect Lx, then the results obtained here remain of considerable interest, because of their implications for theories of perception of sequences such as Korean CV syllables. If the results presented above can be demonstrated in more extensive tests to be correct, the environment (the vowel in this case) following a phonemic contrast (initial consonants in this experiment) could play a more prominent role in the description and identification of that consonantal distinction. Some realizations of those consonantal phonemes may in fact retain only a few of the features that normally distinguish them, compared with other instances of the same phoneme. Such cases suggest environments that are likely candidates for language change -- the shifting of an item from one perceptual category to another. A related implication for Korean *s*- would be that /s/, as distinct from /s+/, may be perceived as more or less aspirated, or as more or less lenis, depending on the vowel which follows. Some phonemic distinctions might also be in the process of being lost where the inventory of factors that signal the identity of /s/ vs. /s+/ is reduced. Further testing is therefore necessary with additional subjects and for the full inventory of Korean vowels, to assess the behaviour of Lx in varying vocalic contexts.

3.1. Summary of Korean Manners of Articulation

The literature on Korean obstruents and the results of Esling (1984, 1988) and of the present study indicate that the three contrastive manners of articulation are distinguished primarily by onset characteristics but also by following-vowel phonatory characteristics. The aspirated stop demonstrates longest voice onset time (VOT), around 100 msec; the lenis stop is partially aspirated, around 40 msec VOT; and the fortis stop demonstrates shortest VOT, around 20 msec with virtually no aspiration noise present. These observations conform to the temporal changes in glottal width found by Kagaya (1974) for the three types of consonants. Aspirated and lenis stops generally begin with similar glottal width, followed by an increase in width for [h] for the former but reduced width during [h] of the latter. Fortis stops begin with narrower glottal opening and decrease rapidly to tight closure of the glottis prior to articulatory release and voicing. Electromyographic studies have shown, in addition, that the vocalis muscle is distinctly active in the fortis or 'forced' series immediately prior to the release of oral stop closure (Hirose et al., 1974; Fujimura, 1977). "It is noted also ... that the Korean aspirated ... stops show some momentary activity of the vocalis muscle immediately preceding voice onset" (Fujimura, 1977, p. 286).

It has been reported that the vowels associated with lenis stops sound *breathy* to Western ears (Kim, 1965, p. 349), and that the fortis stops are accompanied by a vowel of 'laryngealized' quality (Abberton, 1972; Ladefoged, 1973, p. 76). These observations have been confirmed in experimental studies by Hardcastle (1973) who argues for the recognition of a glottal 'tensity' feature, and by Iverson (1983, p. 198) who identifies the presence of 'murmur' in lenis consonants, and especially in /s/ which "correlates well with various reports of its amorphous 'breathy' quality (Kim-Renaud, 1974, p. 14, 16)." In examining the vowels themselves, vowels in the lenis environment are indeed distinct from vowels in the aspirated environment in most cases. It is the aspirated condition, however, which produces vowels with Lx characteristics of *breathiness* (or greater *whisperiness*) rather than the lenis condition. Only in the case of the open vowel /ɑ/ are Lx ratios higher after lenis consonants (similar to the *breathy* model) than after aspirated consonants. Vowels following fortis consonants are clearly distinct from the other two conditions for all vowels; lower in Lx ratio, which would correspond with either increased *creakiness* or *harshness* of phonation. As vowels in the lenis environment, independent of their articulatory identity, are known to reveal the manner of articulation of a preceding consonant (Esling, 1988, p. 7), the most likely interpretation to be drawn at this stage in the research on Korean is that the phonatory cue present in post-lenis-consonant vowels such as /o/, /u/ or /i/ is not *breathy voice,* as it is in the case of /ɑ/, but rather *whispery creaky voice* or *harsh whispery voice.*

Abberton's (1972) laryngographic study found that

> breathy voice onset is not consistently present for any stop series but does occur sometimes following the medium and most aspirated stops; in contrast with the

least aspirated at the same place of articulation for which breathy voice does not occur (p. 75).

The present results are consistent with Abberton's findings in that *breathy* (or *whispery*) Lx features characterize the lenis and aspirated series but not the fortis series. It now appears that whether the lenis or aspirated series is more *breathy* is a function of the articulatory quality of the vowel involved. Dart (1987, p. 146) reports that 'Korean fortis stops [are] generally characterized aerodynamically by higher oral pressure and lower oral flow than their lenis counterparts,' and postulates that tenser vocal tract walls are responsible for this. The present study adds to this body of knowledge the finding that the 'tensity' of the fortis series has a distinct laryngeal component. Independently of the activity of the supralaryngeal vocal tract, the phonatory quality of post-fortis-consonant vowels is clearly distinct from the other two series and well within the expected range for *creaky voice* or *harsh(/ventricular) voice*. As the vowels of this series are consistently higher in pitch than in the other two series (the lenis environment having the lowest pitch), *harshness* would seem to be a more likely candidate for describing fortis CV sequences. Details of Lx-waveform shape beyond the computed Lx ratio have not been analyzed further here, although the 'double-peak' phenomenon reported in Esling (1984) could be a useful way of describing differences in phonation that are associated with degrees of *harshness* as distinct from *creakiness*.

3.2. Theoretical Implications

In second-language-acquisition theory, the argument has been advanced that CV sequences may be critical in the process of acquiring accent (Tarone, 1978). The more traditional view gives more importance to the acquisition of individual consonants or vowels as feature bundles minimally distinct from other members of the inventory. This is a view implicit in phoneme theory but which, stated in this extreme way, leaves many questions unanswered in second-language-acquisition research. With respect to Korean, it seems clear that either the consonant or the vowel of the CV sequence can provide adequate information for correct identification. That is, although place of articulation cannot be perceived very reliably from vowel quality alone, manner of consonantal articulation can be perceived quite accurately by vowel quality alone, most distinctly for the fortis series, then for the lenis series, then (less reliably) for the aspirated series. Results from Esling (1988) further indicate that while aspiration and 'tensity' may be significant consonant indicators, information present in a vowel from a lenis context overrides these indicators in identifying a CV sequence. Together with the finding that phonatory cues differ for the aspirated and lenis series depending on the identity of the particular vowel, these results imply that the consonant of the sequence cannot be considered separately from its vowel, as a perceptual entity or, especially, in explaining acquisitional processes. On purely vocalic criteria (in this case the specification of Lx), several categories, including /æ/ and /ə/ following *p-*, *t-*, *k-*, *tʃ-* and three vowels following *s-*, present evidence that in some cases there may be no clear distinction whatsoever between lenis and aspirated consonant onsets.

These findings suggest that properties of isolated consonants may not be the critical elements in successful recognition (and perhaps production) of the syllable-long or word-long items that convey lexical meaning. Such a position reinforces second-language-acquisition studies (see Tarone, 1978) which identify the CV sequence as a minimum critical unit of choice by learners, and which focus on the importance of word identification rather than phoneme identification as the basis for building a phonological system in a second language. Such evidence suggests that it is unwise to concentrate on consonant details in second-language pronunciation teaching when it may be the vowel of the syllable that carries a large portion of the cues that learners rely on to distinguish meaning.

NOTES

This study was supported by research grant funding made available by the Social Sciences and Humanities Research Council of Canada, and by the University of Victoria. Analysis software was developed and modified with the cooperation of Craig Dickson, Roy Snell, Steve Eady, Allan Wynrib and Jim Woolsey of Speech Technology Research, Ltd., in Victoria.

In technological terms, this research represents a step in the development of automatic laryngographic analysis software for microcomputer facilities equipped to process acoustic signals. The analysis algorithm described here has been designed as an *EDIT320* program, to be used in conjunction with *Computerized Speech Lab* speech-analysis software developed by Speech Technology Research, Ltd., available as the *CSL 4300* laboratory research system from Kay Elemetrics Corporation.

REFERENCES

Abberton, Evelyn. (1972). Some laryngographic data for Korean stops. *Journal of the International Phonetic Association, 2*, 67-78.

Dart, Sarah N. (1987). An aerodynamic study of Korean stop consonants: Measurements and modeling. *Journal of the Acoustical Society of America, 81*, 138-147.

Dickson, B. Craig, Clayards, J., Wynrib, A., & Snell, R. (1990). *User's guide to the CSL program.* Victoria: Speech Technology Research, Ltd.; and Pine Brook, NJ: Kay Elemetrics Corporation.

Esling, John H. (1984). Laryngographic study of phonation type and laryngeal configuration. *Journal of the International Phonetic Association, 14*, 56-73.

Esling, John H. (1987). Teaching phonetics using the Phonetic Data Base on microcomputer. *Proceedings of the XIth International Congress of Phonetic Sciences*, vol. 4, pp. 243-246. Tallinn: Academy of Sciences of the Estonian SSR.

Esling, John H. (1988). Phonetic analysis of Korean obstruents. *Working Papers of the Linguistics Circle, 7*(1), 1-11. Victoria: University of Victoria.

Fourcin, A.J. (1974). Laryngographic examination of vocal fold vibration. In B. Wyke (Ed.), *Ventilatory and phonatory control systems* (pp. 315-333). London: Oxford University Press.

Fujimura, O. (1977). Control of the larynx in speech. *Phonetica, 34*, 280-288.

Hardcastle, W.J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics, 1*, 263-272.

Harmegnies, B., & Landercy, A. (1986). Comparison of spectral similarity indices for speaker recognition. *Proceedings of the 12th International Congress on Acoustics* (Toronto), vol. 1, A1-4.

Hirose, H., Lee, C.Y., & Ushijima, T. (1974). Laryngeal control in Korean stop production. *Journal of Phonetics, 2*, 145-152.

Iverson, Gregory K. (1983). Korean *s*. *Journal of Phonetics, 11*, 191-200.

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. *Journal of Phonetics, 2*, 161-180.

Kim, Chin-W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). *Word, 21*, 339-359.

Kim-Renaud, Y. (1974). *Korean consonantal phonology*. Ph.D. dissertation, University of Hawaii.

Ladefoged, Peter. (1973). The features of the larynx. *Journal of Phonetics, 1*, 73-83.

Tarone, Elaine. (1978). The phonology of interlanguage. In J.C. Richards (Ed.), *Understanding second and foreign language learning* (pp. 15-32). Rowley, MA: Newbury House.

Fig. 1. *CSL* display of Lx signal: partial text with *breathy voice*.
Top screen (A): original Lx waveform.
Bottom screen (B): normalized, preemphasized signal.

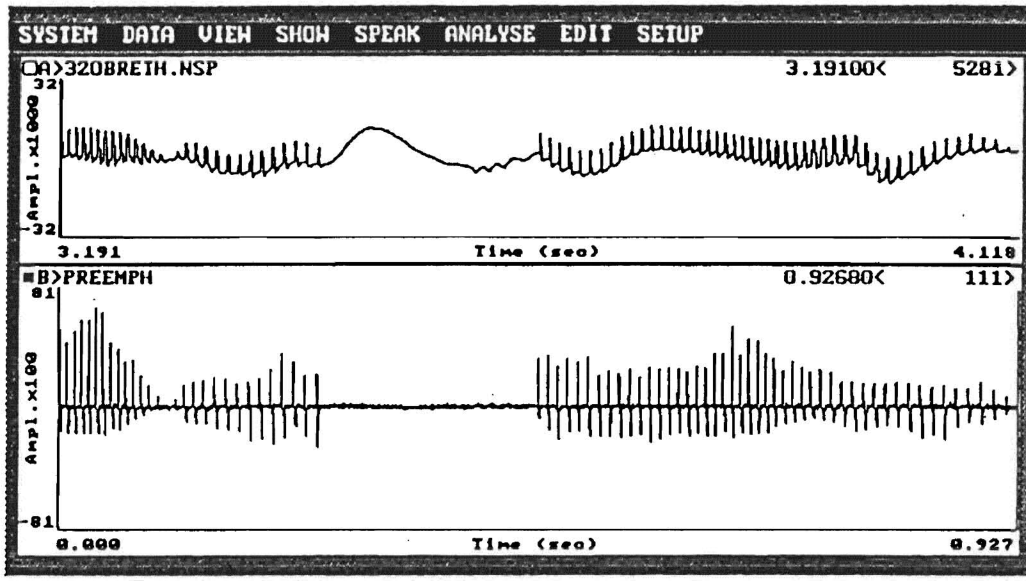

Fig. 2. Lx ratio ranges and frequency ranges of six model phonation types at 65% confidence. Ranges intersect at the mean Lx ratio and midpoint of the estimated frequency range. Median Lx ratios are represented by an open square.

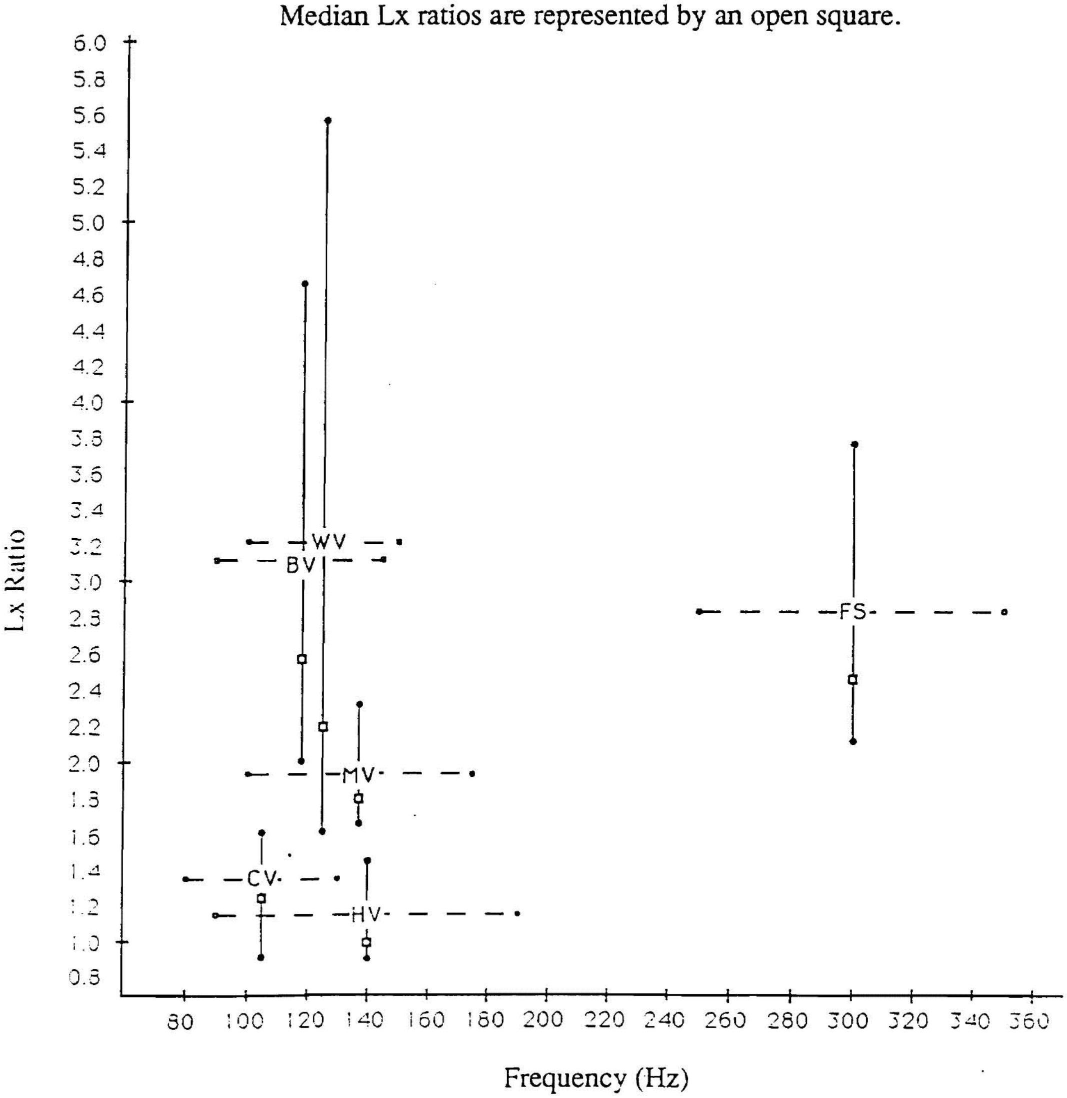

Fig. 3. Mean Lx ratios, differenced waveforms, by vowel category.

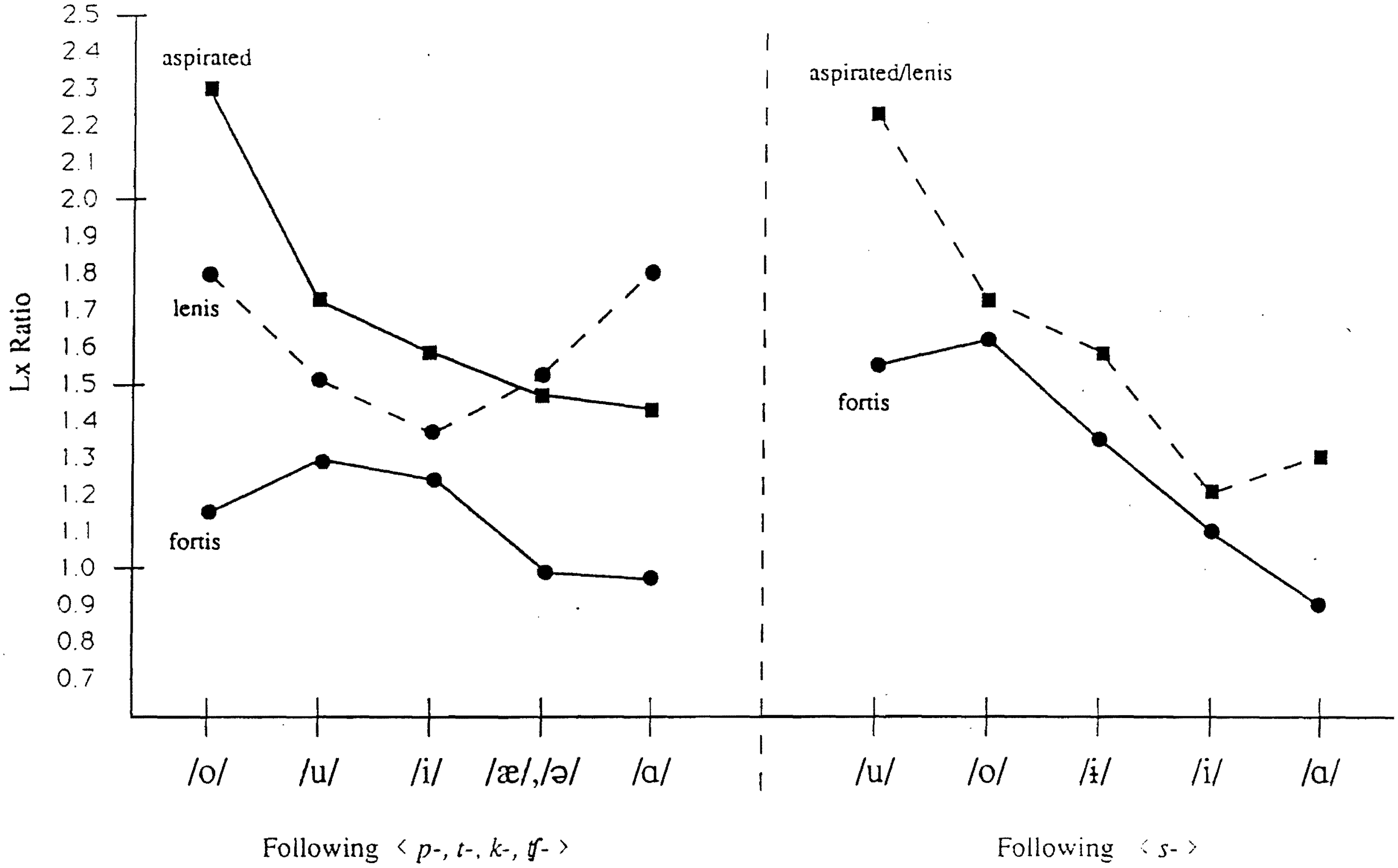

Table 1. Mean and median Lx ratios and ranges for six model phonation types at 95% (bold) and 65% confidence levels.

Phonation Type		Minimum	Median	Mean	Maximum
Breathy Voice	95%	1.50	3.03	3.41	9.15
	65%	2.04	2.57	3.17	4.66
Whispery Voice	95%	1.34	3.00	3.59	10.05
	65%	1.62	2.19	3.23	5.57
Falsetto	95%	1.69	2.78	2.96	6.25
	65%	2.16	2.47	2.84	3.75
Modal Voice	95%	1.35	1.91	2.03	4.16
	65%	1.67	1.80	1.93	2.32
Creaky Voice	95%	0.21	1.34	1.33	2.82
	65%	0.91	1.25	1.35	1.61
Harsh/Ventricular Voice	95%	0.77	1.13	1.20	2.87
	65%	0.90	1.00	1.15	1.46

Table 2. Mean and median Lx ratios for all vowels following Korean fortis, lenis and aspirated consonants.

Consonantal Environment	Tokens	n Periods	Mean Lx	Median Lx
Fortis /p^{+}, t^{+}, k^{+}, ʧ$^{+}$/	15	343	1.16	1.08
Fortis /s^{+}/	6	131	1.31	1.24
Lenis /p^{-}, t^{-}, k^{-}, ʧ$^{-}$/	16	259	1.61	1.53
Lenis/Aspirated /s/	6	91	1.60	1.49
Aspirated /p^{h}, t^{h}, k^{h}, ʧ^{h}/	17	266	1.60	1.51