PHONETIC ANALYSIS OF KOREAN OBSTRUENTS

John H. Esling

Department of Linguistics
University of Victoria

1. INTRODUCTION

Korean obstruent consonants are examined to identify the prosodic relationship between aspirated, lenis and fortis manners of consonantal articulation and adjacent vowels at the syllable level. The hypothesis that consonantal manner of articulation can be identified by vowel quality alone is tested in a series of perceptual experiments which examine the relative contribution of vowel quality and timing differences to the perception of consonants. Micro Speech Lab and accompanying software are used to process data on the IBM-XT/AT and construct perceptual tasks. A concatenation and editing program (MSLEDIT) developed in the Centre for Speech Technology Research permits random-access auditory and visual waveform comparison of phonetic data as well as recombination of selected sections from up to five separate input files for auditory presentation in psycholinguistic listening experiments. Results support the hypothesis that lenis and fortis consonants can be identified by vowel quality cues alone, suggesting that laryngographic analyses of vocalic phonatory quality should be included in statistical procedures to assess the relative contribution of timing of aspiration and phonation type in the identification of CV syllable sequences.

2. LABORATORY METHODS AND PROCEDURES

To allow audio and visual waveform comparison of phonological contrasts within the Korean sound system, and construction and presentation of perceptual listening tasks, development of the data manipulation routine, MSLEDIT, was undertaken on the IBM-PC/XT/AT-based Micro Speech Lab speech processing system in the Centre for Speech Technology Research at the University of Victoria. Micro Speech Lab (MSL) contains a program diskette, internally mounted data acquisition hardware including anti-aliasing filters, A/D and D/A circuitry, and a user's manual with instructions and descriptions of theory of use and applications (Dickson 1985). Software allows user control of signal input, waveform displays, audio output, analysis (amplitude, pitch, spectrum) and file management.

MSLEDIT is a program written to supplement MSL's speech-capturing, storage and processing capabilities, which allows phonetic researchers to access and display graphic waveforms of sampled data files, listen to words or several-second samples of text in any language selected, vary listening sequences, edit existing files, and combine elements of old files into new files. "Designed as a supplementary package to accompany the Micro Speech Lab, the purpose of the program is to provide a highly flexible method for auditory examination and manipulation of digitally stored signals" (Dickson 1987). Up to five sampled data (speech) files can be displayed and monitored individually, in reverse, or in continuous repetition of sequences composed of parts of any displayed file.

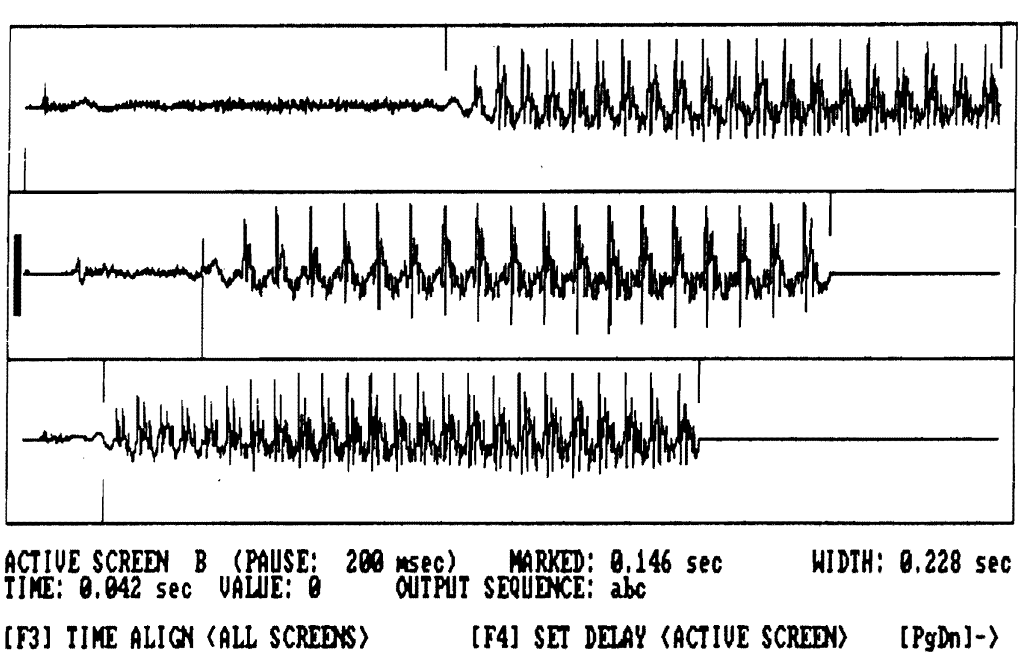
Data for acoustic analysis in Korean are provided from a Phonetic Data Base (PDB) of linguistically organized speech samples assembled using MSL. Words and text samples in the PDB are drawn from numerous linguistic, sociolinguistic and dialect survey sources to reflect a wide range of speech sounds of languages of the world. Samples are digitally encoded using the MSL capturing routine and stored by language on diskette or hard disk and documented on paper by number for reference to phonetic, phonemic and orthographic representations, and English gloss. Phonetic sounds that are normally difficult to obtain, and phonemic inventories of a range of languages not usually encountered, are available in the PDB, including: Egyptian Arabic, Cantonese, Modern Standard Chinese, Inuktitut, Japanese, Korean, Miriam, Nitinaht, Nyangumarta, Rutooro, Runyoro, Scots Gaelic, Skagit (Lushootseed), Spokane, Turkish, Umpila, Xhosa, and Yoruba.

3. ACOUSTIC ANALYSIS

The minimal contrast for dental stops in Korean is illustrated in the MSLEDIT display in figure 1. The aspirated stop demonstrates longest voice onset time (VOT), 0.095sec; the lenis stop is actually partially aspirated with 0.032sec VOT; and the fortis stop demonstrates shortest VOT, 0.015sec, with virtually no aspiration present. All three screens, A B C, are time-aligned to the longest file (A, 0.228sec), and split cursors mark the onset of voicing in each screen. The active cursor is set to screen B where the time read-out is 0.042sec at the onset of voicing, with value 0 displayed as the waveform crosses the baseline, 0.032sec after the initial consonant burst. The output sequence specified in lower case, a b c, represents the marked portions of each screen (between cursors), extracted for display purposes at this stage from original full-length files. A menu of function key operations scrolls across the bottom of the display.

Figure 1. MSLEDIT display of Korean CV sequences: Dental stops.

A: /t^hɑ/ B: /t^-ɑ/ C: /t^+ɑ/



These observations parallel and confirm 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. Electro-myographic 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: 286).

Figures 2 and 3 illustrate the extent to which phonatory mechanisms can be seen in the MSL display to differ in their effect on the acoustic waveform. In the three words, a minimal triplet, the strong aspiration of the first example carries into the vowel which remains voiceless. The second example shows reduced intensity of glottal friction, which is much shorter and more intense in the third example, while both have fully voiced vowels. Marking only the first CV sequence of each word (in figure 3) highlights these differences, as well as the contrast in phonatory quality that marks each postconsonantal vowel. Note that the vowel in A is voiceless (whispered).

Figure 2. MSLEDIT display: Initial bilabial stops.

A: /p^hitɑ/ 'bloom' B: /p^-itɑ/ 'be empty' C: /p^+itɑ/ 'be dislocated'

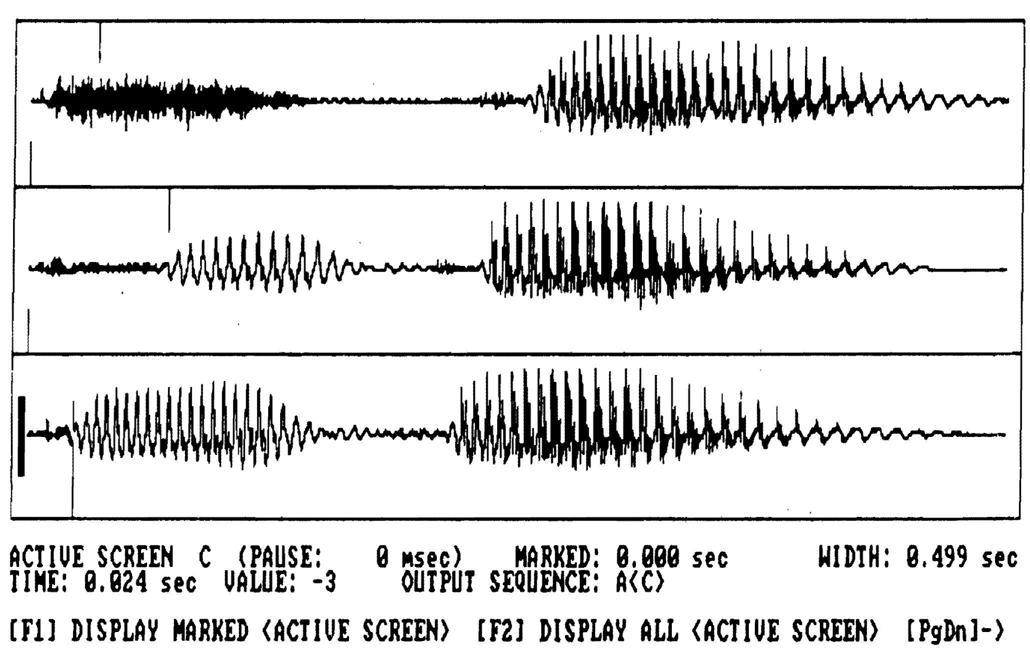


Figure 3. Initial bilabial stops: Marked CV contrasts.

A: /p^hi/ B: /p^-i/ C: /p^+i/

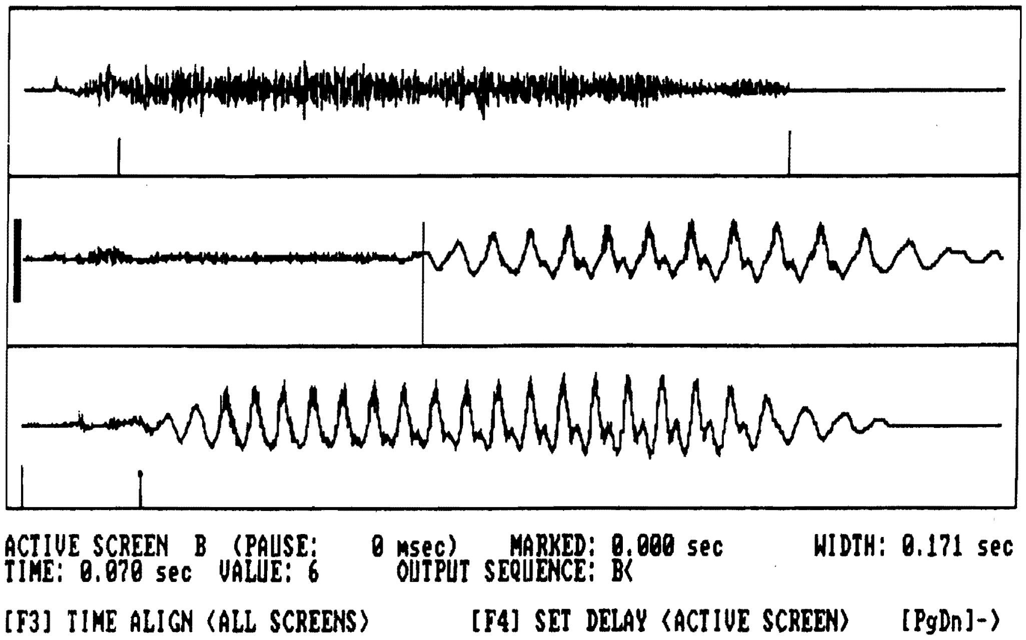


Figure 4. MSLEDIT display: Initial bilabial stops.

A: /p^hul/ 'grass' B: /p^-ul/ 'fire' C: /p^+ul/ 'horn'

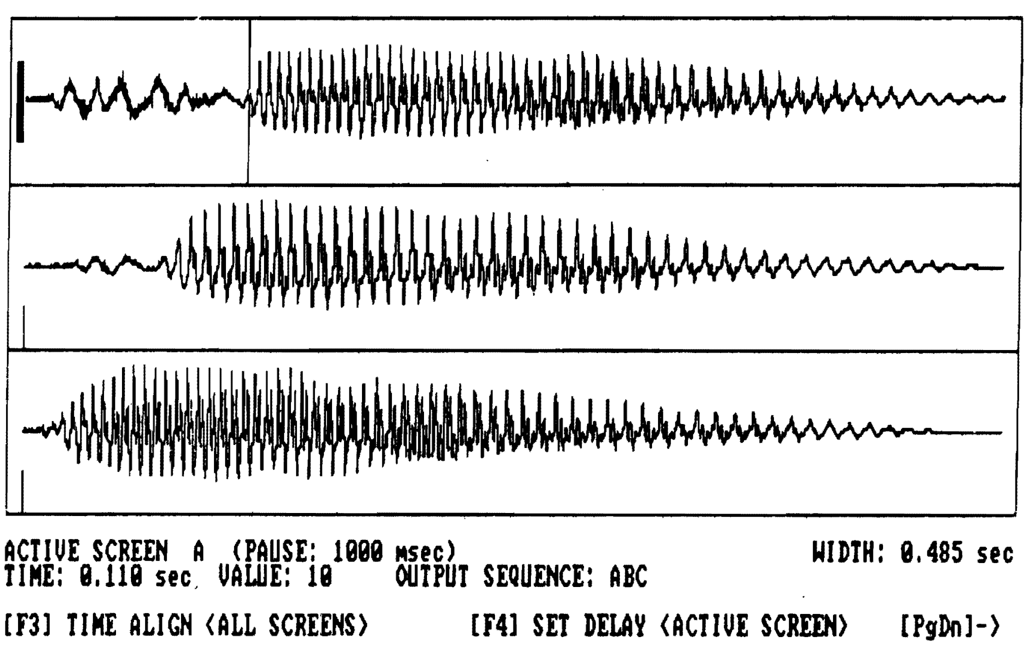
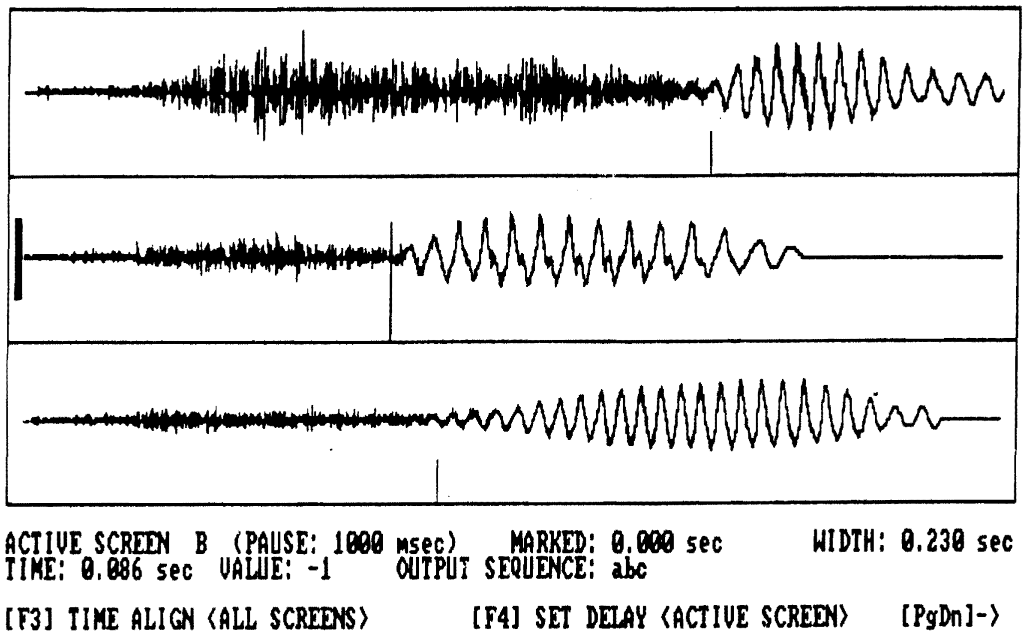


Figure 5. Affricate contrasts: Initial CV sequences.

A: $/\check{c}^{h}i/$ B: $/\check{c}^{-}i/$ C: $/\check{c}^{+}i/$



Timing differences are particularly clear preceding an /u/-quality vowel as in figure 4. These distinctions begin to be obscured, however, with the affricate [č] in figure 5, especially where friction is prolonged in the fortis example. Nevertheless, the qualities of the three postconsonantal vowels appear to remain consistent and distinctive. It has been reported that the lenis stops are often heard to be breathy (Kim 1965: 349), and that the fortis stops are accompanied by a vowel of laryngealized quality (Ladefoged 1973: 76). These observations have been confirmed in experimental studies by Hardcastle (1973) who argues convincingly for the recognition of a glottal "tensity" feature, and by Iverson (1983: 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: 14,16)". It has not been documented experimentally, however, that vowels themselves which accompany lenis or fortis consonants differ systematically in phonation type. Because of this possibility, and the implications it would have for second language acquisition theory, a series of perceptual tests was organized to evaluate the role of vowel quality in perceiving Korean consonant contrasts.

4. PERCEPTUAL DISCRIMINATION TESTING

In second language acquisition theory, the argument can be made that CV sequences are critical in the process of acquiring accent (Tarone 1978). A 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 view is implicit in phoneme theory, but leaves many questions unanswered in second language acquisition research. The assumption followed here is that syllables are not necessarily divided into constituent segments for the acquisition process to operate. The questions examined with respect to Korean are whether consonantal place of articulation can be perceived from vowel quality alone; whether manner of articulation can be perceived by vowel quality alone; and whether CV sequences are identified primarily on the basis of vowel quality or of consonantal features such as differences in timing and aspiration.

Test design and presentation are facilitated by MSL companion software, RANDOM and MSLAUDIO. RANDOM, written for this project by CSTR, arranges a specified set of sampled speech data files into random order and assigns new, numbered filenames in that order. For each perception test, a new randomized order of presentation is created. MSLAUDIO allows a command file to be written by the researcher to load and speak specified files automatically in a listening-test battery (in randomized order in each case here), with pre-set delay times and number of repetitions.

In tests 1 and 2, only a vowel is heard, extracted using MSLEDIT as delimited by the marked cursor positions in figure 1, for example, from the complete set of Korean test words in the Phonetic Data Base. The subject is a Korean graduate research assistant, who is also the original speaker in the PDB recordings. A second subject, the speaker's wife, a non-linguist and non-English speaker, was also tested subsequently. Initial tests are therefore commutation tests, where utterances produced by the subject are presented as data, assuming a criterion of familiarity, in a listening discrimination task by self or spouse (Labov 1972). The task is to identify the consonant which precedes the vowel. Each item is repeated continually on high quality external speakers until the subject chooses to continue by pressing any key, but items are not returned to. The tests begin with a practice run of three items for familiarization with the equipment. The first test is a battery of 39 vowels where place of articulation is not known. In the second test, the 39 vowels are separated into their five respective categories for presentation: labial stops, dental stops, velar stops, palato-alveolar affricates, and dental fricatives.

The results of test 1 indicate that information present in the vowels is not alone adequate to identify place of articulation of the consonant, or whether it is a stop, affricate or fricative. Subject 1 scored 26% and subject 2 only 23% correct in identifying consonants from vowel quality alone. Consonant transition information is usually a strong indicator of consonant identity, suggesting that little transition data is present in these examples. Nevertheless, identification of secondary manner of articulation (aspirated, lenis, or fortis) relying on vowel quality alone is very high, 59% and 67% for the two subjects respectively. Applications of chi-squared with 2 d.f. are significant for both tests at the $p<0.05$ level, i.e., there is significant evidence for an association between subjects' (correct) selections and the three secondary manners of articulation, even where the consonants themselves could not be identified. It appears from these preliminary data that items with fortis articulation (presence of the "tensity" feature) are easiest to identify (100%, 100%), followed by lenis items (64%, 71%), while aspirated items are virtually impossible to identify correctly (9%, 18%) in the absence of consonant timing information. It is worth pointing out that aspirated consonants also appear twice as difficult to identify correctly for place of articulation as are fortis consonants when there is only vocalic resonance and no timing evidence present. A tentative interpretation is that closure and friction information prior to voicing are important in identifying aspirated consonants, while phonatory quality during vocalic voicing retains identifying clues of lenis and fortis consonants.

Tests 3 and 4 assess the auditory recognition of CV sequences, including 39 original items and 72 manipulated sequences consisting of all possible permutations of C+V made up from the original set. These "artificial" sequences are easy to obtain using MSLEDIT marking, sequence-combining, and file-saving capabilities. Test 3 contains all 111 items, and test 4 includes only the 72 re-ordered items. Each test is preceded by a 10-item practice set. Instructions are to indicate the consonant-vowel sequence heard, with all possible Cs and Vs listed on the response sheet. For subject 1, choices and responses are in phonemic transcription, while for subject 2, all instructions, choices and responses are in Korean orthography. Auditory stimuli are presented only twice in the CV tests, with a brief interval before the next token is heard. Items are loaded in 10-file batches, creating regular pauses in the performance of the task. After the practice run, subjects demonstrated no difficulty in making a choice on the basis of two hearings, whereas in tests 1 and 2, unlimited listening had caused some equivocation.

Results of test 3 can be divided into two parts: identification of original CV sequences, and identification of re-ordered CVs. Both subjects are able to recognize original items correctly (92% and 90%, respectively). The small margin of error is a measure of the interfering effect of the equipment and testing situation, and that the CVs are not complete morphemic units. Re-ordered CVs are scored according to whether auditory identification matches the consonant (C) element or the vowel (V) element from which they were constructed. A third category is scored (N) if identification corresponds to neither element in the constructed CV sequence. In general, recognition is divided fairly equally between C and V cues: C, 43%, 36%; V, 43%, 47%; N, 14%, 17%, for the two subjects. There is a pattern, however, which differentiates these assignments systematically. C assignments predominate for C(aspirated)+V(fortis) and C(fortis)+V(aspirated) constructions. V assignments, on the other hand, predominate for C(aspirated)+V(lenis), C(lenis)+V(aspirated) and C(fortis)+V(lenis) combinations. For both speakers, these 6x2 distributions represent a significant association, at $p<0.05$ in a chi-squared test with 5 d.f., between the constructed CV test sequences and their auditory identification. C(lenis)+V(fortis) combinations were predominantly identified as aspirated (neither). These observations can be summarized as follows: (1) "Lenis" vowels are perceived as the dominant element in a sequence; (2) aspirated consonants are perceived as the dominant element in a sequence except when the vowel is "lenis"; (3) fortis consonants are perceived as the dominant element in a sequence except when the vowel is "lenis"; (4) "aspirated" vowels and "fortis" vowels are both perceived as being preceded by aspirated consonants when the consonant is lenis.

These findings are supported in test 4, where only the constructed CVs are presented in a different order. The interpretation is that differences of consonant timing and aspiration are significant where vowels have been extracted from aspirated or fortis contexts. Vowels from lenis contexts, on the other hand, signal a lenis consonant whatever consonant was actually heard. There is little difference in these tests between a vowel from an aspirated or a fortis context. If aspiration is long, C is heard as aspirated; and if C is short and tense, it is heard as fortis. Even a semi-aspirated lenis C is heard as aspirated when followed by a vowel from an aspirated or a fortis context. Vowels from a lenis context, however, are readily identified and change the perception of the consonant that precedes them.

This supports the contention that while aspiration and "tensity" are significant consonant indicators, information present in a vowel from a lenis context overrides these indicators in identifying a CV sequence. Laryngealization has been reported by Ladefoged (1973) and by Abberton (1972) as a vocalic prosody of fortis stops, but neither the impression of breathiness ascribed by Kim (1965) to lenis (medium-aspirated) stops nor its prosodic nature have been demonstrated experimentally. In Abberton's laryngographic study,

> 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 (1972: 75).

It is hypothesized here that a distinctive phonatory quality ("breathy" in auditory terms) is associated prosodically with lenis consonants as a principal property of vowels in such CV sequences, and acts as a principal cue in identifying meaning in those sequences. This quality remains distinct from phonation found in aspirated sequences which have been shown in the perceptual tests to be indistinct from their fortis vocalic equivalents. This proposal conforms with recent evidence that "Korean fortis stops [are] generally characterized aerodynamically by higher oral pressure and lower oral flow than their lenis counterparts" (Dart 1987: 146), where tenser vocal tract walls are postulated to account for the tensity observed for fortis stops. It is proposed in the case of lenis consonants to examine in detail, using updated and revised laryngographic techniques (see Esling 1984), the suprasegmental characteristics of phonation type that carry the prosodic cues to consonant identification and, potentially, lexical item recognition.

5. RESEARCH RESULTS

The analysis of Korean aspirated, lenis, and fortis obstruent contrasts has had several encouraging results both theoretically and in practical terms.

1. The Micro Speech Lab system has been shown to illustrate clearly the acoustic waveform contrasts between the three types of articulation. Data access is rapid, and by random selection.

2. Timing differences are confirmed for the three-way contrast, as is the glottalized "tensity" feature of fortis items reported in previous research.

3. Investigation of vowel quality differences, suggested but not addressed explicitly in previous research, reveals consistent low pitch and breathy phonation for vowels following lenis consonants. Fortis consonants are followed by a tense, higher-pitched vowel.

4. Development of MSLEDIT software provides an excellent research format for examining phonetic detail both visually and auditorily (see Dickson 1985, 1987). The recombination capability permits splicing and arranging of data that could only be achieved before with separated linear recordings, tape loops, and laborious splicing of magnetic tape.

5. The system's efficiency has allowed rapid and easy construction of perceptual auditory testing tasks using MSLAUDIO, and the additional development of a program, RANDOM, to automatically randomize speech data files for presentation in listening tasks. Sets of 39-, 72- and 111-file lists have been prepared and presented with minimal effort.

6. Isolated and randomized single vowels presented auditorily to Korean listeners for identification reveal that recognition of place of articulation of initial consonants is very poor, while recognition of manner of articulation is very high. This suggests that vocalic information alone is sufficient to indicate the secondary articulation of a consonant even when that consonant cannot be identified.

7. Presentation of CV syllable sequences, spliced together using MSLEDIT in all C+V permutations, results in a systematic pattern of identifications which confirms (a) that vowel quality (presumably breathiness) and/or low pitch is the indicator of a lenis consonant and (b) that duration of aspiration relative to length of vowel in a syllable determines whether a consonant is perceived as aspirated. Tests have been verified with a second Korean listener, and rating sheets have been prepared in Korean to be used with a wider sample of subjects.

8. Findings suggest that properties of isolated consonants alone 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 recent second language acquisition work which identifies the CV sequence as a critical unit of choice by learners, and which focuses on the importance of word identification rather than phoneme identification as the basis for building a phonological system in an L2. Such evidence suggests that it is unwise to concentrate on consonantal details in L2 pronunciation teaching when it may be the vowel of the syllable that carries a large portion of the clues that learners rely on to distinguish meaning.

6. FURTHER RESEARCH

The development of microcomputer-based speech processing software in Micro Speech Lab format, including MSLEDIT, MSLAUDIO, and MSLSPECT, and continued updating and improvement of the Phonetic Data Base, facilitates the collection, presentation and comparison of speech sound material for phonetic research. The system permits expedient access to large amounts of diverse data, active manipulation and organization of speech items, and rapid presentation of data for aural discrimination testing. The system also contributes significantly to phonetics instruction (see Esling 1987) and to the training of teachers in the use of technological aids for the delivery of speech sound information. Additional applications include the computer-network transmission and sharing of speech data for collaboration in phonetic research.

In the further study of Korean obstruents, electrical impedance laryngographic analysis (Fourcin 1974; Esling 1984) will be combined with Micro Speech Lab procedures to evaluate the "breathy" and "tense" phonatory features of vowel quality indicated in present research, and to differentiate vowel quality from the effect of pitch. Current results illustrate and strengthen the possibility of developing speech analysis packages around the Phonetic Data Base theme. Research is also being carried out to apply Micro Speech Lab technology to the comparison, analysis and on-screen phonetic transcription of glottalized consonants of Salish and Wakashan languages stored in the PDB. This research potential is made possible through a cooperative effort involving resources in the Department of Linguistics and the technological expertise of the Centre for Speech Technology Research at the University of Victoria.

NOTES

I am indebted to my co-researchers, Geoff O'Grady, Thom Hess, Barry Carlson, and numerous colleagues and graduate students who are the source of the majority of linguistic material in the Phonetic Data Base. Particular appreciation is extended to Yong-Bum Kim and to Seung-Hee Kim for their cooperation in this project. The indispensable assistance and collaboration of Craig Dickson and Roy Snell of the Centre for Speech Technology Research at the University of Victoria in conceiving and developing microcomputer-based speech processing systems is gratefully acknowledged, as are general university research grants from the Social Sciences and Humanities Research Council of Canada.

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. (1985). *User's Manual for Micro Speech Lab.* Victoria, B.C.: Software Research Corporation.

Dickson, B. Craig. (1987). *User's Manual for the MSL Comparison and Editing Program: MSLEDIT*. Victoria, B.C.: Centre for Speech Technology Research Society.

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. Pp. 298-301 in *Proceedings of the XIth International Congress of Phonetic Sciences.* Vol. 5. Tallinn: Academy of Sciences of the Estonian SSR.

Fourcin, A.J. (1974). Laryngographic Examination of Vocal Fold Vibration. Pp. 315-333 in *Ventilatory and Phonatory Control Systems.* B. Wyke, ed. 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.

Hirose, H., C.Y. Lee, and T. Ushijima. (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.

Labov, William. (1972). *Sociolinguistic patterns*. Philadelphia: The University of Pennsylvania Press.

Ladefoged, Peter. (1973). The Features of the Larynx. *Journal of Phonetics* 1: 73-83.

Tarone, Elaine. (1978). The Phonology of Interlanguage. Pp. 15-32 in *Understanding Second and Foreign Language Learning*. J.C. Richards, ed. Rowley, Mass.: Newbury House Publishers, Inc.