
The Ohio State University

Working Papers in Linguitics No. 43

Papers from the Linguistics Laboratory

Edited by

Sook-hyang Lee

and

Sun-Ah Jun

The Ohio State University

Department of Linguistics

222 Oxley Hall
1712 Neil Avenue
Columbus, Ohio 43210 USA
lingadm@ling.ohio-state.edu

January 1994

Information Concerning the OSUWPL

The Working Papers in Linguistics is an occasional publication of the Department of Linguistics of the Ohio State University and usually contains articles written by students and faculty of the department. There are generally one to three issues per year. Information about available issues appears below. Numbers 1, 5, 10 and 23 are out of print and no longer available.

There are two ways to subscribe to WPL. The first is on a regular basis: the subscriber is automatically sent and billed for each issue as it appears. The second is on an issue-by-issue basis: the subscriber is notified in advance of the contents of each issue, and returns an order blank if that particular issue is desired.

Volume

18 $5.00. 183 pp. (June 1975): Papers by Michael Geis, Sheila Geoghegan, Jeanette Gundel, G.K. Pullum, and Arnold Zwicky.

19 $5.00. 214 pp. (September 1975), edited by Robert K. Herbert: *Patterns of Language, Culture and Society: Sub-Saharan Africa* contains eighteen papers presented at the Symposium on African Language, Culture, and Society, held at The Ohio State University on April 11, 1975.

20 $5.00. 298 pp. (September 1975), edited by Robert K. Herbert: *Proceedings of the Sixth Conference on African Linguistics* contains twenty-seven papers presented at the Sixth Conference on African Linguistics, held at The Ohio State University on April 12-13, 1975.

21 $5.00. 252 pp. (May 1976), edited by Arnold M. Zwicky: *Papers on Nonphonology*. Papers by Steven Boer and William Lycan, Marian Johnson, Robert Kantor, Patricia Lee, and Jay Pollack.

22 $5.00. 151 pp. (February 1977), edited by Olga Garnia: *Papers in Psycholinguistics and Sociolinguistics*. Papers by Sara Garnes, Olga Garnica, Mary Louise Edwards, Roy Major, and John Perkins.

24 $5.00. 173 pp. (March 1980), edited by Arnold M. Zwicky: *Clitics and Ellipsis*. Papers by Robert Jeffers, Nancy Levin (OSU Ph.D. Dissertation), and Arnold Zwicky.

25 $5.00. 173 pp. (January 1981), edited by Arnold M. Zwicky: *Papers in Phonology*. Papers by Donald Churma, Roderick Goman (OSU Ph.D. Dissertation), and Lawrence Schourup.

26 $5.00. 133 pp. (May 1982), edited by Brian D. Joseph: *Grammatical Relations and Relational Grammar*. Papers by David Dowty, Catherine Jolley, Brian Joseph, John Nerbonne, and Amy Zaharlick.

27 $5.00. 164 pp. (May 1983), edited by Gregory T. Stump: *Papers in Historical Linguistics*. Papers by Donald Churma, G.M. Green, Leena Hazelkorn, Gregory Stump, and Rex Wallace.

28 $5.00. 119 pp. (May 1983). Lawrence Clifford Schourup, *Common Discourse Particles in English Conversation*, OSU Ph.D. Dissertation.

29 $5.00. 207 pp. (May 1984), edited by Arnold Zwicky and Rex Wallace: *Papers on Morphology*. Papers by Belinda Brodie, Donald Churma, Erhard Hinricks, Brian Joseph, Joel Nevis, Anne Steward, Rex Wallace, and Arnold Zwicky.

30 $5.00. 203 pp. (July 1984). John A. Nerbonne, *German Temporal Semantics: Three Dimensional Tense Logic and a GPSG Fragment*, OSU Ph.D. Dissertation.

31 $6.00. 194 pp. (July 1985), edited by Michael Geis: *Studies in Generalized Phrase Structure Grammar*. Papers by Belinda Brodie, Annette Bissantz, Erhard Hinricks, Michael Geis and Arnold Zwicky.

32 $6.00. 162 pp. (July 1986). *Interfaces*. 14 articles by Arnold M. Zwicky concerning the interfaces between various components of grammar.

33 $6.00. 159 pp. (August 1986). Joel A. Nevis, *Finnish Particle Clitics and General Clitic Theory*, OSU Ph.D. Dissertation.

34 $6.00. 164 pp. (December 1986), edited by Brian Joseph: *Studies on Language Change*. Papers by Riita Blum, Mary Clark, Richard Janda, Keith Johnson, Christopher Kupec, Brian Joseph, Gina Lee, Ann Miller, Joel Nevis, and Debra Stollenwerk.

35 $10.00. 214 pp. (May 1987), edited by Brian Joseph and Arnold M. Zwicky: *A Festschrift for Ilse Lehiste*. Papers by colleagues of Ilse Lehiste at The Ohio State University.

36 $10.00. 140 pp. (September 1987), edited by Mary Beckman and Gina Lee: *Papers from the Linguistics Laboratory 1985-1987*. Papers by Keith Johnson, Shiro Kori, Christiane Laeufer, Gina Lee, Ann Miller, and Riita Valimaa-Blum.

37 $10.00. 114 pp. (August 1989), edited by Joyce Powers, Uma Subramanian, and Arnold M. Zwicky: *Papers in Morphology and Syntax*. Papers by David Dowty, Bradley Getz, In-hee Jo, Brian Joseph, Yongkyoon No, Joyce Powers, and Arnold Zwicky.

38 $10.00. 140 pp. (July 1990), edited by Gina Lee and Wayne Cowart: *Papers from the Linguistics Laboratory*. Papers by James Beale, Wayne Cowart, Kenneth deJong, Lutfi Hussein, Sun-Ah Jun, Sook-hyang Lee, Brian McAdams, and Barbara Scholz.

39 $15.00. 366 pp. (December 1990), edited by Brian D. Joseph and Arnold M. Zwicky: *When Verbs Collide: Papers from the 1990 Ohio State Mini-Conference on Serial Verbs* contains eighteen papers presented at the conference held at The Ohio State University, May 26-27, 1990.

40 $15.00. 440 pp. (July 1992), edited by Chris Barker and David Dowty: *Proceedings of the Second Conference on Semantics and Linguistic Theory* contains twenty papers from the conference held at The Ohio State University, May 1-3, 1992.

41 $12.00. 148 pp. (November 1992), edited by Elizabeth Hume: *Papers in Phonology*. Papers by Benjamin Ao, Elizabeth Hume, Nasiombe Mutonyi, David Odden, Frederick Parkinson, and Ruth Roberts.

42 $15.00. 237 pp. (September 1993), edited by Andreas Kathol and Carl Pollard: *Papers in Syntax*. Papers by Christie Block, Michael Calcagno, Chan Chung, Qian Gao, Andreas Kathol, Ki-Suk Lee, Eun-Jung Yoo, Jae-Hak Yoon, and a bibliography of published works in and on Head-Driven Phrase Stucture Grammar

43 $12.00. 130pp. (January 1994), edited by Sook-hyang Lee and Sun-Ah Jun: *Papers from the Linguistics Laboratory*. Papers by Benjamin Ao, Islay Cowie, Monica Crabtree, Janet Fletcher, Ken de Jong, Sun-Ah Jun, Claudia Kurz, Gina Lee, Sook-hyang Lee, Ho-hsien Pan, and Eric Vatikiotis-Bateson.

The following issues are available through either: The National Technical Information Center, The U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22151 (PB), or ERIC document Reproduction Service (ED) Center for Applied Linguistics, 161 N. Kent St., Arlington, VA 22209.

2 November 1968, 128 pp. (OSU-CISRC-TR-68-3). PB-182 596.

3 June 1969, 181 pp. (OSU-CISRC-TR-69-4). PB-185 855.

4 May 1970, 164 pp. (OSU-CISRC-TR-70-26). PB-192 163.

6 September 1970, 132 pp. (OSU-CISRC-TR-70-12). PB-194 829.

7 February 1971, 243 pp. (OSU-CISRC-TR-71-7). PB-198 278.

8 June 1971, 197 pp. (OSU-CISRC-TR-71-7). PB-202 724.

9 July 1971, 232 pp. (OSU-CISRC-TR-71-8). PB-204 002.

11 August 1971, 167 pp. ED 062 850.

12 June 1972, 88 pp. (OSU-CISRC-TR-72-6). PB-210 781.

13 December 1972, 255 pp. ED 077 268.

14 April 1973, 126 pp. ED (parts only)

15 April 1973, 221 pp. ED 082 566.

16 December 1973, 119 pp. ED (parts only)

Foreword

This volume includes papers that take an experimental approach to issues in linguistics. It is the third of a series of progress reports from the Linguistics Laboratory (see OSUWPL No. 36 and No. 38.) All papers were presented at meetings of the Acoustical Society of America between 1989 and 1991. We are very grateful to Eliza Segura-Holland for her help in proof-reading the English and in making the final layout. We also would like to thank Mary Beckman, Gayle Ayers, Stefanie Jannedy, Andreas Kathol and Jennifer Venditti for their physical and emotional support. Many aspects of the experiments and of the production of this volume were supported by the National Science Foundation under grants IRI-8617852 and IRI-8858109 to Mary Beckman.

Sook-hyang Lee
Sun-Ah Jun
January 1994

Ohio State University Working Papers in Linguistics No. 43

Papers from the Linguistics Laboratory

Table of Contents

Initial tones and prominence in Seoul Korean*

Ken de Jong

Department of Linguistics, The Ohio State University

Abstract

The present paper reports on an investigation of the informal observation that Seoul Korean nouns have an initial prominence and often have an initial prominence entirely due to an initial intonational prominence, or could other aspects of the signal be involved? second, are initial tones best considered pitch accents (prominence-lending tones associated with a particular syllable) or are they better seen as phrase tones (markers of the edge of a prosodic group) Two subjects recited various sentences at different rates with different levels of overall emphasis. Word initial syllables are not longer than their medial counterparts; the only durational lengthening found is the consonant initial to the phrase. F0 measurements on initial nouns of various lengths suggests that initial high tones are underlyingly always present, but are obscured in short noun phrases by a following, phrase final tonal prominence. These results suggest that Korean nouns do not have initial stress, but rather occur in prosodic phrases which are marked in part by an initial high tone.

Introduction

Temporal duration and intonational accents have been noted to act together in prosodic system to lend prominence to a particular syllable (Lehiste, 1970; Beckman, 1986). For example, in English, it has been shown that the location of word stress can be cued both by intonational and durational differences (Fry, 1958; Bolinger, 1958, 1965; Lieberman, 1960).

A similar situation might be true of Seoul Korean. Native speakers have suggested informally that there might be stress on the initial syllable of nouns; though this claim does not seem to surface in the literature. Furthermore, Lee (reported on in this volume) has also found initial intonational prominences. Figure 1 is an f0 track from this study of a sentence with a five syllable noun in subject position. The feature of interest is the initial f0 maximum, which is realized on the initial syllable of the subject noun. These f0 maxima appear in all of repetitions of this sentence.

However, if the sentence also contains an adjective previous to the noun, the peak occurs over the adjective, and no peak then appears on the first syllable of the following noun. Thus, the peak location is not a property of the word, but is more likely a property of some larger prosodic domain.

This paper deals with two related questions. First, is the perception of initial stress simply due to the intonational prominence illustrated in Figure 1? Or is there some other aspect of the speech signal which is involved in the perception of stress?

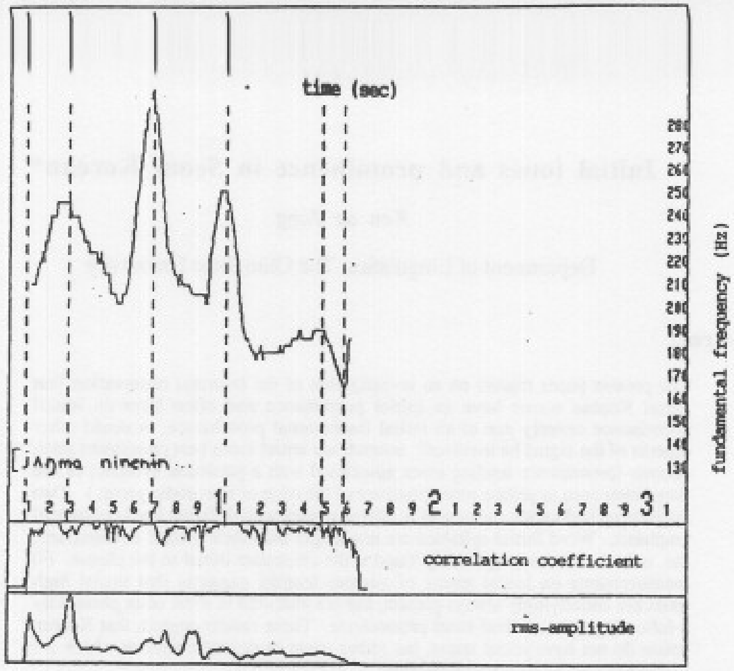


Figure 1. An f0 trace of a Seoul Korean sentence with a five syllable noun in subject position. The first two f0 maxima occur over the subject. The first occurs over the initial syllable of the subject noun, the second, more prominent peak over the subject marker, /nɨn/.

Stress in English, besides being involved with the occurrence of pitch accents, has its own correlates, duration, amplitude, and more distinct formant structures (e.g., as shown in Fry, 1958; Harris, 1978; Huss, 1978; and Summers, 1987). Second, is the pitch peak apparent in Figure 1 indicative of a prominence lending pitch accent, such as those found in English (Bolinger, 1965; Pierrehumbert, 1980)? To obtain an answer to the first question, I measured the durations of segmentally identical CV sequences placed in word medial and word initial positions. If Korean does have word initial stress, one would expect consistent durational differences between all word initial syllables and medial syllables. If, however, there is no syllabic stress, there would be no differences between word medial and word initial syllables.

A third and a fourth possibility also exist. The durational difference may be a correlate of a pitch accent. Perhaps the f0 peak is indicative of a pitch accent which has durational lengthening associated with it. Pitch accents in Swedish and English have such lengthening associated with them, (Stälhammar, et al. 1973; Nord, 1986, and Engstrand, 1988 for Swedish; Beckman, Edwards, and Fletcher, 1992, de Jong, 1991, for English).

Alternatively, durational differences may be an edge marker of some prosodic unit not isomorphic with the word, for example, an accentual or intonational phrase of the type proposed in Pierrehumbert (1980), Selkirk (1980), and Pierrehumbert and Beckman (1988). I am assuming in this scenario that the initial f0 maximum illustrated in Figure 1 also marks the edge of this domain. Thus, if either of these possibilities is correct, word initial syllables should only be longer than word medial syllables when they are phrase initial as well. When the noun is preceded by an adjective, the lengthening should occur on the first syllable of the adjective, not on the first syllable of the noun.

To obtain an answer to the second question, I must address two additional questions. First, what is the relationship between the tone and the segmental material with which it appears? Pitch accents in English are phonologically associated with a stressed syllable. Thus, they seem to be temporally aligned with the sonority peak of the stressed syllable (Silverman and Pierrehumbert, 1990). Second, can the initial pitch peak be construed as the property of a prosodic domain? Pitch specifications, called phrase tones or phrase accents, can appear as (usually peripheral) markers of a prosodic domain. English pitch accents contrast with phrase tones by not being associated with a prosodic domain. They simply appear optionally before a nuclear accent in an intonational or intermediate phrase (Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988).

With regard to this second issue, the initial intonational peaks found by Lee occurred consistently in the five syllable tokens of which Figure 1 is an example. However, as can be seen in Figure 2, the initial peak did not always appear in three syllable tokens. Thus, the pitch peak might be phonologically optional, just as English pitch accents are. In English, prenuclear accents tend to appear near the beginning of an intonational phrase, especially when the nuclear accent is placed somewhat later in the phrase. With much shorter phrases, the subject will often forgo the prenuclear accents.

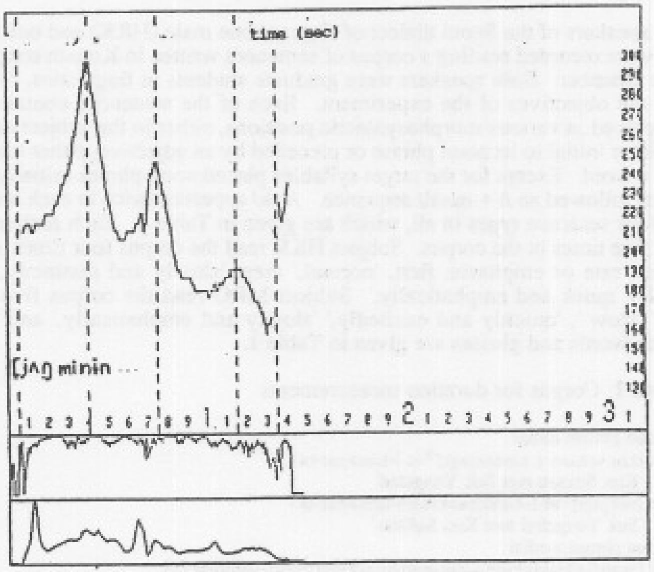


Figure 2. An f0 trace of a Seoul Korean sentence with a three syllable subject. The first maximum occurs over the subject marker, /nin/, which appears immediately following the noun.

However, there is another possible interpretation of the Korean patterns. One might claim that the initial pitch peaks are indicative of phrase tones which occur consistently at the beginning of some prosodic domain. To do so, one must say that the initial peak does exist underlyingly even in the utterance shown in Figure 2. However, it does not appear on the surface due to the proximity of the following, much higher tonal prominence on /nɨn/.

The initial tonal prominence appears clearly in the utterance shown in Figure 1, because the high tone on /nɨn/ is too far away temporally to obscure the initial peak. If this is, in fact, what is occurring here, one should find a continuum from clear f0 peaks to completely obscured peaks, with intermediate forms with partially obscured peaks. If, however, the optional pitch accent analysis is correct, there should be no transitional forms either the accent is there, or it is not.

In order to answer these questions about the status of the initial pitch peak, a series of sentences were recorded similar to those recorded for the durational analysis, except that the subject nouns varied in length. Since the utterances in Lee's corpus consistently have a major intonational prominence on the subject marker appended to the right edge of the subject noun, this corpus shows the behavior of the initial tonal maximum as the following tonal prominence becomes temporally closer. Of interest is the alignment of the initial f0 maximum with respect to the initial syllable of the noun, and the height of the initial f0 peak as the following f0 peak draws closer. In order to explore a range of prosodic possibilities, emphasis and rate were also manipulated explicitly.

Methods

Two speakers of the Seoul dialect of Korean, one male (HKK) and one female (MRO), were recorded reading a corpus of sentences written in Korean script in an anechoic chamber. Both speakers were graduate students in linguistics, but were naive to the objectives of the experiment. Each of the sentences contained the syllable placed in various morphosyntactic positions, either in the subject or in the object, either initial to its noun phrase or preceded by an adjective, either medial or initial to a word. Except for the target syllables placed noun phrase initially, all of the targets followed an /i + nasal/ sequence. /sək/ appeared twice in each sentence, making four sentence types in all, which are given in Table 1. Each sentence was repeated five times in the corpus. Subject HKK read the corpus four times, each at a different rate or emphasis; first, 'normal,' then, 'slowly and distinctly,' 'fast,' and finally 'quick and emphatically.' Subject MRO read the corpus five times, 'normal,' 'slow ', 'quickly and excitedly,' 'slowly and emphatically,' and finally 'fast.' The words and glosses are given in Table 1.

Table 1. Corpus for duration measurements

1. Noun phrase initial
   a: [kimsəksan-i-nɨnsəkjəŋtʃʰəl-i-lɨlmanat-ta]
   'Kim Suksan met Suk Yungchul'
   b: [səkjəŋtʃʰəl-i-nɨnkimsəksan-i-lɨlmanat-ta]
   'Suk Yungchul met Kim Suksan'
2. Noun phrase medial
   a: [məttʃinkimsəksan-i-nɨnməttʃinsəkjəŋtʃʰəl-i-lɨlmanat-ta]
   'Good looking Kim Suksan met good looking Suk Yungchul'
   b: [məttʃinsəkjəŋtʃʰəl-i-nɨnməttʃinkimsəksan-i-lɨlmanat-ta]
   'Good looking Suk Yungchul met good looking Kim Suksan'

The sentences were then digitized and durational measurements were made on the target syllables using a waveform editor. The target syllable duration was measured as the time from the onset of perceptible frication for the /s/ to the last nonsinusoidal pitch pulse at the offset of the vowel. The acoustic duration of the vowel was also measured as the time from the first glottal pulsation without high frequency frication to the last nonsinusoidal pitch pulse at the offset of the vowel. For a reference to factor out differences due to rate, the duration of the verb /mannat/ was also measured. Its duration was the time from the onset of sinusoidal pulsation for the /m/ to the last perceptible pitch pulse at the offset of the second /a/.

The two subjects were again recorded reading a corpus of sentences in which the subject noun varied in length from three to five syllables. Also, the sentences differed in that half contained an adjective previous to the noun, and half did not. The sentence types are listed in Table 2. Subjects HKK and MRO were instructed to read through the corpus four and five times, respectively, varying rate and emphasis as is described above for the durational corpus. The token sentences were then digitized and analyzed for f0 using an autocorrelation routine on a PC6300 under MSDOS.

Table 2. Corpus for f0 measurements

No. of syll.	
3	[jəŋmi-nɨnjəŋa-lɨltʃoɑ-hæ] 'Youngmee likes Young-Ah'
4	[jəŋman-i-nɨnjəŋa-lɨltʃoɑ-hæ] 'Youngman likes Young-Ah'
5	[jəŋman-i-ne-ninjəŋa-lɨltʃoɑ-hæ] 'Youngman's people like Young-Ah'
6	[uɑhɑnjəŋmi-ninjəŋa-liltʃoɑ-hæ] 'Graceful Youngmee likes Young-Ah'
7	[uɑhɑnjəŋman-i-ninjəŋa-liltʃoɑ-hæ] 'Graceful Youngman likes Young-Ah'
8	[uɑhɑnjəŋman-i-ne-ninjəŋa-liltʃoɑ-hæ] 'Graceful Youngman's people like Young-Ah'

First, to examine the alignment of the initial tonal peak with respect to the initial syllable, and to determine the temporal proximity of the following intonational prominence on /nɨn/, the time of the initial intonational peak, the time of the major peak on /nɨn/, the onset of the utterance (taken as the onset of the first perceptible pitch pulse in the /j/), and the offset of the first syllable (taken as an amplitude minimum characteristic of the offset of the /ŋ/) were noted. An illustration of these measurements is given in Figure 3.

To examine the relative height of the f0 peaks, three f0 values were noted. The value of the initial f0 peak was taken as the first reliable f0 maximum, i.e. the first f0 value in or after the first vowel, which was higher than subsequent values. The value of the final f0 peak on /nɨn/, and the time and value of the lowest f0 value between the two peaks which occurred during a vowel were taken.

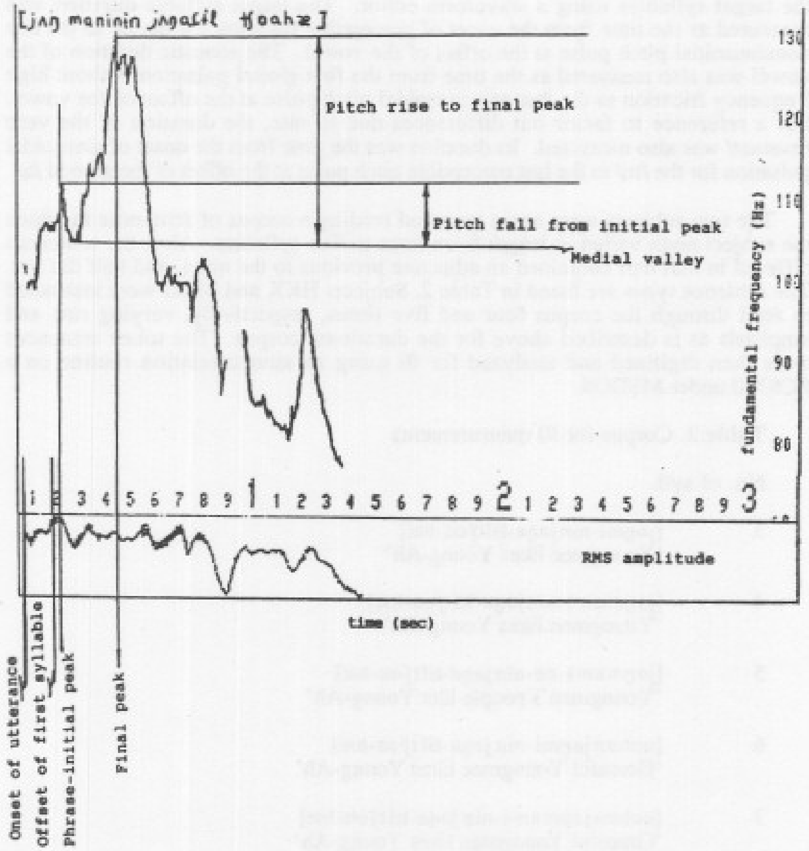


Figure 3. An f0 trace and intensity envelope for one of the token sentences with the temporal and f0 measures used to analyze the status of the initial f0 peak.

Results

Figure 4 shows that phrase initial tokens both in subject and object position are longer at a given rate than are medial tokens in the same sentences for subject HKK. The same difference appears for subject MRO, however not as consistently. Figure 5 shows that both subjects have a durational difference between word initial and word medial tokens of the same sentence also when there is an adjective preceding the word in the phrase. Word initial syllables tend to be longer than word medial tokens, regardless of the linguistic context of the word.

Closer inspection of these durational differences reveals that they are not the same as the durational differences correlated with stress in English. If one measures only vowel durations, the picture is somewhat different. Figure 6 shows that, for both speakers, there are no clear differences in vowel duration between word initial and medial syllables. At any given rate, word initial vowels and word medial vowels have the same duration. Thus, the word initial durational differences found in syllable durations seem to be localized to some extent in the initial consonant.

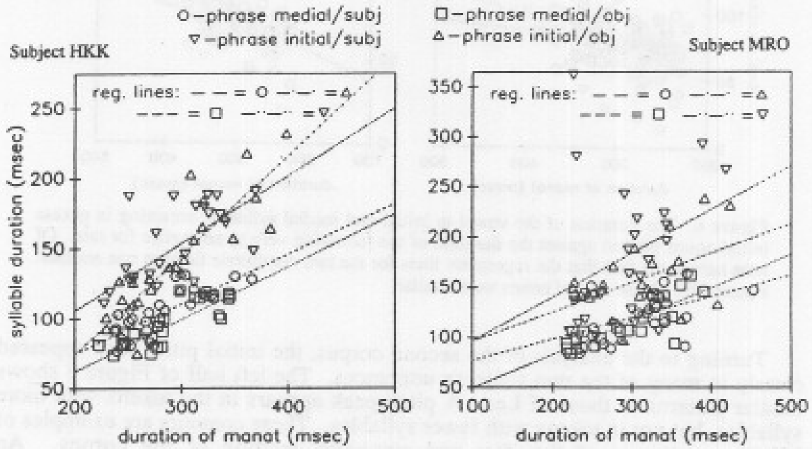


Figure 4. The duration of word initial and medial syllables occurring both in phrase initial and phrase medial positions plotted against the duration of the following verb, 'mannat' to normalize for rate. That phrase initial tokens tend to be longer than phrase medial tokens is evident in the correlation line for the phrase initial tokens being much higher in the plot than that for the medial tokens.

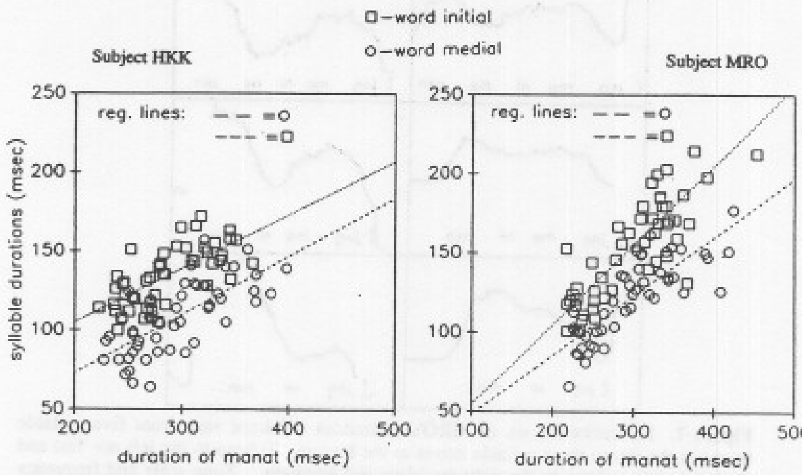


Figure 5. The duration of word initial and medial syllables occurring in phrase medial nouns plotted, as in Figure 4, against the duration of the following verb to normalize for rate.

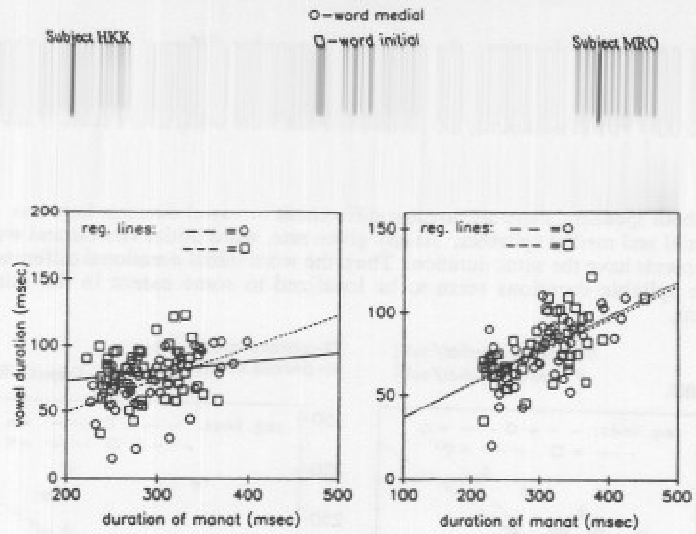


Figure 6. The duration of the vowel in initial and medial syllables occurring in phrase initial nouns plotted against the duration of the following verb to normalize for rate. Of note here is the fact that the regression lines for the two conditions lie atop one another. Patterns for phrase medial nouns were similar.

Turning to the analysis of the second corpus, the initial pitch peak appeared clearly in many of the two subject's utterances. The left half of Figure 7 shows similar patterns to those of Lee. A pitch peak appears in the tokens with more syllables, but not in tokens with fewer syllables. These contours are examples of MRO's renditions of the 'fast and emphatic' reading of the corpus. An investigation of the traces on the right, which are examples of her 'slowly and distinctly' reading of the corpus, shows a somewhat different pattern. All three of these traces have a contour similar to that in Figure 1. The initial peak is apparent even in the trisyllabic 'slowly and distinctly' tokens. The same was true of HKK's tokens.

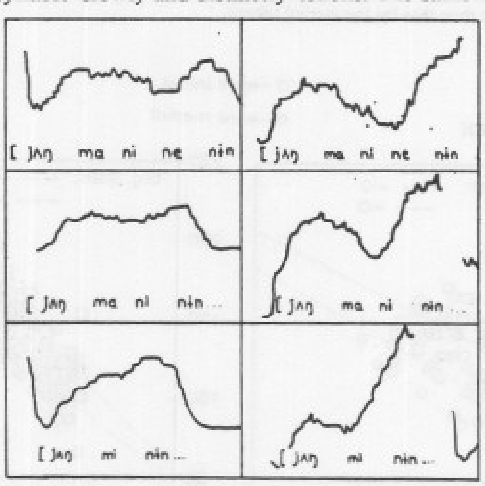


Figure 7. F0 tracks for six of MRO's utterances. Tokens vary from five syllable nouns at the top to three syllable nouns at the bottom. Tokens to the left are 'fast and distinct,' while tokens to the right are 'slow and emphatic.' Time scale and frequency scale vary from panel to panel.

I turn now to the alignment of the initial f0 peak. There is a fairly tight alignment of the initial tone with the end of the initial syllable for both subjects. Figure 8 shows a fairly restricted range of variation in tone position. Though the tone does not always appear over the initial syllable (as is shown by positive latencies in Figure 8), it never appears more than 50 ms. past the offset of the first syllable for HKK. Except for a few of MRO's "normal" renditions, her tokens also cluster around the offset of the first syllable (within 100 ms. or so). A few of the tokens also come somewhat earlier. MRO occasionally has early peaks in her "slow" renditions. Roughly half of HKK's slow and emphatic tokens also have an unusually early pitch peak, along with tokens from other conditions where the following f0 maximum occurs very closely following the initial peak (shown to the left in Figure 8).

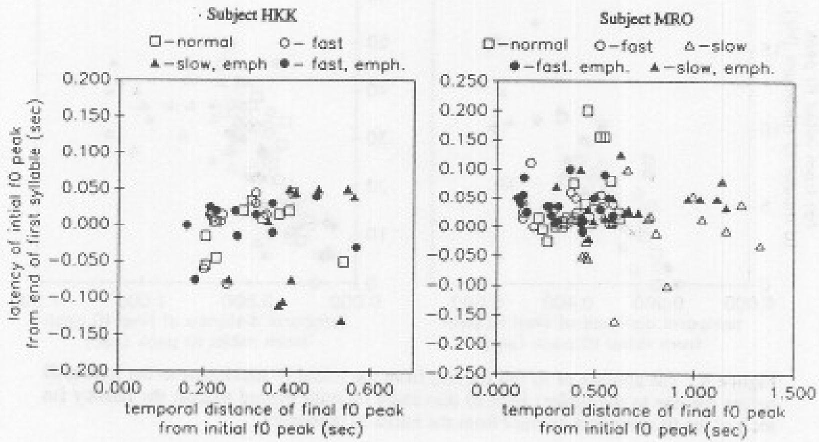


Figure 8. The latency of the initial f0 maximum from the offset of the first syllable of the phrase plotted against the latency of the f0 maximum on /nin/ from the initial f0 maximum.

The latter tokens may be reflecting tonal repulsion, which has also been shown to exist in English (Silverman and Pierrehumbert, 1990). The occurrence of the tones on or after the offset of an initial syllable has also been noted of English pitch accents (also in Silverman and Pierrehumbert, 1990). There seems to be some tendency for f0 peaks to be repulsed away from utterance initial edges. Thus, the tight clustering of f0 peaks around the offset of the initial syllable suggests that the initial pitch peak is associated with the initial syllable just as English pitch accents are associated with particular syllables.

Turning to the prominence of the pitch peak, Figure 9 shows that there is a fairly strong correlation between the prominence of the initial peak, measured as the amount of subsequent fall in f0, and the proximity of the later peak (HKK, Pearson's $r = 0.737$; MRO, $r = 0.689$). In addition to the reduction of fall from the initial high apparent in HKK's tokens, MRO's slower tokens with more than 500 ms. between the two tonal peaks show little variation in the height of fall from the initial high. The difference in the patterns for these two speakers seems to have arisen because HKK simply did not have tokens long enough to show initial highs without the effect of the following high. The leveling off of fall from the initial high for MRO's extremely long tokens shows a consistency in the height of the initial high when the effect of the following high is eliminated.

The behavior of the subject final high is in striking contrast to the initial high. Figure 10 shows that, for HKK, bringing the two tones closer together has no appreciable effect on the amount of rise to the subject final high. However, there is a relationship between the temporal distance between the two peaks and the amount of rise to the final peak in MRO's tokens.

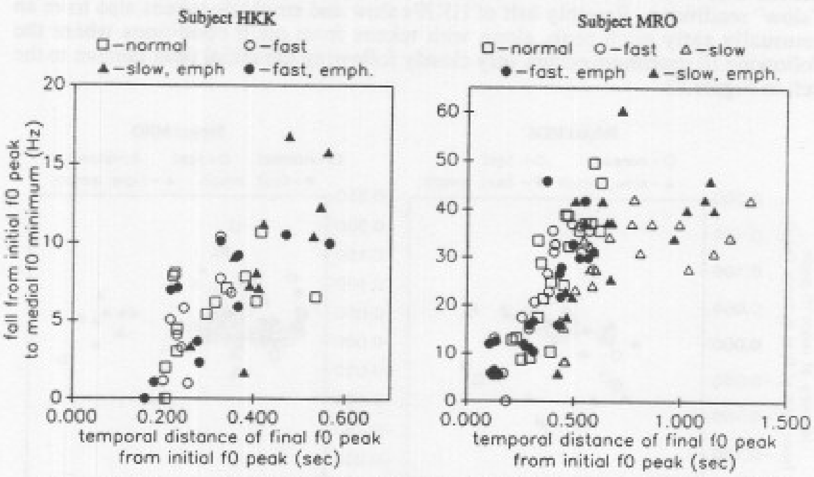


Figure 9. The amount of f0 fall (in Hz) from the initial f0 maximum to the lowest f0 before the rise to the subject final f0 maximum on /nin/ plotted against the latency (in sec.) of the f0 maximum on /nin/ from the initial f0 maximum.

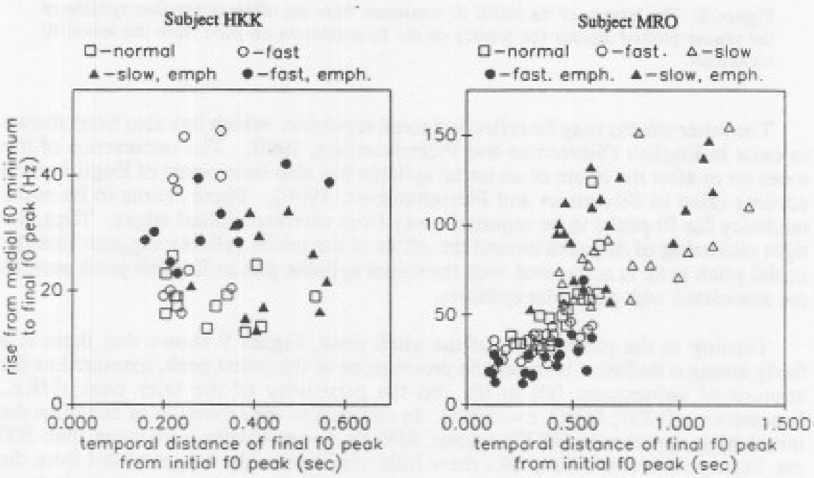


Figure 10. The amount of rise from the f0 minimum following the initial f0 maximum to the subject final f0 maximum on /nin/ plotted against the latency of the f0 maximum on /nin/ from the initial f0 maximum.

Looking closer at MRO's tokens, there is clear tendency toward separation between tokens in different conditions. MRO performed much greater changes in pitch range between the five readings of the corpus than did HKK. So, some of the correlation of pitch rise and time between f0 peaks seems to be due to concurrent effects of the stylistic manipulations on both rate (affecting time between f0 peaks) and pitch range (affecting amount of pitch rise). However, not all of the relationship is due to condition differences, because splitting the tokens by condition of the corpus does not eliminate the relationship. A positive correlation is found across the tokens within each condition (normal, r = 0.746; slow, r = 0.605; slow, emphatic, r = 0.545; fast, r = 0.678; fast, emphatic, r = 0.570).

These correlations disappear when one does not consider the amount of fall from the initial high. As can be seen in Figure 11, if one looks at the difference in the height of the two peaks, any relationship with the distance between the two peaks essentially disappears. Thus, it seems that the depth of the valley is the crucial portion of the f0 rise that is being affected by changes in the temporal distance between the two peaks.

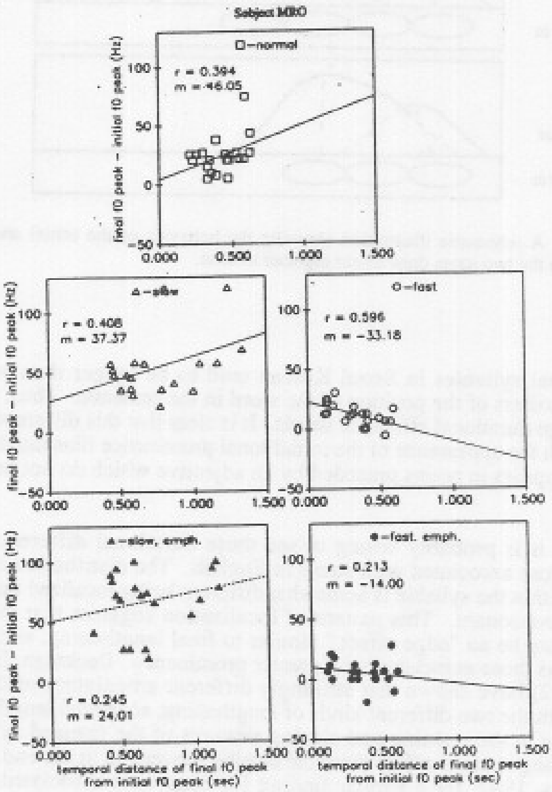


Figure 11. The difference between the value of the initial f0 maximum and subject final f0 maximum on /nin/ plotted against the latency of the f0 maximum on /nin/ from the initial f0 maximum. Tokens from MRO shown here are broken down by rate and emphasis condition.

The pattern, then, of the present data can be schematized as in Figure 12. The subject final high is obscuring the initial peak as it draws closer in time. This being the case, the disappearance of the peak in very short subject phrases under fast conditions is not positive evidence against the phrase accent analysis of the initial tones.

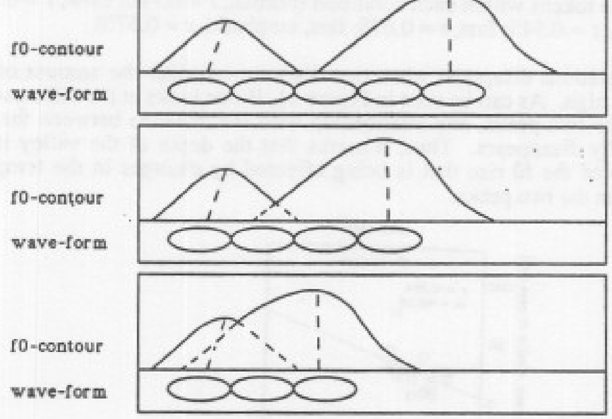


Figure 12. A schematic illustration showing the behavior of the initial and subject final tones as the two tones draw closer together in time.

Discussion

Word initial syllables in Seoul Korean tend to be longer than word medial syllables, regardless of the position of the word in the sentence. Thus, there seems to be a left edge durational effect for words. It is clear that this difference cannot be associated with the appearance of the initial tonal prominence illustrated in Figure 1, since it also appears in nouns preceded by an adjective which do not bear an initial tone.

However, it is probably wrong to see these durational differences as being identical to those associated with stress in English. The distribution of durational differences within the syllable is somewhat different being localized rather strongly in the initial consonant. This pattern of localization suggests that the durational differences may be an "edge effect," similar to final lengthening, and not a "head effect," such as those associated with greater prominence. Beckman, Edwards, and Fletcher (1992) have shown that strikingly different articulatory mechanisms are associated with the two different kinds of lengthening; accentual lengthening arises from a change in the opening and closing gestures of the stressed syllable, while final lengthening tends to affect more the closing movement at the end of the vowel (see Summers, 1987, for a similar finding for consonant associated lengthening versus stress). Thus, the durational effects found here seem to be a case of "initial lengthening."

Some of the f0 analyses reported here suggest that the initial tone is similar to the prominence lending pitch accents in English. The peak tends to be closely aligned in time with the offset of the initial syllable. This association, thus, could partially account for the perception of 'stress' on the initial syllable of nouns..

However, other f0 analyses reported here suggest that the initial peak is not to be seen as a prominence lending pitch accent. Unlike prenuclear accents in English, the peak always occurs underlyingly on initial syllables, and is not optional. The peak consistently appears when the final f0 peak is relatively far away in all renditions of the corpus for both speakers. Between these examples and those in which no initial peak is apparent, there is a continuous and predictable string of intermediate forms varying in the depth of the valley intervening between the peak and its right neighbor.

Conclusion

Cross-linguistically, prosodic effects which appear very similar on the surface, might turn out to have very different roles in the prosodic systems of the two languages. Only measurement of the right aspects of the speech production activity and signal emitted across appropriate linguistic and stylistic manipulations can tease out these differences. Of such an analysis of Seoul Korean, this paper is but a small part.

Acknowledgments

I thank Sook-hyang Lee and Sun-Ah Jun for constructing the corpora and assisting in the elicitation procedures described herein.

* An earlier version of this paper was presented at the 119th meeting of the Acoustical Society of America, May, 1990.

References

Beckman, M.E., J. Edwards, and J. Fletcher. 1992. "Prosodic structure and tempo in a sonority model of articulatory dynamics." In Docherty, G. and D.R. Ladd, eds. *Papers in Laboratory Phonology II: Gesture, Segment Prosody*. Cambridge University Press.

Bolinger, D. 1958. "A theory of pitch accent in English." *Word*, 7: 199-210.

Bolinger, D. 1965. *Forms of English: Accent, Morpheme, Order*. Harvard University Press.

Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. (Travaux de L'institut de Linguistique de Lund) CWK Gleerup.

de Jong, K.J. 1991. *The Oral Articulation of English Stress Accent*. Unpublished dissertation. Columbus, Ohio. Ohio State University.

Edwards, J., M.E. Beckman, and J. Fletcher. 1991. "The articulatory kinematics of final lengthening." *Journal of the Acoustical Society of America*, 89(1): 369-382.

Engstrand, O. 1988. "Articulatory correlates of stress and speaking rate in Swedish CV utterances." *Journal of the Acoustical Society*, 88(5): 1863-1875.

Fry, D.B. 1958. "Experiments in the perception of stress." *Language and Speech*, 1: 126-152.

Harris, K.S. 1978. "Vowel duration changes and its underlying physiological mechanisms." *Language and Speech*, 21(4): 354-361.

Huss, V. 1978. "English word stress in the postnuclear position." *Phonetica*, 35: 86-105.

Lehiste, I. 1970. *Suprasegmentals*. Cambridge, MA.: MIT Press.

Nord, L. 1986. "Acoustic studies of vowel reduction in Swedish." *STL Quarterly Progress Status Report*, 1986. (4): 19-36.

Pierrehumbert, J.B. 1980. *The Phonology and Phonetics of English Intonation*. Unpublished dissertation. Cambridge, MA.: MIT.

Pierrehumbert, J.B., and M.E. Beckman. 1988. "Japanese Tone Structure". *Linguistic Inquiry*, Monograph 15. Cambridge, MA: MIT Press.

Selkirk, E.O. 1980. "The role of prosodic categories in English word stress." *Linguistic Inquiry*, 11: 563-605.

Silverman, K. and J.B. Pierrehumbert. 1990. "The timing of prenuclear high accents in English." In Kingston, J., and M.E. Beckman, eds. *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge, MA. Cambridge University. Press.

Stälhammar, U., I. Karlsson, and G. Fant. 1973. "Contextual effects on vowel nuclei." *STL - Quarterly Progress Status Report*. 1973 . (4): 1-18.

Summers, W.V. 1987. "Effects of stress and final consonant voicing on vowel production: Articulatory and acoustic analyses." *Journal of the Acoustical Society of America*, 82: 847-863.

The domains of laryngeal feature lenition effects in Chonnam Korean

Sun-Ah Jun

Department of Linguistics, The Ohio State University

Abstract

This paper investigates the domain of two aspects of laryngeal features (voicing and voicing onset time) in the Chonnam dialect of Korean. In Korean, voiceless lenis stops, /p, t, k/, sometimes become voiced between voiced segments. Traditionally, this voicing has been discussed as occurring "within a word". Recently, Cho (1989) suggested that lenis stop voicing happens within a phonological phrase. To test this, utterances of various constructions (mostly from Cho 1989) were produced by three Chonnam speakers at three different tempi (slow, normal, fast) with three repetitions. An electroglottograph (EGG) was recorded simultaneously with an audio wave. The results show that word initial lenis stops were almost always voiced within the "accentual phrase" but not within the Phonological Phrase. A second experiment was run for the other laryngeal feature, VOT, to see whether the domain is an accentual phrase as in the case of lenis stop voicing or something else. One, two, or three-syllable words, where a test syllable [p^ha] was either at a word initial or medial position and phrase initial or medial position, were put in the frame sentence. Sixty four sentences were read with two different accentual phrasings. The result shows that VOT durations of [p^h] are significantly different between word initial and medial as well as accentual phrase initial and medial. This suggests that there is a hierarchy of prosodic level in Chonnom Korean: a prosodic word and an accentual phrase.

Introduction

Phonetic studies show that the pronunciation of segments depends on word and phrase level prosody (e.g., Lehiste, 1960). Many of these effects can be subsumed under the notion "lenition". For example, Pierrehumbert and Talkin (1989) found that the 'gestural magnitude' of /h/ is less in 'weak' positions, word medially or in deaccented words; (overall amplitude is smaller, energy is more concentrated in the first harmonic).

Keating et al.(1983) survey phonetic studies to show that many languages have different allophones of voiced or voiceless stops depending on position within a word or a phrase and on degree of stress. A cross linguistic tendency seen in these phonetic studies is that voicing-related phonetic features, such as the glottal opening gestures for voicelessness or aspiration, become lenited or weakened depending on position at some prosodic level.

Thus, in languages with voiceless stops only, voiceless unaspirated stops tend to be voiced word medially (e.g., Mandarin, Choctaw, Cuna, Korean, Tamil, cited in Keating et al.). In addition, in languages whose initial stop contrasts involve short lag with long lag VOT values, there is a common pattern of medial deaspiration of initial voiceless aspirated stops (e.g., Lisker & Abramson 1964 for English). A question that arises in looking at these studies is what is the domain of the lenition effects. Clearly it is not always the word. Word-initial voiceless stops in Burmese becomes voiced in phrases, especially after a "weakened" (toneless and reduced) syllable. In Polish, word final stops become voiced before a vowel-initiated word (Keating et al., 1983). That is, the domain of the lenition effects can be larger than a word.

This paper investigates the domain of two phonetic aspects of laryngeal features (voicing during closure and VOT after release) in the Chonnam dialect of Korean. Korean has no voicing contrast for obstruents, but it has phonemically a 3-way contrast among voiceless obstruents.

Aspirated:	p^h	t^h	k^h	$tʃ^h$	
Fortis:	p'	t'	k'	tʃ'	s'
Lenis:	p	t	k	tʃ	s

As mentioned in Keating et al.(1983), Korean is believed to show some lenition phenomena word medially. That is, slightly aspirated voiceless stops (or lenis stops) become voiced intervocalically. Lisker and Abramson (1964), however, noticed that word-initial lenis stops are also voiced when preceded by a monosyllabic modifier which ends in a sonorant (e.g., {i} 'this' + {tal} 'a moon' => [idal] 'this moon').

Therefore, we must look to a larger unit as the domain of this effect. Recently, Cho (1989) suggested that, in the Seoul dialect of Korean, lenis stop voicing happens within a phonological phrase whose boundaries are determined by the syntactic structure of a sentence or constituent. Since Chonnam dialect is the same as Seoul in its syntactic structures (even though it differs from Seoul in its intonation patterns and by some lexical items) , the domain of voicing, if we follow Cho's analysis, should be the same for both dialects. A pilot study showed , however, that the domain of voicing in the Chonnam dialect changed depending on speech rate and intonation pattern, suggesting that the domain cannot be a unit defined entirely by syntax.

In the Chonnam dialect of Korean, there is a phrase which can be defined by the tonal pattern, called "accentual phrases" by Jun (1989). An accentual phrase in Chonnam has either of two tonal patterns, Low-High-Low or High-High-Low, as determined by the laryngeal feature specification of the first segment of the phrase: if the segment is either [+constricted glottis] or [+spread glottis], the phrase has a High-High-Low accentual pattern and otherwise a Low-High-Low pattern. Figure 1 (from Jun, 1989) is an example of an utterance showing accentual phrases.

The second and the fourth accentual phrases have H-H-L accentual patterns since their phrase initial sounds are [+spread glottis] ([k^h] and [s], respectively). The other accentual phrases have L-H-L patterns. (The final H-L pattern is a boundary tone(%) for the intonational phrase.)

An accentual phrase in Chonnam can contain more than one word. This can be seen later in figure 5. To determine whether the accentual phrase is the domain of lenis stop voicing, an experiment was designed in such a way that many sentences contain an accentual phrase having more than a word and, at the same time, most words begin with a lenis stop.

Since any word in a phrase did not have its own tonal pattern nor have any influence in the tonal pattern of the phrase, the lowest level of prosodic structure in Chonnam was claimed to be an accentual phrase. Therefore, the level of prosodic word was not motivated before. But it is possible that a prosodic word is a domain of other phonological features and becomes the lowest level of prosodic hierarchy.

As noticed by Lisker & Abramson (1964), cross-linguistically aspirated stops tend to be deaspirated word medially. Since Korean has voiceless aspirated stops word initially and medialy, the duration of VOT after release was examined to see whether there is any similar lenition phenomenon depending on positions in a word or a phrase and, if there is, whether the domain of VOT lenition is also an accentual phrase or something else.

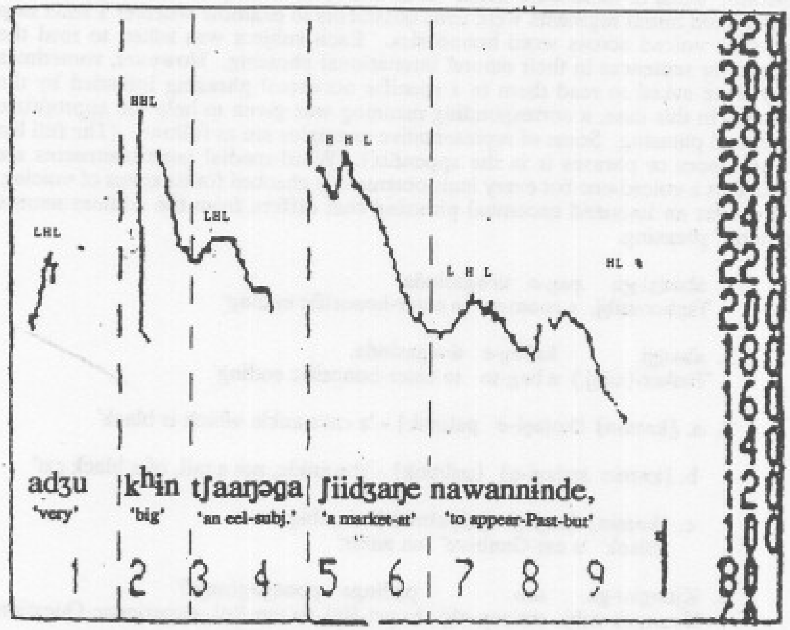


Figure 1. A pitch track of a phrase, {adʒu} {kʰin} {tʃaaŋəga} {ʃiidʒaɲe} {nawanninde}, 'very' 'big' 'an eel-subj.' 'a market-at' 'appeared- but' => 'A very big eel came out at the market, but...'.

Experiment 1 : Lenis Voicing

Methods

Subjects:

Two male speakers and one female speaker of Chonnam dialect participated. They were in their late twenties. Two of them, one male and one female, had lived in Kwangju, the main city of Chonnam province, for 26 years and one other male had been in Kwangju for 19 years and moved to Seoul and lived there for 7 years until he came to the States. This last subject sometimes showed a mixed intonation pattern between Chonnam and Seoul intonation. Thus, to elicit the Chonnam intonation, about 30 minutes conversation was made between him and a native Chonnam speaker before recording, and contexts which include many lexical items typical of the Chonnam dialect were given for the test phrase or sentences. Recording was made when the intonation of his utterances was very close to that of a native Chonnam speaker's.

Material

Thirty-four phrases or sentences with various syntactic structures were selected. Most of these were taken from Cho's(1989) corpus of examples. Most of the word initial segments were lenis obstruents to examine whether a lenis stop becomes voiced across word boundaries. Each subject was asked to read the phrases or sentences in their natural intonational phrasing. However, sometimes they were asked to read them in a specific accentual phrasing intended by the author. In this case, a corresponding meaning was given to help the appropriate accentual phrasing. Some of representative examples are as follows. (The full list of sentences or phrases is in the appendix). Word-medial lenis obstruents are spelled as a voiced one but every lenis obstruent is checked for its status of voicing. { } means an intended accentual phrasing that differs from the context-neutral "natural" phrasing.

1. abədʒi-ga paŋ-e tirəgasinda.
'Father-subj. a room-to to enter-honorific ending'

2. abədʒi kabaŋ-e tirəgasinda.
'Father-(subj.) a bag-to to enter-honorific ending'.

3. a. {kəmin} {kojaŋi-e palmok} - 'a cat's ankle which is black'

b. {kəmin kojaŋi-e} {palmok} - 'the ankle, not a tail, of a black cat'

c. {kəmin} {kojaŋi-e} {palmok} - 'ambiguous'
'black' 'a cat-Genitive' 'an ankle'

4. Kjungsu-ga ton padingə pondʒəginnja?
'Kyungsu-subj. money-obj. to get-Rel. to see-Rel.-experience-Question'
=> 'Have you seen Kyungsu get or receive money?'

5. Uri sənseŋin Suni-hantʰe tʃəmsuril tʃal tʃuningət kat'ira.
'our teacher-subj. Suni-to a score-obj. well to give-Rel. to seem.'
=> 'It seems that our teacher gives a good score to Suni.'

Procedure

Subjects were instructed to say each sentence or phrase in its 'natural' intonation pattern unless marked : they were asked to read a sentence as if they are talking to someone in a natural conversation. Subjects repeated each phrase three times at three self-selected rates, normal, slow, and fast, in a sound-attenuated room. Utterances were blocked by rate.

In order to ascertain more directly the intended voicing of a stop, an Electroglottograph (henceforth EGG) recording was made. For EGG, a high frequency electrical current is passed through the larynx, between electrodes placed on the neck surface on the left and the right sides. The electrical impedance (resistance) between the electrodes depends on the glottal area, being small when the vocal folds are pressed firmly against each other and large when the folds are completely separated. To get EGG data, The EGG band was held tightly around the subject's neck while he or she was reading. The audio wave and EGG signals were simultaneouslytracked and digitized simultaneously. Using computer programs, the two signals were separated and viewed synchronically.

For every lenis obstruent position, the two signals were checked for sinusoidal periodicity, indicative of voicing. The voicing status of each lenis obstruent was ascertained independently from the audio and the EGG signal. Almost all of the time, the two signals agreed in voicing status: either both showed periodisity or neither did. However, there were a few cases when the two disagreed; either the audio showed voicing and the EGG did not, or (very rarely) the EGG showed voicing and the audio did not. Figure 2 shows sample traces for each type of case.

In Figure 2 , (a) + (b) are taken from two utterances of the same sentence with different accentual phrasings.: 1. {Kjungsu-ga} {tonɨl} {} {pondʒəginnja}? 2. {Kjungsu-ga} {tonɨl badingə} {pondʒəginnja}? For each picture, the upper two windows are for Audio signals and the lower two windows are for EGG signals and they are synchronized. The second and the fourth windows are expanded views beginning from the small tick points in the first (Audio wave) and the third (EGG wave) window, respectively. The horizontal dimension in each window indicates the time dimension and the vertical dimension indicates the amplitude of signal. In Figure 2(a), the verb-initial lenis stop, /p/ of {padi-ngə}, is voiceless as shown by the circles of both Audio and EGG wave forms, whereas it is voiced by both signals in Figure 2(b). Figure 2(c) is part of sentence 5 above, '*Uri sənsegɨn Sunihantʰe tʃəmsurɨl tʃal tʃuningət kat'ɨra'*. The second and the fourth window show an expanded view for {tʃəmsurɨl}; as shown by the waveforms inside the circle, the fricative has a voiced Audio signal but a voiceless EGG signal.

For these few cases where the two signals disagreed, the EGG signal was assumed to indicate the speaker's intention, since it can give information about glottal adduction and abduction unfiltered by the vocal tract. It is already noticed in the literature that the same glottal configuration can produce voiced or voiceless sounds depending on the vocal tract function and air pressure difference. That is, when vocal folds are abducted and supraglottal air pressure is higher due to some constriction within a vocal tract, then vocal folds are not vibrating and a voiceless sound occurs. On the other hand, when there is a high subglottal air pressure and a vocal tract is open, then vocal folds can vibrate while abducted. A good example is a breathy or murmured /h/ and figure 3 shows this. The sentence is:*Youngsunin Hekjungirɨl* tʃoahe 'Youngsu likes Hekyung.'

In Figure 3, the second (Audio) and fourth (EGG) window show waveforms around [h]. The EGG signal shows low-amplitude vibrations for an open glottis while the Audio signal shows high-amplitude vibrations continuously through the /h/. Thus, it is the EGG signal which gives more reliable information about glottal adduction and abduction.

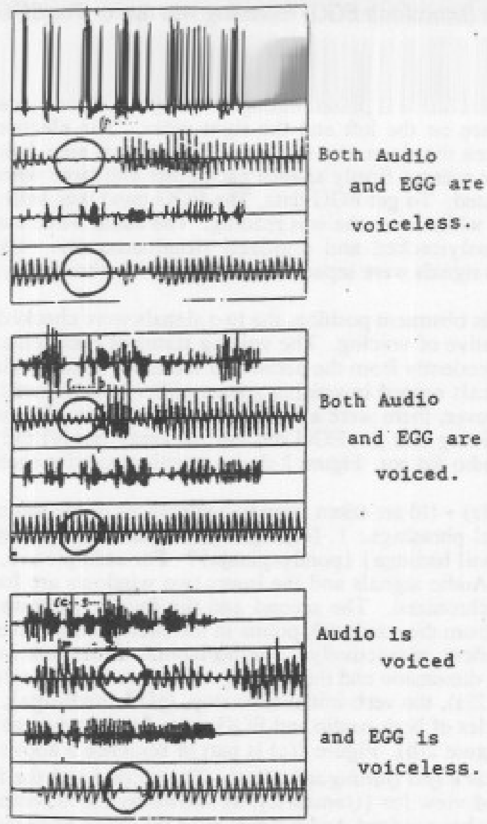


Figure 2. Audio & EGG wave forms: (a) When both are voiceless (b) When both are voiced (c) When audio is voiced and EGG is voiceless.

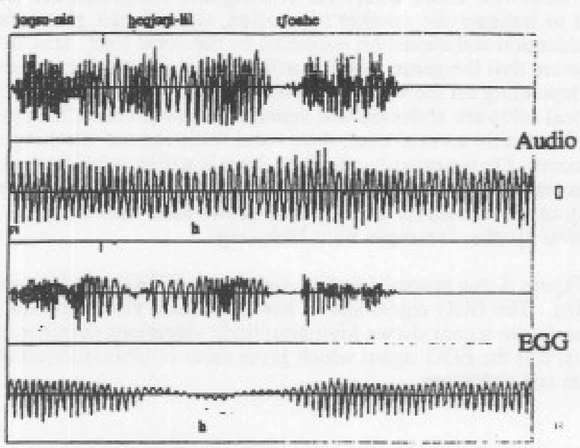


Figure 3. Audio and EGG waveforms of *yəŋsunin <u>hegjəŋril</u> tʃoahe*. The second and fourth window shows waveforms around underlined part of the sentence.

Results and Discussion

Both EGG and Audio wave form data showed that lenis obstruents in word initial position between voiced segments become voiced almost all the time in accentual phrase medial position but not at accentual phrase initial position. Figure 4 and 5 illustrate these contrasting positions.

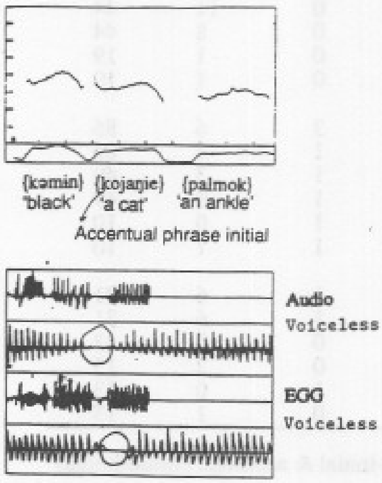


Figure 4. Pitch tracks of *kəmin kojaŋie palmok* "a black cat's ankle" uttered in three accentual phrases. Audio & EGG waveforms show a voiceless /k/ of *koyaŋie* at an accentual phrase initial position.

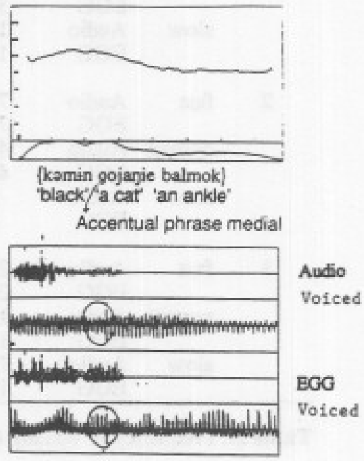


Figure 5. Pitch tracks of *kəmin kojaŋie palmok* "a black cat's ankle" uttered in one accentual phrase. Audio & EGG waveforms show a voiced /k/ of *koyaŋie* at an accentual phrase medial position.

They show pitch tracks and Audio/EGG waveforms for 'kəmin koyaŋi-e palmok' example phrase 3 above in the material section. Figure 4 shows an utterance in which the phrase is broken into three accentual phrases,{kəmin}{koyaŋie}{palmok}, while Figure 5 shows an utterance where the whole phrase forms one accentual phrase, {kəmin goyaŋie balmok}. When the lenis stop, /k/ of /koyaŋi/ "a cat" is in accentual phrase initial position in figure 4, it is voiceless in both waveform signals, but as an accentual medial /k/ in figure 5, it is voiced.

For cases where there was a discrepancy between signals, the EGG showed what was expected from the position in terms of accentual phrasing. The following two tables show the frequencies of occurence of lenis obstruent voicing for each speaker for three speech rates: Table I is that of word-initial & accentual phrase-medial position and Table II is that of word-initial & accentual phrase-initial position. A total of 195 tokens (three repetitions of 65 word) were examined for each signal. Due to some missing tokens for each speaker, the sum of totals in tables I & II are not exactly 195. Ambiguous cases are where the signal shows so weak vibration that the decision of voicing was difficult.

Table I: Frequency of voicing (word-initial and accentual phrase-medial)

Speaker	Rate	signal	Clearly voiced	Clearly voiceless	Ambig-uous	total
1	fast	Audio	63	0	3	66
		EGG	62	0	4	66
	normal	Audio	33	0	11	44
		EGG	36	0	8	44
	slow	Audio	18	0	1	19
		EGG	18	0	1	19
2	fast	Audio	77	3	6	86
		EGG	77	1	8	86
	normal	Audio	45	1	3	49
		EGG	45	1	3	49
	slow	Audio	9	1	0	10
		EGG	8	1	1	10
3	fast	Audio	24	1	6	31
		EGG	24	1	6	31
	normal	Audio	11	0	2	13
		EGG	11	0	2	13
	slow	Audio	12	1	0	13
		EGG	12	0	1	13

Table II: Frequency of voicing (word-initial & accentual phrase-initial)

Speaker	Rate	signal	Clearly voiced	Clearly voiceless	Ambig-uous	total
1	fast	Audio	7	97	5	109
		EGG	8	94	11	113
	normal	Audio	1	124	6	131
		EGG	1	125	11	137
	slow	Audio	0	170	4	174
		EGG	0	173	1	174
2	fast	Audio	4	82	21	107
		EGG	4	96	7	107
	normal	Audio	1	137	2	140
		EGG	0	138	2	140
	slow	Audio	0	169	9	178
		EGG	0	168	10	178
3	fast	Audio	4	131	6	141
		EGG	4	131	6	141
	normal	Audio	2	151	4	157
		EGG	2	149	6	157
	slow	Audio	1	165	0	166
		EGG	1	165	0	166

As seen in Table I, voiceless lenis obstruents in word-initial position within an accentual phrase are mostly voiced and are rarely voiceless. At the same time, the total number of tokens which are both word-initial and phrase-medial tokens is smaller as the rate decreases from fast to slow, indicating faster speech contains fewer accentual phrases. This tendency is again shown in the Table II. The slower the rate, the more accentual phrases there are in an utterance. This tendency is very clear for the first two speakers, but less clear for the third speaker, whose general speech rate is very slow compared with others.

The observation that a fast rate utterance tends to have more voicing can be supported by the idea about phonetic representation expressed in Browman and Goldstein(1987). They assume that the gestures are invariant across different contexts. However, using the inherently spatiotemporal characteris-tics of gestures, they claim that "it is possible for gestures to overlap in time. Such overlapping activation of several invariant gestures results in context-varying articulatory trajectories when the gestures involve the same articulators, and in varying acoustic effects even when different articulators are involved. That is, much coarticulation and allophonic variation occur as an automatic consequence of overlapping invariant underlying gestures" (p.2).

Therefore, the intervocalic voicing in casual or fast speech may involve not only reduction of gestural magnitude of the glottal opening-and-closing gesture responsible for the voicelessness but also the blending of closely phased voicing gestures. Moreover, this intervocalic lenition is very highly likely to occur within an accentual phrase instead of across accentual phrases. This indicates that the magnitude of voicing gesture is reduced in the middle of an accentual phrase, but not reduced (maybe increased) at the accentual phrase initial position. This may be possible due to the existance of an accentual phrase boundary. That is, there would be enough time for the voicing gesture of the lenis stop to reach its target position at an accentual phrase initial position. (As will be shown in the next experiment, the dura-tion of features is longer at an accentual phrase initial position than non-initial position.) As a result, it would be less likely that the voicing gesture of the preceding vowel and that of the following lenis stop are overlapping with each other.

This interpretation is supported by the results in Table II which shows the opposite pattern to table I. That is, most word-initial & phrase-initial lenis obstruents are voiceless and are rarely voiced. The results in these two tables constitute strong evidence that a word-initial voiceless lenis obstruent becomes voiced accentual-phrase medially but remains voiceless accentaul phrase initially. Therefore, the domain of voicing is the accentual phrase in the Chonnam dialect.

Experiment 2 : VOT lenition

In this experiment, the VOT duration of /p^h/ was measured to see whether there is any lenition effect at all and, if there is, what is the domain of lenition effect. Three outcomes are possible, with different interpretations:

Outcome 1 : If there is no lenition effect of aspiration, there would be no significant difference in VOT duration depending on position.

Outcome 2 : If there is a lenition effect of aspiration and if its domain is an accentual phrase, then there would be two significant groupings of the VOT duration; one group in accentual phrase initial and the other group in medial.

Outcome 3 : If the domain is some unit smaller than an accentual phrase (such as the prosodic word), there would be at least two groupings of VOT duration within an accentual phrase.

Method

Subjects

The two speakers from the first experiment other than the male speaker who had been in Seoul participated in this second experiment.

Material

1, 2, or 3 syllable words where /p^{h}/ was either word initial onset or word medial onset were put in the frame sentence:

'igəsɨn ____ hago ____ gɨman'
'This is ____ and ____.'

Twenty-eight words were put in either position of the frame sentence which was read with either of the following accentual phrasings.

1. {igəsɨn} { ___ hago} {___ gɨman}
2. {igəsɨn} { ___ hago ___ gɨman}

There were 32 frame sentences for each of the two accentual phrasings above. The first accentual phrasing above is often found in normal reading without focusing any special lexical item. The second accentaul phrasing is found when the first conjunct is narrowly focused. To get this phrasing, each subject was told to emphasize (or to put a focus on) the first conjunct item before recording. The words with initial and medial /p^{h}/ are shown in Table III:

Table III : Words with initial and medial /p^{h}/

< 11 Words with initial /p^{h}/ >

p^{h}a	'green onions'	p^{h}ado	'surge'
p^{h}are	'sea lettuce'	p^{h}adʒaŋ	'wave length'
p^{h}ari	'a fly'	p^{h}adʒən	'a grilled food with onions'
p^{h}araŋse	'a blue bird'	p^{h}atʃ^{h}uls'o	'a police station'
p^{h}adʒama	'pajamas'	p^{h}atʃ^{h}ulpu	'a maid'
p^{h}agimtʃ^{h}i	'pickled green onions'		

< 17 Words with medial /p^{h}/ >

jaŋp^{h}a	'onions'	tanp^{h}a	'a short wave'
ɨnp^{h}a	'silvery waves'	tʃaŋp^{h}a	'a long wave'
jəp^{h}a	'aftershock'	kjəkp^{h}a	'destruction'

sop^{h}a	'ripples'	jaŋp^{h}ak'aŋ	'onion snack'
ʧup^{h}asu	'frequency'	hwip^{h}aram	'a whistle'
map^{h}aram	'the south wind'	kodʒup^{h}a	'a high frequency'
ʧ^{h}odanp^{h}a	'ultra-sonic'	ʧəŋip^{h}a	'defender of justice'
p^{h}ump^{h}ari	'street vender'	insaŋp^{h}a	'impressionist art'
naŋmanp^{h}a	'romantic art'		

Depending on the position of /p^h/ within a word and within an accentual phrase eight groups can be defined. These are shown in Table IV.

Table IV: 8 groups of prosodic positions

Group 1: { } {p^h __} { }	}	Word initial
Group 2: { } { } {p^h __}	}	Accentual phrase initial
Group 3: { } {p^h ________}	}	
Group 4: { } { ____ p^h ___}	--	Word initial Accentual phrase medial
Group 5: { } {__p^h__} { }	}	
Group 6: { } { } {__p^h__}	}	Word medial
Group 7: { } {__p^h________}	}	Accentual phrase medial
Group 8: { } {________p^h__}	}	

Procedures

Subjects read each sentence five times at normal rate in a sound-attenunated booth. The sound was digitized and sound waveforms were displayed using a waveform editing program. The duration of VOT was measured from the release of stop to the beginning of complex waveform for the following vowel. The duration of VOT was measured using waveform editing program.

Results and Discussion

A one-way ANOVA was run on the VOT durations for the 8 groups and showed a significant main effect of group ($p < .01$). A Tukey test was performed between groups at alpha = 0.05 level. Each speaker shows somewhat different relationships among groups but both speakers show a significant difference between word initial and word medial groups. At the same time, each speaker also showed a significant ($p < .01$) difference between accentual phrase initial and medial.

This suggests that the smallest domain of VOT duration is not an accentual phrase but a prosodic word. Figure 6 shows the mean duration(in ms) of each group for each subject. There was no significant difference between word medial groups, group 5 to 8, and between word-initial and accentual phrase initial groups, group 1 to 3. Therefore, there were three groupings depending on position within a word or a phrase: 1. word-initial and phrase-initial, 2. word-initial but phrase-medial, 3. word-medial and phrase-medial.

For both subjects, the duration of VOT is significantly longer word initially,(in group 4), than word medially (groups 5 to 8). Also the duration of VOT is significantly longer accentaul phrase initially (groups 1 to 3), than accentual phrase medially (group 4).

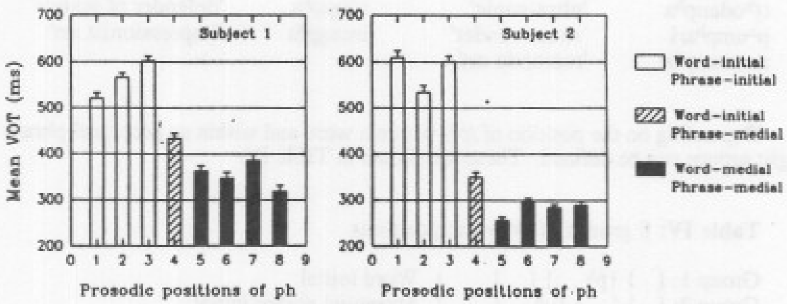


Figure 6. Mean duration of VOT in different prosodic positions.

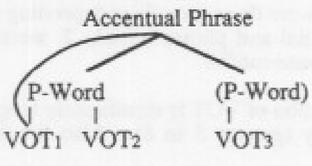
Conclusion

This investigation of two phonetic aspects of laryngeal features of Chonnam Korean suggests that there are a voicing lenition effect intervocalically in a phrase medial position and a reduced aspiration duration word medially as compared to word initially. The domain of voicing of the voiceless lenis obstruents is shown to be an accentual phrase, which was defined by Jun(1989) based on the intonational pattern of a sentence. That is, word initial lenis stops become voiced intervocalically if the word is within an accentual phrase and all accentual phrase initial lenis obstruents are voiceless.

In Jun(1989), an accentual phrase was the lowest level of the prosodic hierarchy and there was no basis for positing a level of prosodic word, which has been proposed in the literature on prosodic phonology by such researchers as Selkirk(1980, 1986) and Nespor & Vogel(1986).

In the second experiment, however, the duration of VOT of aspirated consonants in Chonnam showed that the domain of VOT is smaller than an accentual phrase suggesting that there is a level of prosodic word. That is, the duration of VOT in Chonnam was longer word-initially than word-medially even within an accentual phrase.

On the other hand, the duration of VOT was also longer accentual phrase initially than phrase medially. This suggests that there is a hierarchy of prosodic levels: an accentual phrase is higher than a prosodic word (P-word) level. Since there is an edge effect (longer VOT at left edge) of each prosodic level, we can represent the realization of different VOT durations using a metrical tree as follows.



VOT in P-word initial boundary (VOT_1 or VOT_3) is longer than that within the P-word (VOT_2) and this word-initial VOT is further longer if it occurs at an accentual phrase initial boundary (VOT_1) than at the accentual phrase medial position (VOT_3). And this may sugest that the word-initial laryngeal feature is also linked to the higher prosodic level, an accentual phrase.

Acknowledgement

I would like to thank Prof. Mary Beckman for her valuable comments and I. Park and Y.S. Chung for serving as subjects in the experiments.

* An earlier version of this paper was presented at the 119th meeting of the Acoustical Society of America, State College, Pennsylvania, May, 1990.

Appendix

Thirty four sentences or phrases used in the first experiment.
Word-initial lenis obstruents examined for voicing are underlined.
Word medial lenis stops, /s/ and /h/ are also examined for voicing.

1. abədʒi-ga paŋ-e ɨirəgasi-ngə pwannja?
'Father-subj. a room-loc. to enter-honor.-rel. to see-question'
==> 'Did you see Father entering the room?'
2. abədʒi kabaŋ-e ɨirəgasi-ngə pwannja?
'Father-subj. a bag-loc. to enter-honor.-rel. to see-question'
3. Kjəŋman-iga tʃal kat'a.
'Kyəŋman-sujb. well to go-past' ==> 'Kyəŋman went safely.'
4. Kjəŋman-iga tʃal kat'a. ==> 'It is good for us that Kyəŋman left.'
5. kəmɨn kojaŋi-e palmok
'black a cat-poss. an ankle' ==> 'a black cat's ankle'
6. kəmɨn kojaŋi-e palmok ==> 'a cat's ankle which is black'
7. kit'e Seoul-esə pon kɨ kɨrim-i katʃ'a-radɨra.
'then Seoul-loc. to see-rel. that a picture-sujb. fake-they said-decl.'
==> 'They said the picture (we) saw at Seoul then was fake.'
8. ne kɨrim-i tʃeil nat'adɨra.
'my a picture-subj. very best-they said-decl.'
==> 'They said my picture was the very best.'
9. tʃatoŋtʃʰa-hago kɨrim
'a car-and a pitcure' ==> 'a car and a picture'
10. kɨgən adʒu tʃoin kɨrim-igɨman.
'that-subj. very good a picture-decl.'
==> 'That is a very good picture.'
11. kɨ tʃoin kɨrim
'the good a pitcure' ==> 'the good picture'
12. Kisu-ga pon kɨrim
'Kisu-subj. to see-rel. a picture' ==> 'a picture Kisu saw'
13. Kisu-ga kɨrin kɨrim-ɨl pwannɨnde, tʃəŋmal tʃal --
'Kisu-subj. to draw-rel. a picture-obj. to see-past-and, really well
kɨrjət'əra.
to draw-past-decl.'
==> '(I) saw a picture Kisu drew and it was really well drawn.'
14. na abadʒi-ga tʃusi-n kɨrim pogois'əs'ə.
'I Father-subj. to give-honor.-rel. a picture to look at-past prog.'

==> 'I was looking at the picture Father gave to me.'

15. na abədʒi-ga tonseŋ-hant^{h}e tʃun kɨrim pogois'əs'ə.
'I Father-subj. brother-dat. to give-rel. a picture to look at-past prog.'
==> 'I was looking at the picture Father gave to my brother.'

16. ke-ga tʃa-nga?
'a dog-subj. to sleep-question' ==> 'Does the dog sleep?'

17. ke-ga kin-ga?
'a dog-subj. to crawl-question' ==> 'Does the dog crawl?'

18. ki kirim tʃemog-i adʒu tʃoin tʃandi-radira.
'the a picture a title-subj. very good lawn-they say-decl.'
==> 'They said the title of the picture was 'very good lawn'.'

19. Kjəŋsu-ga ton padɨn-gə ponjəginnja?
'Kyəŋsu-subj. money to receive-rel. to see-experience-question.'
==> 'Have you seen Kyəŋsu received money?'

20. Kjəŋsu-hant^{h}e ton tʃu-ngə pondʒəginnja?
'Kyəŋsu-dative money to give to see-experience-question.'
==> 'Have you seen someone gave Kyəŋsu money?'

21. uri sənseŋin Suni-hant^{h}e tʃəmsu-lɨl tʃal tʃu-ningə gat'ira.
'our teacher-subj. Suni-dat. a score-obj. good to give is likely-decl.'
==> 'It seems to be that our teacher gives a good score to Suni.'

22. pabɨl tʃaju məgəra.
'rice-obj. often to eat-imperative.'
==> 'Eat (your) rice often!'

23. isaŋhage kojaŋi-ga ul-go ke-ga tʃisəs'at^{h}ira.
'strangely a cat-subj. to cry-and a dog-subj. to bark-habitual-decl.'
==> 'Strangely, a cat has been crying and a dog has been barking.'

24. nə kojaŋi tʃ'onnin kaŋadʒi t'eri-n pabo pwannja?
'you a cat o follow-rel. a puppy to bit-rel. a fool to see-experience'
==> 'Have you seen a fool who is bitting a puppy which follows a cat?'

25. Kjəŋsu-nɨn pimat-k'oinnɨn kojaŋi-lɨl poat'a.
'Kyəŋsu-subj. to be in the rain-rel. a cat-obj. to see-past.'
==> 'Kyəŋsu saw a cat in the rain.'

26. na indʒesa tʃ^{h}egɨl tu-gwən pwat'a.
'I now a book-obj. two to read-past.'
==> 'I now just read two books.'

27. na indʒesa tʃ^{h}ek tu-gwənɨl pwat'a.
'I now a book two-classifier to read-past.'
==> 'I now just read two books.'

28. jədʒa-ga tu-mjəŋ tʃinaga-nda.
'women-subj. two-people to pass by-prog.'
==> 'There are two women passing by.'

29. jədʒa tu-mjəŋ-i tʃinaga-nda.
'women two-people-subj. to pass by-prog.'
==> 'There are two women passing by.'

30. pab is'ə.
'rice to have' ==> '(We) have rice.'

31. abənim kjesjə
'Father-honor. to be-honor.-decl.'
==> 'Father is inside (or at home).'

32. Kosənseŋnim-i kjesjə.
'Ko teacher-honor-subj. to be-honor.-decl.'
==> 'Teacher Ko is inside.'

33. abenim kjesimnik'a?
'Father-honor. to be-honor.-question.'

==> 'Is Father inside (or at home)?'
34. K̲osənseŋnim-i k̲jesimnik'a?
'Ko teacher-honor-subj. to be-honor-question.'
==> 'Is Teacher Ko inside?'

References

Abberton E. 1972. "Some Laryngographic Data for Korean Stops," *International Phonetic Association*. 2: 67-78.

Abramson, A.S. and L. Lisker. 1971. "Voice Timing in Korean Stops," *Status Report on Speech Research* 27. (Haskins Laboratories, New Haven).

Beckman, M. and J. Pierrehumbert. 1988. "Intonational Structure in Japanese and English," *Phonology Yearbook*. 3: 255-309.

Browmand C. and L. Goldstein. 1987. "Tiers in Articulatory Phonology, with Some Implications for Casual Speech," *Status Report on Speech Research*, Haskins Laboratories.

Cho, Y.-M. Y. 1987. "The Domain of Korean Sandhi Rules", Paper presented at LSA meeting.

Cho, Y.-M. Y. 1988. "Syntax and Phrasing in Korean", *The Phonology-Syntax Connection*. Stanford, CA: Stanford University.

Dart, S. 1986. "An Aerodynamic Study of Korean Stop Consonants: Measurements and Modeling," *Journal of Acoustical Society of America* 81(1): 138-147.

Han, M.S. and R.S. Weitzman. 1970. "Acoustic Features of Korean /P, T, K/, /p, t, k/ and /ph, th, kh/", *Phonetica* 22:112-128.

Hirose, H., Lee, C.Y. and T. Ushijima. 1974. "Laryngeal Control in Korean Stop Production," *Journal of Phonetics* 2: 145-152.

Jun, S.-A. 1989. "The Accentual Pattern and Prosody of the Chonnam dialect of Korean," *Harvard Studies in Korean Linguistics III*, S. Kuno, et al. eds., Harvard Univ. Cambridge, Mass.

Keating P. & W. Linker and M. Huffman. 1983. "Patterns in allophone distribution for voiced and voiceless stops," *Journal of Phonetics* 11: 277-290.

Koo, H.S. 1986. *An Experimental Acoustic Study of the Phonetics of Intonation in Standard Korean.*. Ph.D. Diss. The University of Texas at Austin.

Lehiste, I. 1960. "An Acoustico-Phonetic Study of Internal Open Juncture." *Phonetica* 5. Supplement.

Lisker, L. and S. Abramson. 1964. "A Cross-language Study of Voicing in Initial Stops: Acoustical Measurements," *Word* 20, No. 3: 384-422.

Nespor, M. and I. Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris Publication.

Pierrehumbert J. and D. Talkin. 1989. "Lenition of /h/ and Glottal Stop." In, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. ,G. Docherly & D.R. Ladd, eds. p 90-116, Cambridge Univrersity Press.1992

Selkirk, E.O. 1980. "Prosodic Domain in Phonology: Sanskrit Revisited",M. Aronoff and M.K. Kean eds., *Juncture*. Saratoga:Anma Libri.

Selkirk, E.O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge. Mass.:Mit Press.

Selkirk, E.O. 1986. "On Derived Domains in Sentence Phonology," *Phonology Yearbook*. 3.

Silva, D.J. 1989. "Determining the Domain for Intervocalic Stop Voicing in Korean," *Harvard Studies in Korean Linguistics III*, S. Kuno et al. eds., Harvard University. Cambridge, Mass.

The Timing of Lip Rounding and Tongue Backing for /u/

Gina M. Lee

Department of Linguistics, The Ohio State University

Abstract

A small corpus of X-ray microbeam data was examined to test the predictions made by two well-known views of anticipatory coarticulation: time locking and feature spreading. Lip rounding and tongue backing associated with English /u/ were investigated. The token types were VCnV sequences where V = /i, u/ and C = /s, t/, consonants which are assumed to be neutral with respect to the features under study. The lip and jaw pellets were rotated in such a way that their principal component of movement was in the vertical dimension. In the /iCnu/ tokens there was no evidence for either time locking or feature spreading of the onset of lip rounding. However, other articulatory phenomena are temporally regulated. The timing of both the maximum point of lip rounding and of tongue backing was fixed relative to the acoustic onset of /u/. These results may provide evidence for the idea of fixed targets in speech production.

Introduction

Different accounts of the timing of gestures in coarticulation have been put forth. Under a time locking hypothesis, such as that proposed by Bell-Berti and Harris (1979, 1982), timing is an intrinsic part of the speech motor plan. They found that the onset of EMG activity for lip rounding begins at a fixed interval relative to the onset of the voicing of a rounded vowel. Regardless of the number of (or duration of) preceding consonants, anticipatory activity begins at a point that is temporally fixed. Under this view, units of speech production are dynamic gestures.

Under a feature spreading hypothesis, such as that first proposed in Henke (1966), timing is not specified in the segmental description, but, rather, is extrinsic to the execution of the speech motor plan. Anticipatory activity begins at a point where there are no contradictory demands on the articulators. For /u/ speakers begin the anticipatory rounding gesture at the earliest point possible. The longer the cluster of feature-neutral consonants (or the larger the number of consonants) preceding the rounded vowel, the earlier the onset of rounding with respect to the acoustic period of the vowel. The feature is said to "spread" to the beginning of the consonant string, even as early as six consonants preceding the acoustic onset of the vowel. Under this view, the units of speech production are invariant in nature.

Figure 1, adapted from Perkell (1986: 49), illustrates the difference between these views. Lip protrusion (rounding) is shown as a function of time. The onset of rounding is indicated by a tick mark. The dashed lines represent acoustic boundaries of the intervocalic consonant(s). In (A) time locking, the point of onset relative to the voicing for the rounded vowel is the same regardless of the number of consonants. In (B) feature spreading, the points of onset differ. If the rounding begins at the end of the period for the preceding unrounded vowel, then the longer the string of intervocalic consonants, the earlier the onset of rounding.

To test the predictions made by these two views of coarticulation, the anticipatory gestures associated with /u/ were examined using X-ray microbeam data. Both lip rounding and tongue backing were investigated.

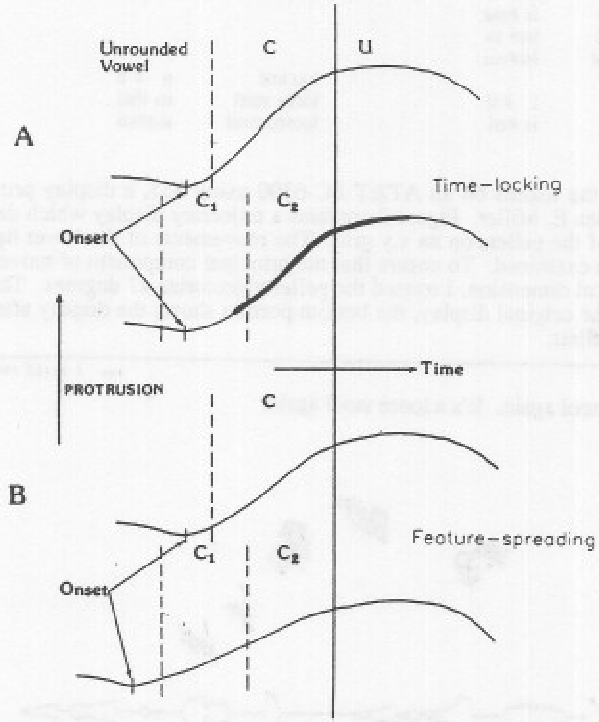


Figure 1. A comparison of time locking and feature spreading (Perkell 1986).

Methods

The tokens were taken from a set of X-ray microbeam data bases (Kiritani et al. 1975, Kiritani 1986) made under the direction of Osamu Fujimura at the University of Tokyo. The utterance types were identical to those used in Bell-Berti and Harris (1979); the data base had been made according to their specifications. The subject was a female native speaker of English. As indicated in Table 1, the token types consisted of VCnV sequences where each vowel was either /i/ or /u/. The consonants were /s/, /t/ and various combinations of the two up to a maximum of four consonants. For some token types only the placement of the word boundary differed, as in lee stool /li#stul/ and lease tool /lis#tul/. All sequences were embedded within actual English words, and were placed in the frame sentence 'It's a _______ again.' Each token type was repeated at least twice. There were thirteen /iCnu/ tokens and six of each of the other token types /iCni/, /uCni/ and /uCnu/ for a total of thirty-one tokens. Each sentence was read slowly with preservation of word boundaries, sometimes to the extent that [ ] could be heard between them, as in one reading of leased tool [list tul]. Across utterances, there was no difference in rate.

Table 1. Token types

lee tool	i # tu	loo tool	u # tu
lee stool	i #stu	loose stool	us #stu
lease tool	is # tu	loosed stool	ust#stu
lease stool	is #stu		
leased tool	ist# tu		
leased stool	ist#stu		
		loo teal	u # ti
lee teal	i # ti	loose steel	us #sti
lease steel	is #sti	loosed steel	ust#sti

I examined the tokens on an AT&T PC-6300 using XD, a display program developed by Joan E. Miller. Figure 2 contains a trajectory display which depicts the movement of the pellets on an x,y grid. The movements of the lower lip and jaw pellets were examined. To ensure that the principal component of movement was in the vertical dimension, I rotated the pellet trajectories 47 degrees. The top portion shows the original display; the bottom portion shows the display after the rotation of the pellets.

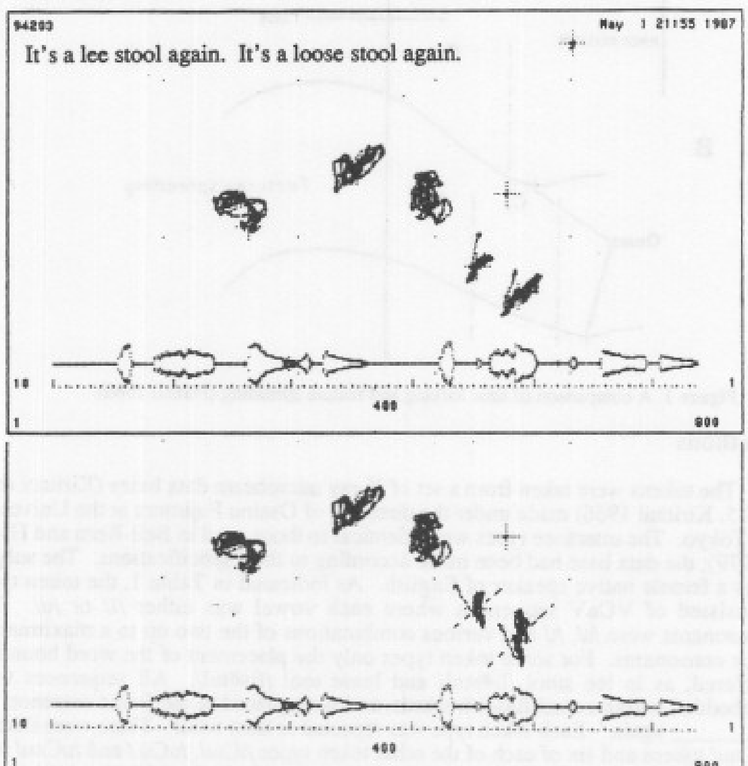


Figure 2. A trajectory display from the XD program. "Before" and "after" diagrams show the effects of rotating the lip and jaw pellets.

Figure 3 contains a trace display which depicts the vertical movement of the rotated pellets as a function of time. Since the lower lip and jaw are mechanically linked, there is a possible effect of confoundment of lip and jaw raising in the movement of the lower lip pellet. To determine the movement of the lower lip alone, I subtracted the value of jaw movement from the combined trajectory value of the jaw and lower lip movement, resulting in a trace which reflected, as much as possible, pure lip movement. In the diagram, this trace is represented as 'U1'.

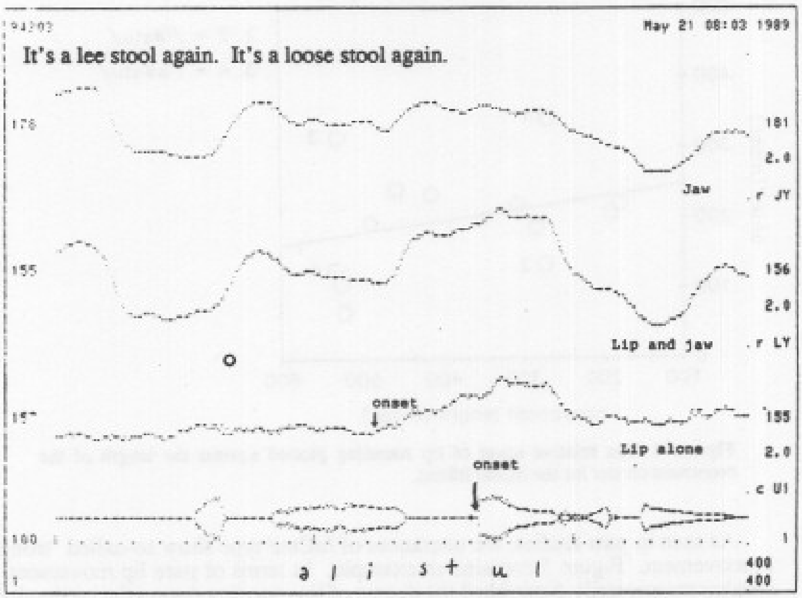


Figure 3. Traces showing the movement of the lip and jaw pellets, and a user-defined trace showing movement of the lip alone. The utterance is lee stool.

Results

Lip Rounding

Three kinds of measurement were made in the /iCnu/ tokens: the onset of lip rounding, defined as the point in u1 at which the trace begins to rise; the acoustic onset of /u/; the rounding peak within /u/; and the acoustic length of the intervocalic consonant(s). The first three are labeled in Figure 3.

In Figure 4 the onset of lip rounding is plotted against consonant length for the thirteen /iCnu/ tokens. The results showed no evidence for time locking of lip rounding. If the onset is time locked, one would expect to find values clustering within a narrow range. In this study, the values vary substantially in the vertical dimension. Moreover, the relative onsets vary in some tokens of the same consonant length and consonant type. Both /i#stu/ type tokens, labeled 1 and 2, show widely differing relative onsets (134 and 341.7 ms.) but the length of the /st/

cluster is virtually the same (273.6 and 270.0 ms. respectively). Both /is#stu/ type tokens, labeled 3 and 4, likewise show quite different relative onsets (127.3 and 314.9 ms.) but almost the same consonant length (532.8 and 529.2 ms. respectively). In a case of feature spreading, the relative onset is directly proportional to consonant length, and so we would expect a regression line of the form y=x. But the results clearly do not support this view either.

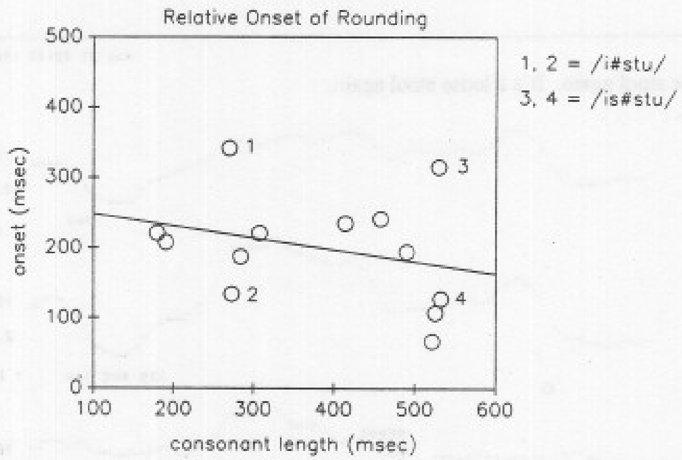


Figure 4. The relative onset of lip rounding plotted against the length of the consonant cluster for the /iCnu/ tokens.

As seen in past studies, the utterances of /uCnu/ type show so-called "troughs" in movement. Figure 5 contains an example. In terms of pure lip movement, the speaker momentarily diminished the degree of lip rounding during the production of the consonants. There were no troughs in the /utu/ type tokens. This may be due to the duration of the intervocalic consonant; the consonant was shorter in /utu/ than in the other /uCnu/ types. It would have been interesting to examine /usu/ type tokens for comparison, but since /sul/ and /ul/ are not actual English words, there were no tokens of this type.

The occurrence of troughs in the conditions specified is consistent with what has been found in earlier investigations. This is problematic for the notion of feature spreading in that there is no clear reason why lip rounding activity during this sequence should diminish during the production of consonants compatible with (and presumably) neutral for lip rounding.

Although no connection between rounding onset and acoustic onset can be seen here, there is a relationship between the acoustic onset of the vowel and the peak value of lip rounding. The peak value was the first cursor position at which the trace reached its maximum value during the production of /u/, as shown in Figure 6. Only twelve /iCnu/ tokens were examined. In one token, the timing of the peak could not be determined. The trajectory display ended shortly after the acoustic onset of the /u/, making it impossible to identify the point of maximum protrusion.

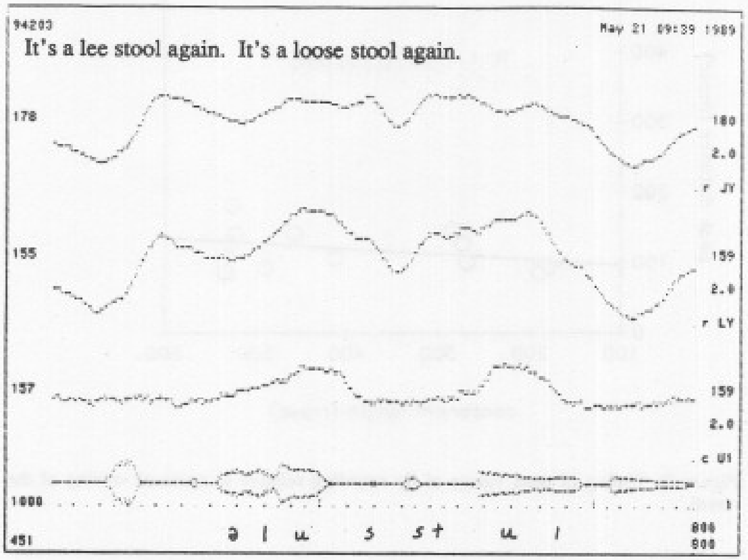


Figure 5. A "trough" in a /uCnu/ utterance.

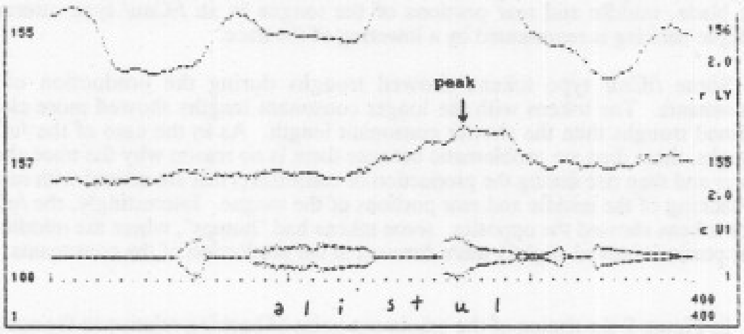


Figure 6. Measuring the peak of lip rounding.

As Figure 7 shows, there is no relationship between the peak distance and consonant length. The values of the peak distance range from 87.1 ms. to 180.9 ms. following the onset of voicing. The values generally fall within a narrow range; eleven of the twelve tokens are between 80-140 ms. This suggests that the occurrence of the peak is regulated relative to the acoustic onset of the vowel. The timing of the onset of lip rounding does not matter, as long as the maximum amplitude is reached at a relatively fixed time.

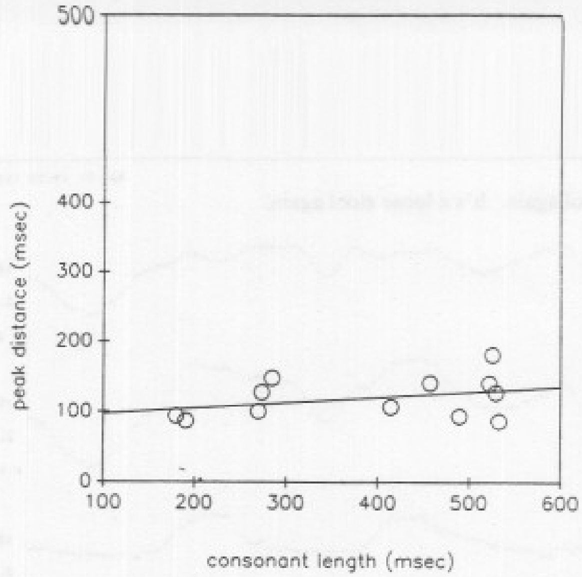


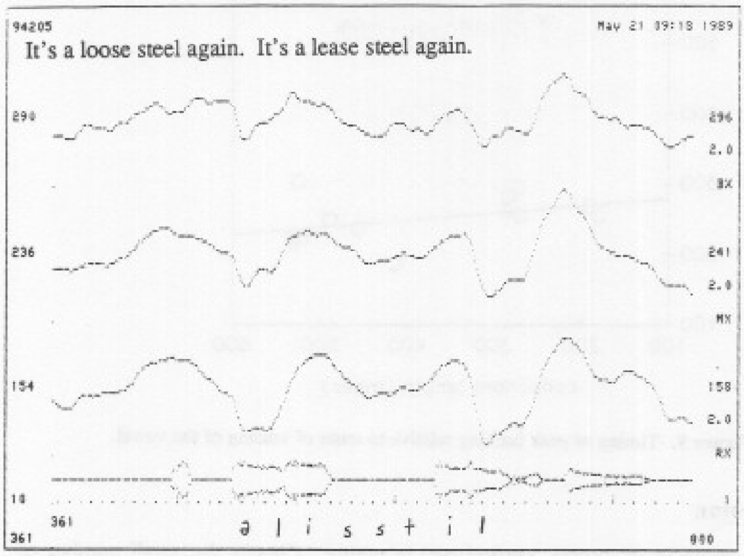
Figure 7. Timing of peak values of lip rounding relative to onset of voicing of the vowel.

Tongue Backing

Because the maximal point of lip rounding for /u/ was found to be "time locked", the question arises as to whether the same phenomenon might be seen in tongue backing as well. Figure 8 contains traces showing horizontal movement of the blade, middle and rear portions of the tongue in an /iCnu/ type utterance. Tongue backing is represented by a lowering of the trace.

Some /iCni/ type tokens showed troughs during the production of the consonants. The tokens with the longer consonant lengths showed more clearly defined troughs than the shorter consonant length. As in the case of the /uCnu/ troughs, these dips are problematic because there is no reason why the trace should lower and then rise during the production of consonants that are neutral with respect to backing of the middle and rear portions of the tongue. Interestingly, the /uCnu/ type tokens showed the opposite: some tokens had "humps", where the middle and rear portions moved slightly more forward in the production of the consonants.

In Figure 9 the timing of the maximum point of backing relative to the acoustic onset of /u/ is plotted against consonant length. As mentioned earlier, one of the tokens could not be analyzed, and only twelve were used. The maximum point was defined as the first position at which the trace was at its lowest. For all /iCnu/ tokens, the maximum point of backing is temporally fixed following the acoustic onset. The values have no connection with consonant length, and fall within a narrow range. The minimum is 187.6 ms. following the onset of voicing; the maximum is 301.5 ms. Eleven of the twelve tokens fall within the range of 210-300 ms. Thus, just as in rounding, the timing of maximal tongue backing was also consistent.



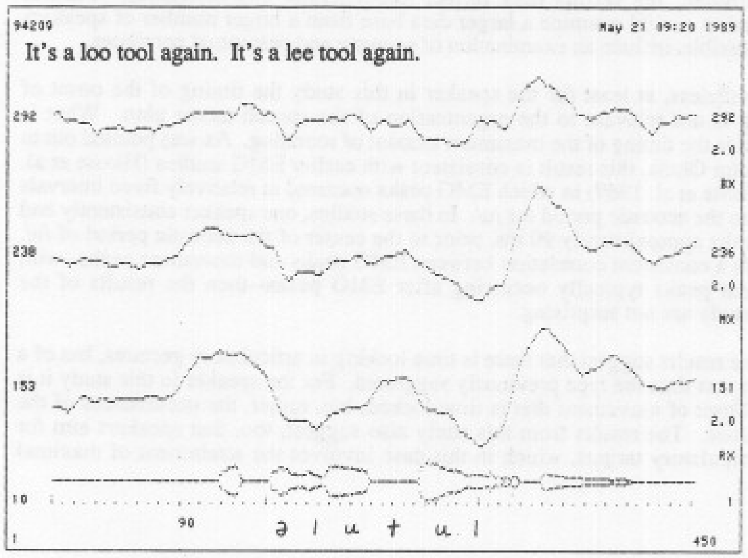


Figure 8. Movement traces of tongue backing in the utterances *lease steel* (top) and *loo tool* (bottom).

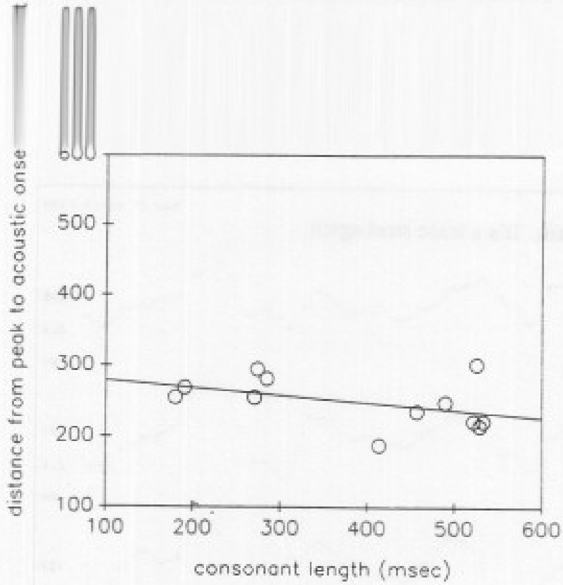


Figure 9. Timing of peak backing relative to onset of voicing of the vowel.

Conclusion

This study is, of course, preliminary in nature. Clearly the small number of tokens makes it difficult to form strong generalizations. Since only one speaker was examined, the results may reflect idiosyncratic characteristics. Further investigation should examine a larger data base from a larger number of speakers and, if possible, include an examination of acoustic and perceptual correlates.

Nonetheless, at least for the speaker in this study the timing of the onset of rounding is not relevant to the organization of the speech motor plan. What is regulated is the timing of the maximum amount of rounding. As was pointed out to me by John Ohala, this result is consistent with earlier EMG studies (Hirose et al. 1968, Hirose et al. 1969) in which EMG peaks occurred at relatively fixed intervals relative to the acoustic period for /u/. In these studies, one speaker consistently had EMG peaks approximately 90 ms. prior to the center of the acoustic period of /u/. If there is a consistent correlation between EMG peaks and movement peaks--with movement peaks typically occurring after EMG peaks--then the results of the current study are not surprising.

These results suggest that there is time locking in articulatory gestures, but of a different sort than the type previously suggested. For the speaker in this study it is not the onset of movement that is time locked, but, rather, the occurrences of the peak values. The results from this study also suggest, too, that speakers aim for fixed articulatory targets, which in this case involves the attainment of maximal points.

Acknowledgments

Earlier versions of this paper were presented at the 113th meeting of the Acoustical Society of America in Indianapolis, IN (1987) and at the 117th meeting in Syracuse, NY (1989), where I received many useful comments. I gained much from discussion with Mary Beckman and other participants in the seminar on articulation and coarticulation. I am also grateful to Jan Edwards for her suggestions.

References

JASA = Journal of the Acoustical Society of America

Bell-Berti, Fredericka and Katherine S. Harris. 1974. RMore on the motor organization of speech gesturesS, *Haskins Laboratories: Status Report on Speech Research* SR 37/38: 73-77.

__________. 1979. RAnticipatory co-articulation: some implications from a study of lip roundingS, *JASA* 65.3: 1268-1270.

__________. 1981. RA temporal model of speech productionS, *Phonetica* 38: 9-20.

__________. 1982. RTemporal patterns of coarticulation: lip roundingS, *JASA* 71.2: 449-454.

Fowler, Carol A. 1980. RCoarticulation and theories of extrinsic timingS, *Journal of Phonetics* 8: 113-133.

Gelfer, Carole E., Katherine S. Harris and Gary Hilt. 1981. RTemporal constraints on anticipatory coarticulationS, Abstract, *JASA Supplement 1*, Vol. 69, Spring: S55.

Gelfer, Carole E., Fredericka Bell-Berti and Katherine S. Harris. 1985. RDetermining the extent of coarticulation: effects of experimental designS, *Haskins Laboratories: Status Report on Speech Research* SR-82/83: 19-31.

Harris, Katherine S., Betty Tuller and J.A. Scott Kelso. 1986. RTemporal invariance in the production of speechS, *Invariance and Variability in Speech Processes*, Joseph S. Perkell and Dennis H. Klatt (eds.): 243-252. Hillsdale, N.J.: Lawrence Earlbaum Associates.

Henke, William L. 1966. *Dynamic articulatory model of speech production using computer simulation*. Ph.D. dissertation, MIT.

Hirose, H., S. Kiritani and S. Shibata. 1968. RAn electromyographic study of articulatory movementsS, *Annual Bulletin No. 2, Research Institute of Logopedics and Phoniatrics*, Univ. of Tokyo.

Hirose, H., Z. Simada and S. Kiritani. 1969. RAn electromyographic study of articulatory movementsS, *Annual Bulletin No. 3, Research Institute of Logopedics and Phoniatrics*, Univ. of Tokyo.

Kiritani, S., K. Itoh and O. Fujimura. 1975. RTongue-pellet tracking by a computer-controlled X-ray microbeam systemS, *JASA* 57.6.II: 1516-1520.

Kiritani, Shigeru. 1986. RX-ray microbeam method for measurement of articulatory dynamics-techniques and resultsS, *Speech Communication* 5: 119-140.

Lubker, J. and Thomas Gay. 1982. RAnticipatory labial coarticulation: experimental, biological, and linguistic variablesS, *JASA* 71.2: 437-448.

Perkell, Joseph S. 1980. RPhonetic features and the physiology of speech productionS, *Language Production* . Brian Butterworth (ed.), Vol. 1: 337-372. Academic Press.

______. 1986. RCoarticulation strategies: preliminary implications of a detailed analysis of lower lip protrusion movementsS, *Speech Communication* 5: 47-68. North Holland.

Sussman, Harvey M. and John R. Westbury. 1981. RThe effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulationS, *Journal of Speech and Hearing Research* 46: 16-24.

Prosody and intrasyllabic timing in French

Janet Fletcher [1] *and Eric Vatikiotis-Bateson* [2]

[1]Speech, Hearing, and Language Research Centre, Macquarie University
[2]ATR Auditory and Visual Perception Research Laboratories

Abstract

Durational variation associated with accentuation and final lengthening is examined in a corpus of articulatory data for French. Both factors are associated with measurable differences in acoustic duration. However two different articulatory strategies are employed to make these contrasts although both result in superficially longer and more displaced gestures.

Introduction

The importance of prosodic organization in speech timing has been acknowledged for a number of years. Stress and accentuation are known to affect the acoustic durations of syllables in a number of languages. For example, in French, syllables with right-boundary tonic accents or stresses and non-emphatic initial accents are longer than unaccented syllables. The basic rhythmic group or prosodic word is defined by right boundary tonic accents, and not by left boundary prominences as in English stress feet. On the other hand, like English, French displays final lengthening. Accented syllables at the right edge of phrases Q units consisting of one or more prosodic words Q are significantly longer than accented syllables interior to the phrase (Crompton 1980, Touati 1987, Fletcher 1990). Intonational analyses (e.g., Martin 1987) would suggest that phrase-final accents are more prominent than phrase-internal accents. It is often assumed, therefore, that acoustic duration is an important cue to these prosodic relations, and that final lengthening is part of the same linguistic process as accentuation.

The articulatory correlates of these high-level prosodic relations are not that well studied in either French or English. Most studies have focused on the relationship between two categories of prosodic strength - stressed and unstressed, and have usually excluded final syllables. One recent study has attempted to redress the situation for English. Edwards, Beckman, and Fletcher (1991) compared lip and jaw articulation in phrase-final accented, and non-final pitch accented and reduced syllables. They found distinct differences in the ways in which accentuation and final lengthening were realized in their corpus and concluded that, a) the kinematic patterns associated with accentuation and final lengthening reflect different underlying articulatory maneuvers, and b) final lengthening does not necessarily belong to the same kind of phonological process that governs prominence relationships per se, but to some kind of local timing process.

The phrasal phonology of French is somewhat different from English in that final syllables are always accented in the former but only optionally accented in the latter. It is of interest, therefore, to see whether the duration increase in phrase-final accented syllables in French (final lengthening) is associated with similar qualitative patterns of articulator motion to those of the phrase-internal accented/unaccented contrast. In an earlier study, Vatikiotis-Bateson (1988) found consistent differences in the kinematic parameters, duration, peak velocity and articulator displacement in the opening and closing gestures of non-final accented and unaccented syllables in French. Accented syllables were associated with bigger as well as longer gestures as would befit a prominence contrast. Final syllables were excluded from the original study. In the present investigation, part of the French corpus recorded for Bateson's earlier study was reanalyzed to look specifically at the final / nonfinal contrast.

Methods

The model sentence, *"L'interet, qui aveugle les uns, fait la lumiere des autres."* ["The curiosity that blinds some, illuminates otheres"] was analyzed.

Three speakers of standard French (without traces of strong regional accents) produced ten repetitions of the sentence using the syllables /ba/ and /ma/ at two self-selected tempi; conversational and fast. A prosodic analysis of each sentence repetition was performed. The speakers consistently recited the sentence with three prosodic phrases, usually pausing after each phrase. All phrase-internal accents were classified as one type, regardless of whether they were right boundary tonic accents or phrase-initial accents, since there was no discernible difference in the movement amplitudes associated with either accent type. The comparison between this group and the opening gestures of phrase-final syllables constitutes the basis of the final/nonfinal contrast in this study. The unaccented/accented contrast is also based on the comparison between this group and the gestures of unaccented syllables.

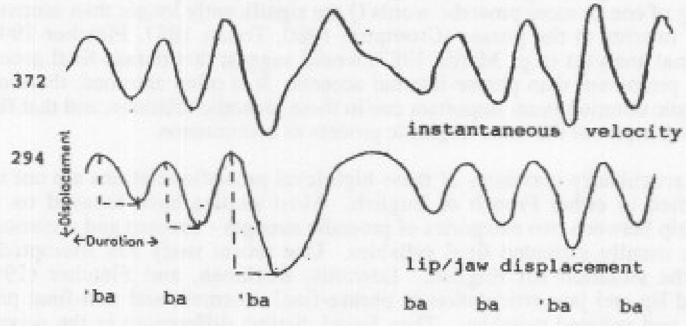


Figure 1. Selspot trace representing lip / jaw position and instantaneous velocity for part of a reiterant version of the sentence: "L'interet qui aveugle les uns, fait the lumiere des autres".

Vertical movements of the lower lip/jaw complex were recorded using the optoelectronic SELSPOT system at Haskins Laboratories. The position files were numerically differentiated to obtain instantaneous velocity. For position of the lower lip/jaw, (Figure 1) peaks and valleys of the movement trace correspond to

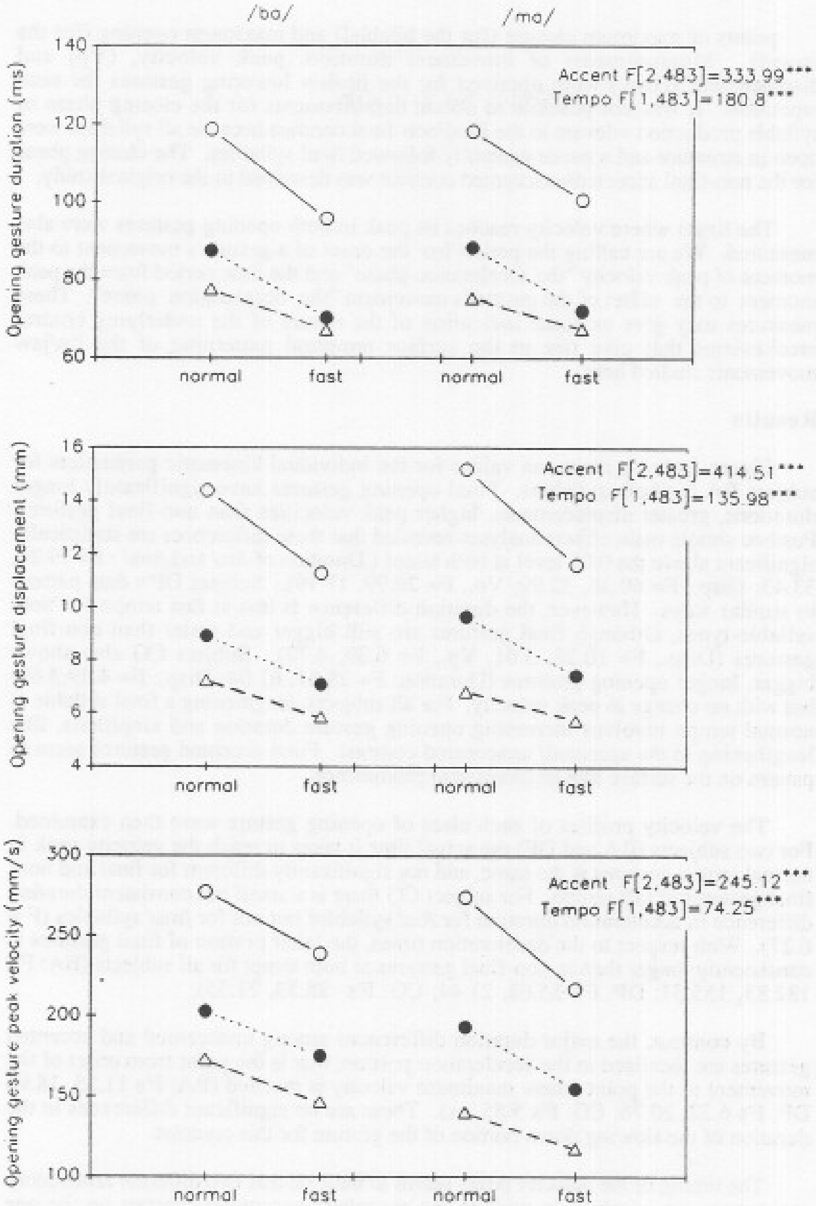


Figure 2. Mean gesture durations, displacements, and peak velocities, for normal and fast tempo /ba/ and /ma/ opening gestures, contrasted for accent and position in utterance. (Subject BA)

points of maximum closure (for the bilabial) and maximum opening (for the vowel). Measurements of movement duration, peak velocity, (Vp) and displacement (Disp.) were obtained for the lip/jaw lowering gestures for each repetition. It was not possible to obtain measurements for the closing phase of syllable production relevant to the final/non-final contrast because all syllables were open in structure and a pause generally followed final syllables. The closing phase for the non-final accented/unaccented contrast was described in the original study.

The times where velocity reaches its peak in both opening gestures were also measured. We are calling the period fro the onset of a gesture's movement to the moment of peak velocity "the acceleration phase" and the time period from the peak moment to the offset of the gesture's movement "the deceleration phase". These measures may give us some indication of the nature of the underlying control mechanisms that give rise to the surface temporal patterning of the lip/jaw movements studied here.

Results

Figure 2 shows the mean values for the individual kinematic parameters for subject BA for both syllables. Final opening gestures have significantly longer durations, greater displacements, higher peak velocities than non-final gestures. Posthoc simple main effects analyses revealed that these differences are statistically significant above the 0.01 level at both tempi (Duration of /ba/ and /ma/ : F= 79.24, 33.43; Disp., F= 60.36, 32.99; Vp., F= 29.99, 17.19). Subject DP's data pattern in similar ways. However, the duration difference is lost at fast tempo for both syllable-types, although final gestures are still bigger and faster than non-final gestures (Disp., F= 10.28, 3.61, Vp., F= 6.39, 4.72). Subject CG also shows bigger, longer opening gestures (Duration: F= 28.81, 61.04; Disp.: F= 4.89,3.69) but with no change in peak velocity. For all subjects, lengthening a final syllable at normal tempo involves increasing opening gesture duration and amplitude, like lengthening in the accented/ unaccented contrast. Final accented gestures seem to pattern on the surface like an increase in prominence.

The velocity profiles of each class of opening gesture were then examined. For two subjects (BA and DP) the actual time it takes to reach the velocity peak in normal tempo gestures is the same, and not significantly different for final and non-final gestures at fast tempo. For subject CG there is a small but consistent duration difference in acceleration duration for /ba/ syllables but not for /ma/ syllables (F = 6.27). With respect to the deceleration times, the latter portion of final gestures is consistently longer than in non-final gestures at both tempi for all subjects (BA: F's 182.83, 155.37; DP: F's 35.02, 21.44; CG: F's 28.53, 21.35).

By contrast, the major duration differences among unaccented and accented gestures are localized in the acceleration portion, that is the point from onset of the movement to the point where maximum velocity is reached (BA: F's 11.35, 18.96 DP: F's 6.22, 20.76; CG: F's 5.55, ns). There are no significant differences in the duration of the slowing down portion of the gesture for this contrast.

The timing of the velocity peaks seems to indicate that two different articulatory maneuvers are involved in making the accented/unaccented contrast on the one hand, and the final/non-final contrast, on the other. The earlier timing of the peaks, together with the smaller observed displacements in unaccented gestures, might indicate that the intention is to produce smaller gestures in unaccented syllables and bigger gestures in accented syllables. In other words the gestures in either case are different in shape and magnitude, suggesting different underlying dynamic parameters - such as force/stiffness and movement amplitude if we were to model these movements in a linear mass-spring model framework (Kelso et al 1985, Vatikiotis-Bateson (1988), Edwards et al. 1991). The small timing difference among accented and unaccented opening gestures falls out directly from the intragestural dynamics.

Another kind of articulatory maneuver is potentially involved in the production of the final/non-final contrast. The shorter, less displaced gestures of non-final syllables could be the result of a change in intergestural timing. The onset of the following opening gesture associated with the upcoming syllable in the phrase truncates the non-final closing gesture. This seems plausible given hat the bulk of the timing difference between final and nonfinal opening gestures is localized in the tail end of final gestures. The gestures for accented final and nonfinal syllables have the same basic shape until the point where peak velocity is reached -- presumably somewhat beyond this point. Non-final gestures are cut short but final gestures are not. Moreover, the magnitude of difference in overall gesture duration is far greater in the final/nonfinal contrast than in the unaccented/accented contrast.

Conclusion

In summary, accentual lengthening and final lengthening are associated with different articulatory maneuvers in this corpus of French. Nonfinal accented and unaccented gestures are probably associated with different underlying intragestural parameter settings (i.e., underlying amplitude, force or stiffness) as suggested in the original study. Final accented and nonfinal accented gestures may or may not have the same intragestural specifications, but observed displacements and associated timing patterns are mainly the product of changes in intergestural timing. Phrase-internal accentuation represents a true change in prosodic prominence in that the primary intention is to produce a bigger syllable. By contrast, final lengthening in French may be more of a targeted durational contrast as suggested by Edwards et al. for English. That is, the intention might be to produce a longer as well as a more prominent syllable. These results support Edwards et al.'s conclusion that acoustic duration is not a sensitive cue to these qualitative differences in linguistic timing.

Acknowledgements

Parts of this research were supported by the National Science Foundation (USA) under Grant no. IRI- 8858109 to Mary Beckman, the Ohio State University, and by the National Institutes of Health (USA) under Grant no. NS-13617 to Haskins Laboratories.

References

Crompton, A. 1980. RTiming patterns in FrenchS *Phonetica* 37: 205-234.

Edwards, J., Beckman, M.E.and Fletcher, J. 1991. RThe articulatory kinematics of final lengtheningS, *Journal of the Acoustical Society of America*, 89.1: 369-382.

Fletcher, J. 1990. RRhythm and final lengthening in FrenchS, *Journal of Phonetics* 18.4.

Kelso, J.A.S., Vatikiotis-Bateson, E., Saltzman, E. L., and Kay, B. A. 1985. RA qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling", *Journal of the Acoustical. Society of America*, 77: 266-280.

Martin, Ph. 1987. "Prosodic and rhythmic structures in French", *Linguistics*, 25: 925-950.

Touati, P. 1987. "Structures prosodiques du suédois et du français", *Travaux de l'institut de linguistique de Lund.* 21

Vatikiotis-Bateson, E. 1988. *Linguistic structure and articulatory dynamics.* Indiana University Linguistics Club.

Lip rounding and vowel formant frequencies in Nantong Chinese

Benjamin Ao

Department of Linguistics, The Ohio State University

Abstract

A study of the vowel system of Nantong Chinese, which has as many as seven high vowels, suggests that instead of the traditional two formant model, a three formant model is needed to approximate the Nantong vowel space. Of the three formants used in the model, F3 appears to be the major acoustic cue of lip rounding. It also appears that the "maximum dispersion" hypothesis is not true with regard to the Nantong vowel system.

Introduction

In the Nantong dialect of Chinese spoken on the northern Yangtze Delta, there are 13 oral vowels, as follows:

i y ʒ̩ ʒ̹̩ z̩ β̩ u

e ɜ o

ɛ ɔ

a

Of the seven high vowels, four sound like syllabic fricatives represented by the IPA symbols used: [ʒ̩] is a syllabic [ʒ], [ʒ̹̩] a syllabic [ʒ̹] (the rounded version of [ʒ]), [z̩] a syllabic [z], and [β̩] a syllabic [β]. Ignoring [z̩], the six other high vowels can be divided into three pairs (front, central and back), each contrasting in rounding.

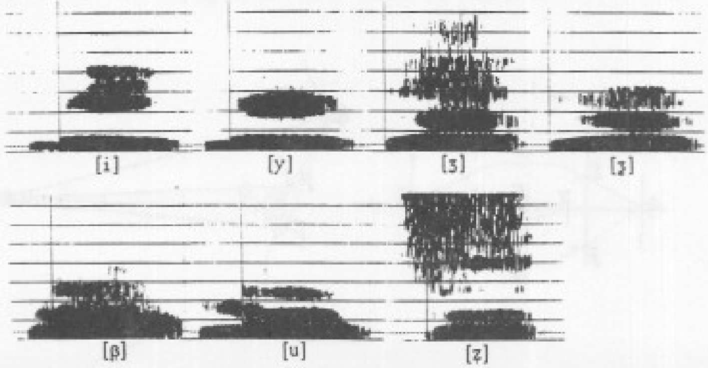


Figure 1. Spectrograms of the seven Nantong high vowels.

A number of questions can be raised about the Nantong vowel system. First of all, is it possible to approximate the crowded Nantong vowel space, especially that of the high vowels, with the traditional two formant model? Secondly, what is the acoustic effect of lip rounding, particularly on the high vowels? Finally, does the "maximum dispersion" hypothesis a la Liljencrants & Lindblom (1972) hold true to the Nantong vowel system, especially in the high vowel area? To answer these questions, I conducted an experiment, which will be explained below.

Experiment

Method

<u>Subjects</u>

Three native speakers of Nantong Chinese, two males (AX and ST) and one female (LP), all in their 20's or 30's, were asked to read out the syllable list.

<u>Material</u>

A list of all the (131) phonotactically possible CV syllables with all the 13 vowels was used as the experiment material. The sequence of syllables was randomized.

<u>Procedure</u>

Video and audio signals of the subjects reading the syllable list were simultaneously recorded with a camcorder placed two feet in front of the mouth of the speaker. A large mirror was placed beside the speaker's head at a 45 degree angle to allow the sideview of the lip movement to be also recorded. A cushion was placed between the speaker's head and the wall behind to prevent any head movement during the recording. The audio recording was fed into a Kay 5500 speech analyzer, and formant frequency values were taken from the average power spectrum across the entire vowel duration.

To quantify lip rounding movements, 11 measurements between geometrically defined points (CF, CE, CD, BF, BE, GJ, GI, PS, LM, LN and LO, see Fig 2) were taken from the playback of the video recording on a 19" TV screen.

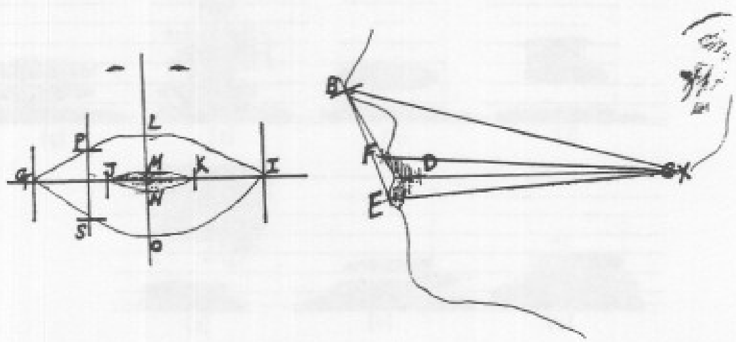


Figure 2. Reference points for lip gesture measurements

Analysis

The traditional two formant (F1 by F2) model for the vowel space proves insufficient for the Nantong vowel system. Overlap between different vowels exists for all three speakers. Figure 3 shows this.

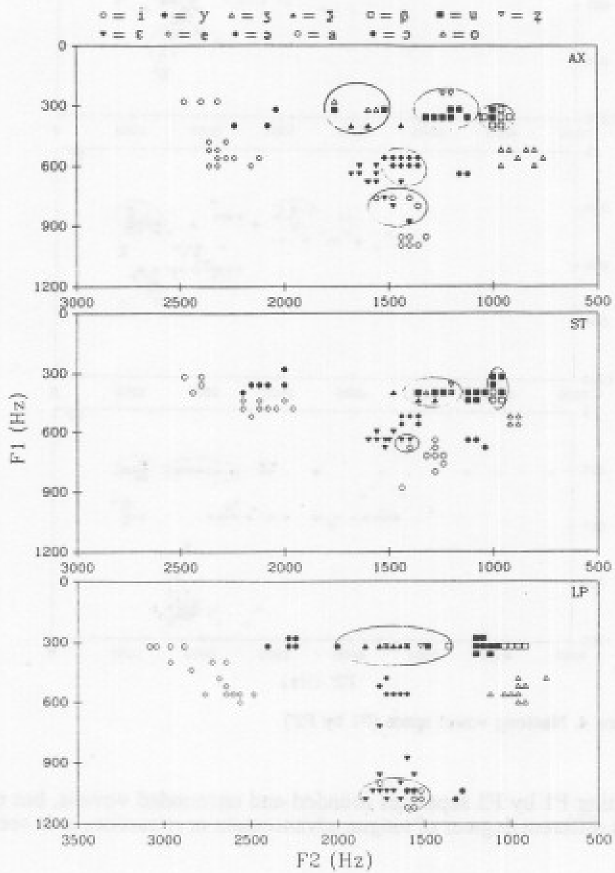


Figure 3. Nantong vowel space in F1 x F2 for three speakers; major areas of overlapping are circled.

Little improvement was achieved by substituting F2 with the weighed average F2' calculated according to the formula proposed by Bladon (1983) as in : $F2' = F2 + C^2 \times (F3 \times F4)^{1/2} / (1+ C^2)$, where $C = [12 \times F2 \times 67 \times F2(1 - F1^2 / F2^2) \times (1 - F2^2 / F3^2) \times (1 - F2^2 / F4^2)] \times [1400 \times (F4 - F3)^2 \times (F3 \times F4 / F2^2 - 1)]^{-1}$. Figure 4 shows Nantong vowel space.

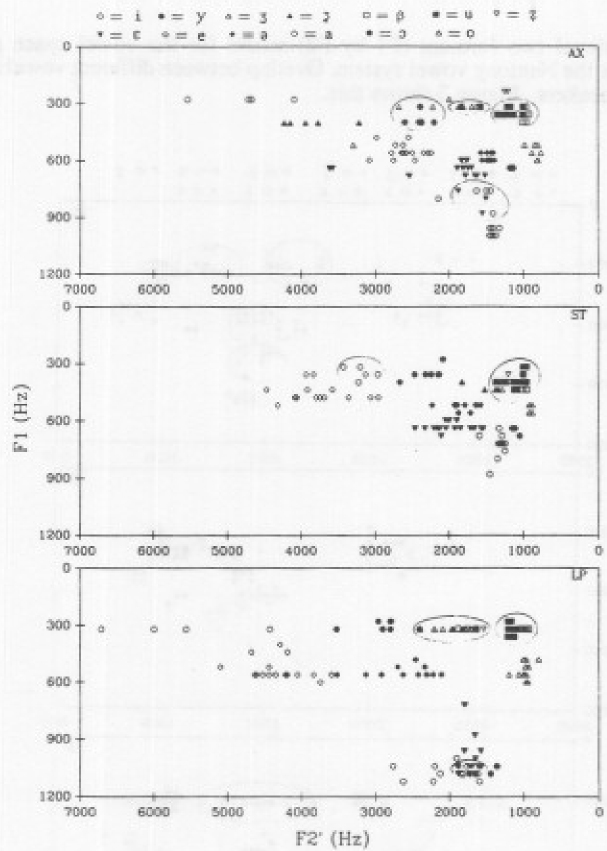


Figure 4. Nantong vowel space (F1 by F2')

Plotting F1 by F3 separates rounded and unrounded vowels, but overlaps vowels with different degrees of tongue advancement or retraction, as is seen in Figure 5.

The best separation of the seven high vowels is obtained by plotting F2 against F3, as shown in Figure 6, where like vowels are encircled. This suggests that a three dimensional model with F1, F2 and F3 as coordinates is necessary for approximating the entire Nantong vowel space, with F1 for vowel height and F2 for tongue advancement.

In order to justify adopting a three formant model instead of the traditional two formant model of vowel space, with F3 as the additional dimension, efforts were made to find out the underlying articulatory parameter responsible for the variation of F3. An ANCOVA test reveals that the single most significant ($p < .01$)

articulatory measurement that covaries with F3 is the horizontal width of horizontal lip opening (JK). Figure 7 shows the mean JK value of the seven high vowels for all three speakers.

Figure 8 shows how JK is correlated with F3. One may notice that there is a big difference between the F3 values of the two members of the rounded and unrounded pairs of front vowels and central vowels, but the difference between the F3 values of the rounded and unrounded back vowels is minimal with speakers ST and LP. In addition to the width of horizontal lip opening, speaker identity has a significant effect on F3 as well ($p < .01$), and lip protrusion (BF, CF) is also a significant covariate (BF and CF, $p < .05$).

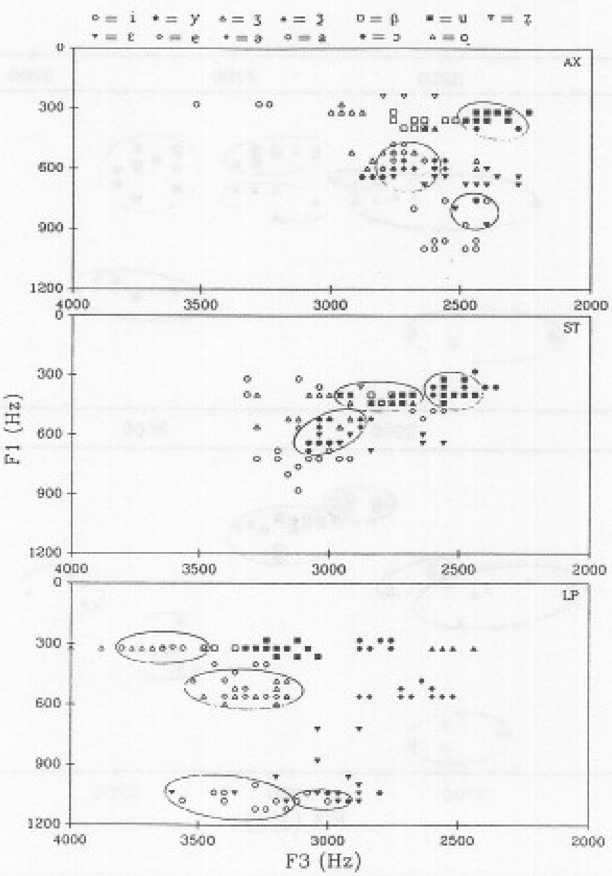


Figure 5. Nantong vowel space (F1 by F3)

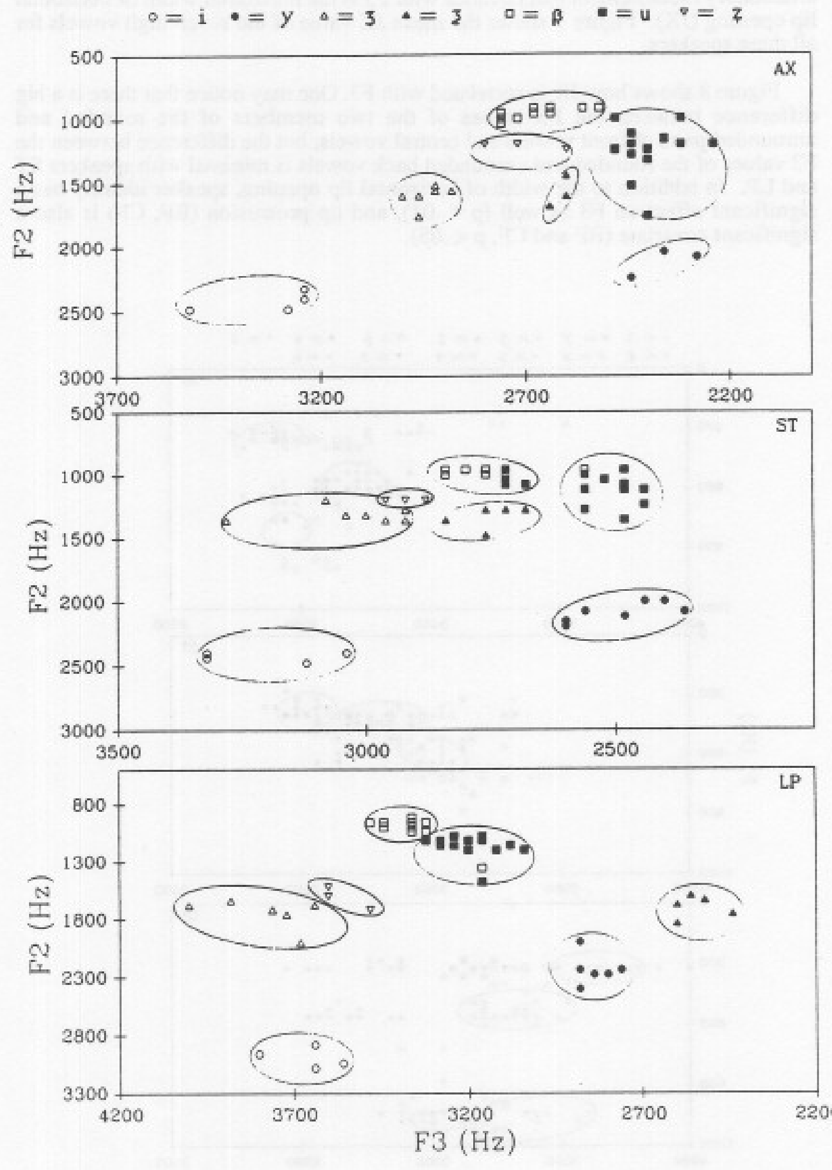


Figure 6. Nantong vowel space for high vowels (F2 by F3)

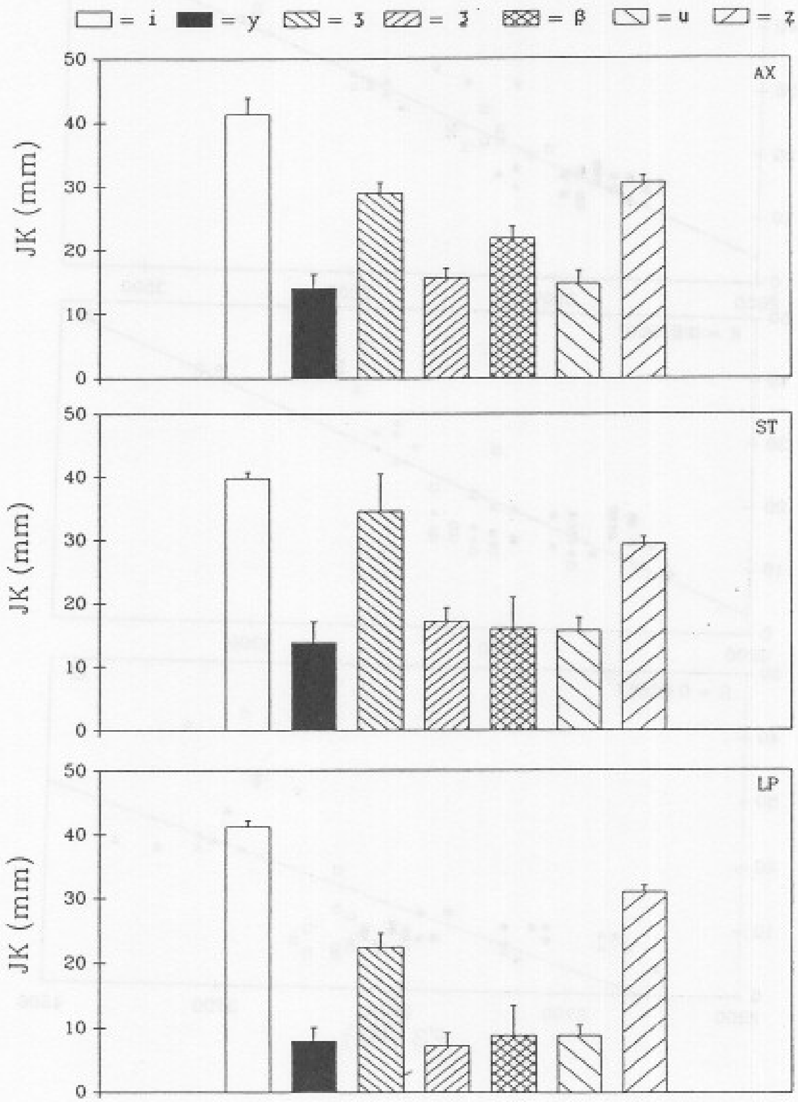


Figure 7. Width of lip opening of the Nantong high vowels

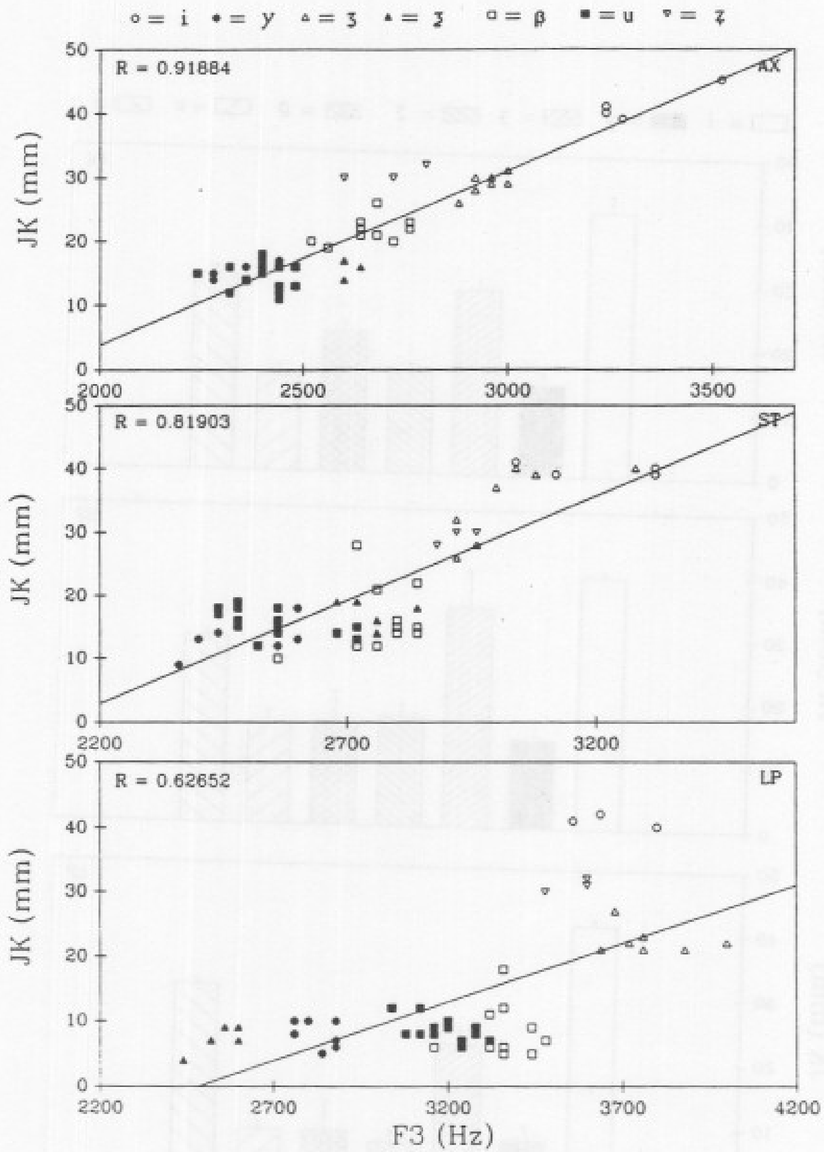


Figure 8. Correlation between F3 and width of lip opening

Discussion

Now we may return to the questions raised at the beginning of this paper. First of all, it is clear by now that the traditional two formant model of vowel space is not universally adequate, for it fails to represent the crowded Nantong vowel space. Instead, a three formant model, with F3 as the third dimension, is needed.

Secondly, it also appears that while tongue height is responsible for F1 and tongue advancement for F2, lip rounding accounts for much of the variation of F3, especially with front and central vowels.

Finally, the massive overlapping of plotted vowel formant frequencies that coexists with vacant areas in a two dimensional model of the Nantong vowel space is contradictory to the maximum dispersion hypothesis. Even in a three dimensional model where overlapping of plotted vowel formant frequencies is by and large eliminated, the distances between the centers of plotted formant frequencies of various vowel are largely unequal. This suggests that the maximum dispersion hypothesis is not universally true.

Acknowledgement

My heartfelt thanks go to Michel Jackson of the Division of Speech and Hearing Science at the Ohio State University. Without his generous help, I could not have completed this project. I am also grateful to the Division of Speech and Hearing Science for allowing me to use their facilities for part of my measurement and computation.

*The original version of this paper was presented in the 119th meeting of the Acoustical Society of America, State College, Pennsylvania, 21-25 May 1990.

References

Bladon, Anthony (1983) "Two-formant model of vowel perception: shortcomings and enhancements". *Speech Communication* 2: 305-13.

Liljencrants, Johan and Bjørn Lindblom (1972) "Numerical solution of vowel quality systems: the role of perceptual contrast". *Language* 48.4: 839-62.

Comparison of lip rounding in German and English Vowels

Monica Crabtree and Claudia Kurz

Department of Linguistics, The Ohio State University

Abstract

This paper describes an experiment which was designed and conducted to test three hypotheses: first, that German back vowels are more rounded than their English counterparts, second, that the degree or type of lip rounding of a given vowel or set of vowels varies with different consonantal contexts, and third, that German back vowels are more rounded than front rounded vowels. The results of the Analysis of Variance suggests strongly that the first claim is true. No contextual variation was detected. The results also give some indication that the third claim is true.

Introduction

Standard German has vowels that English does not have, namely the front rounded vowels [y] and [?] (as well as short (or lax) counterparts) and a long low back vowel. English has a low front vowel which the Standard dialect of German does not (for most speakers). However, German and English are generally considered to share many vowels, that is, similar vowels in English and German are transcribed with the same symbol. For example, the vowel in "Bude" [budη] is transcribed with an [u] as is the English vowel in "booed" [bud]. Although these German and English vowels are transcribed the same, they do not sound the same to many ears. Thus, for example, if a native English speaker produces German words, the vowels might not sound "quite right" to a native German speaker, and vice versa.

The overall vowel quality of vowels in the two languages might be similar enough to warrant their being transcribed with the same symbol, but other details must account for the lack of total identity. One possibility is that German vowels might have slightly more (or less) extreme formant values (i.e., tongue positions would reach more (or less) towards the periphery of the vowel space). While this would have been interesting to pursue, we were more interested in collecting our own data, and we wanted to approach this problem from an articulatory perspective and not an acoustic one, so we considered another possibility.

It has been suggested that German back vowels are more rounded than their English counterparts (Disner 1983 quotes some sources). If this were true, it could account in part for the subtle differences in the vowel qualities of German and English vowels. Since a study of lip rounding would lend itself to collecting data with available equipment (a camcorder, a VCR), we decided to test the hypothesis.

It was also suggested to us that German and English vowels might behave differently from one another in similar contexts. This could also account for the perception that a German vowel spoken by a native English speaker was different from that produced by a native German speaker. It has been claimed that languages with more vowels will manifest less contextual variation than languages with fewer vowels [Keating and Huffman, 1984]. Keating and Huffman suggest that languages with fewer vowels vary more to fill up the "empty" vowel space. One might also imagine that it would be more important for a vowel to reach its articulatory goal in a language with a more crowded vowel space, since if it were to vary more, it would overlap with some other vowel's space. Thus German would define a more crowded vowel space with less room for variation of a given vowel with different contexts than would English since German has four more (monophthongal) vowels than English. This contextual variation could conceivably be realized in the degree or type of lip rounding accompanying the vowel. We therefore decided to test the hypothesis that lip rounding in German vowels varied less in the context of bilabials than did that of English vowels.

Since we were measuring lip rounding, we considered testing one other claim related to this topic. Wood (1986 p.392) predicts that German front rounded vowels must be less rounded than back vowels of like height, so that an acoustically stable "quantal region" is created. With more extreme rounding, Wood's vocal tract models predict that minor variations in the positioning of the tongue in the pre-palatal region where [y] is formed will cause large changes in the second formant frequency associated with the vowel (thus yielding a vowel of inconsistent vowel quality). However, with moderate lip rounding, relatively large variations in tongue position in this region would not affect the second formant. Wood quoted only a single source for the empirical evidence to support his claim—"private communication from Eli Fischer-Jorgensen". We therefore decided to test this as well.

Method

Subjects

Three native English speakers and three native German speakers volunteered to be videotaped as they produced words of their native language which they read as they were presented, one at a time, on index cards. Subjects were seated against a wall and a mirror was placed at a 45 angle from the wall so that a view of the side of the face appeared in the reflection.

Material and Procedure

1 x 1 cm graph paper was hung on the walls behind the subject and opposite the mirror, so that all measurements of a given frame on the VCR screen could be converted to true millimeters.

Subjects were instructed to pronounce the words in a relaxed, normal way. They were also asked to keep their heads as still as possible, and a thick book was placed on the back of the chair so that they could press against it to help immobilize their heads. German speakers read a total of 240 words and English speakers read 160. These totals broken down consist of:

Language	vowels	bilabial context	other contexts
German	all vowels (14 + Käse & bäte)	8 tokens	8 tokens
English	all vowels (10)	8 tokens	8 tokens

For this study only 6 tokens of the 8 available tokens were measured and only the high and mid tense vowels were considered.

The wordlists used are given below in Table I To minimize future measuring time, we chose to precede vowels with only two contexts, and we therefore chose what we considered to be the two most extreme consonantal contexts, that is, one which might be expected to affect the lip rounding in the vowel greatly (bilabials) and one which might be expected to exert little influence on this gesture (velar). We determined to use only real words since subjects would be getting no practice and we wanted no confusion about how something ought to be pronounced in the middle of the taping session. We therefore had to substitute on occasion an alveolar consonant for a velar in the German list. Our preference would have been to use words of CV structure; however, we considered it more important to use words for like pairs in German and English which resembled each other as closely as possible. An alveolar final consonant was selected, preferably voiced, so that the vowel would be longer. Some other minor exceptions were made.

Table I. German and English Wordlists

German vowels

<u>bilabial environment</u>			<u>Other</u>		
biete	müde	Bude	gieß	Güte	gut
bitte	Bütt	Butt	Tick	Küß'	Kuß
bete	Böte	Boot	geh'	töte	Tod
Bett	Pötte	Pott	Keß	Götter	Gott
bäte		bat	Käse		Tat
		Patt			Kacke

English vowels

<u>bilabial environment</u>		<u>Other</u>	
bead	booed	keyed	cooed
bid	put	kid	could
bade	bode	Kate	goad
bed	pawed	Keds	cawed
bad	pod	cad	cod

<u>Measurement</u>

Measurements were later made directly on the monitor while the tape was on "pause". A ruler was used whose smallest units were millimeters. For each vowel spoken, one frame was chosen which represented the apparent culmination of the vowel. For round vowels this was usually the point at which the lips were extended most forward, and for the unrounded vowels this was usually a frame before the jaw or lips began rising for the following consonant.

This approach differs from Linker (1982 p.3), who made a photograph of each vowel as the subjects produced them. Later, measurements were made from enlarged photographs. (Linker reports that a similar study done earlier by Fromkin was conducted in the same fashion.) We used the VCR because it was available, but we believe it is potentially the superior way to study lip movement. Since we could slow the tape down to frame- by-frame speed and reverse the picture as well, we could isolate the articulatory culmination of the vowel. Even with much experience, it was often necessary to reverse the tape to find this spot precisely and it was absolutely necessary to proceed as slowly as frame-by-frame would allow. Furthermore, often the vowel would reach this point in a single frame. But, imagine, there are 30 frames per second! We wonder how accurately one can pinpoint the achievement of an articulatory gesture by attempting to do it in real time, as Linker and Fromkin apparently did when they had to decide when to snap the picture as their subjects were producing vowels.

However, the VCR approach has drawbacks serious enough that until some way is found of circumventing them, it is perhaps no better than Linker's approach. The most important one has to do with the way measurements are made. Since Linker had hard copy, she could fix points reliably and measure distances more accurately probably than we could using a ruler on the video screen. The measures we made involved, for example, distances between a single point and several others. A point could not be marked in any way on the face of the monitor, so no point was fixed permanently. Each time we had to lift the ruler, we would have to determine again where that particular point might be. (We had consistent criteria, of course, and my feeling is that we were fairly accurate and consistent in identifying most points, although it would still be better to be positive.) Also, a point would appear to be in a different place depending on whether one was looking at it head-on or off to one side. Both of these problems could well have contributed to measurement error which was most likely avoided in Linker's approach.

The following points were defined on the lips, and various distances between them became the measures of lip movement. These particular ones were chosen because they seemed like good measures of lip protrusion and spread, lip approximation, and vertical and horizontal opening. Two variables were constructed which reflected the area enclosed by the outer boundary of the lips and by the opening between them. All variables except those involving the outer lip boundary were also chosen because Linker included them in her study.

Points		Distance between points
1,2	corners of the lips	HDO (horizontal distance-outside)
3,4	corners of the inner boundary of the lips	HDI (horizontal distance-inside)
5,6	point on the outer boundary of the lips half-way between 1 and 2	HFO (half-outside)
7,8	points on the inner boundary of the lips halfway between 3 and 4	HFI (half-inside)
9, 10 11,12	points halfway between the left corner and halfway point and right corner and half-way point, respectively	FQO (first quarter-outside) SQO (second quarter-outside)
13,14	points halfway between the left inside corner and the inside halfway point and likewise on the right side	FQI (first quarter-inside) SQI (second quarter-inside)

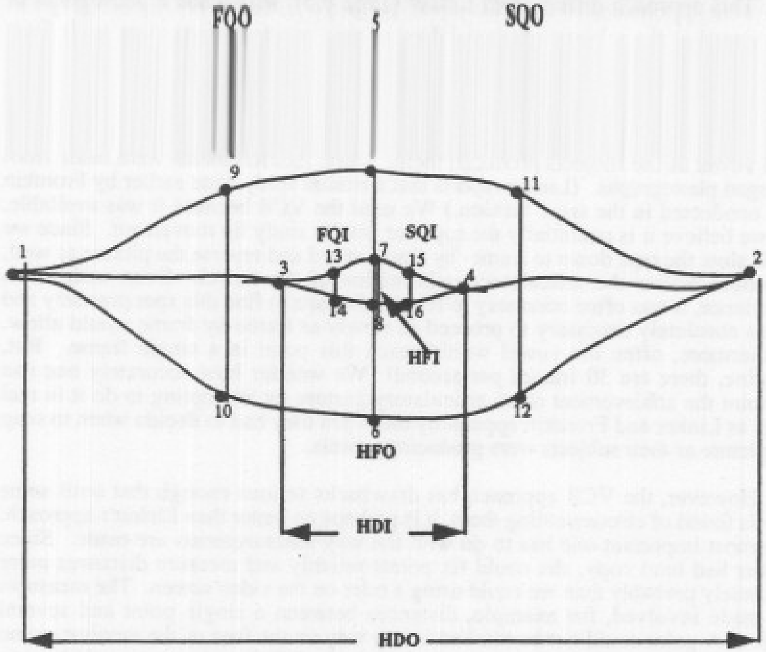


Figure 1. Frontal Measurements

<u>Constructed variables</u>

OAREA—the area enclosed by the outer boundary of the lips
IAREA—the area of the opening between the lips

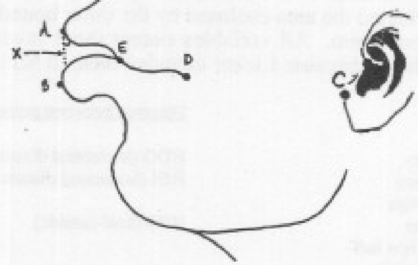


Figure 2. Side Measurements

<u>Points</u>

A	the furthest point out on the upper lip
B	the furthest point out on the lower lip
C	a stationary reference point: where the earlobe meets the face
D	the corner of the mouth
E	the point of forward-most contact of the lips
X	a line dropped vertically from A

Distances between points

AB	distance between the lower and upper lips
DC	distance between the corner of the mouth and the reference point
AC	distance between the upper lip and the reference point
BC	distance between the lower lip and the reference point
AD	distance between the upper lip and the corner of the mouth
BD	distance between the lower lip and the corner of the mouth
ED	length of contact between the lips
AE	distance between upper lip and forward-most contact of lips
BE	distance between lower lip and forward-most contact of lips
XE	distance between E and the reference line dropped from A

Analysis

Analysis of variance (ANOVA) procedures were used to analyze the data collected. Three different models were constructed. The first modeled each dependent variable (i.e., each of the measures identified above) as a function of the classificatory independent variables *round, language, speaker*, and a *round x language* interaction. The second modeled each dependent variable as a function of the independent variables *vowel, language, speaker*, and a *vowel x language* interaction. The third added *context* and the interactions *context x language, context x vowel*, and *context x vowel x language* to the independent variables of the second model.

The main effects of *round, vowel, language*, and *context* were examined for their significance. Here we would be asking if a significant amount of the variance in the dependent variable was being accounted for by that particular effect. However, we could interpret a significant result directly for a given dependent variable only if there were no significant interactions involving the independent variables.

The interactions were also examined for significance. By testing the significance of the *round x language* interaction, we are asking if the difference between the means for the German and English rounded vowels is different than the difference between the means for the German and English unrounded vowels. That is, a given measure might behave differently in German than in English for the different values of "round". In other words, does German implement "round" differently than English?

The results of the *vowel x language* interaction will indicate whether vowels which are transcribed the same in German and English differ in different ways from each other (e.g., round vowels might involve less vertical opening in German but unrounded vowels might involve more).

The test of the *context x language* interaction probes whether German vowels vary overall with context differently than English vowels. The test of the *context x vowel* interaction tells whether different vowels behave differently in different contexts across the two languages. And the *context x vowel x language interaction*, if significant, will say that some individual vowels interact differently with context in German than in English.

For measures with significant interactions, a Newman-Keuls test was performed to determine which simple effects were significant. If a simple effect was significant, the means were compared to determine the direction of the difference.

A one-way ANOVA was done for German front and back rounded high vowels and one was done for the front and back mid vowels, modeling the dependent variables as a function of *vowel*. Results were examined for significance levels.

Results

A summary of some results is given in Table 2. These include the significance of the interactions of language and other independent variables, the significance of the main effect of language, and a comparison of the means of German vowels (without the front rounded vowels) with the English vowels for each dependent variable for which there was at least one interaction which was not significant and for which the main effect of language was significant.

Table II. Summary of selected ANOVA results
(dashes indicate a non-significant result)

variable	round x language	vowel x language	context x language	language	comparison of means
HDO	--	--	--	.0001	E > G
HDI	--	--	--	.0001	E > G
HFO	--	--	--	.0001	E > G
FQO	--	--	--	.01	E > G
SQO	--	--	--	.0001	E > G
AC	--	--	--	--	
BC	--	--	--	--	
AB	.01	.01	--	--	
DC	--	--	--	--	
ED	.0001	.0001	--	.0001	E > G
AD	--	--	.04	.0001	G > E
BD	--	--	--	.0001	G > E
XE	--	.0001	--	.0001	G > E
HFI	--	--	.02	.0001	E > G
FQI	--	.05	--	.005	E > G
SQI	--	--	--	.001	E > G
AE	.0001	.0001	--	.001	E > G
BE	.0001	.0001	--	--	
OAREA	--	--	--	.0001	E > G
IAREA	--	.05	--	.0001	E > G

<u>Main Effects</u>

1.- For all measures taken from the head-on view, English means are consistently larger for all vowels. These are measures of horizontal and vertical opening. The one exception is the variable *FQI*, which shows an interaction of vowel x language.

An analysis of the simple effects shows that the mean for this measure in the English vowel [o] is larger than that of the German, but for all other vowels, there is no significant difference. (The means themselves are given in Appendix A.)

2.- The only measures of lip protrusion which were made from a stationary reference point (AC, BC) are not significant for any measure.

3.- Other measures of lip protrusion, namely AD and BD, are consistently larger for all German vowels. There is an interaction of context and language for the variable AD; however, the means of the German vowels in each context are still larger than those of the English vowels in corresponding contexts.

4.- Two effects not listed in this Table 2 are context and *context x vowel*. Neither was significant for any dependent variable.

<u>Interactions</u>

1.- *round x language* This interaction was significant for 4 dependent variables. Graphs of the means are presented below in Figure 3. Significant differences between the means between languages was determined with Newman-Keuls tests and are marked with an asterisk. (All are significant at the $p < .05$ level.) Ignoring the differences in the unrounded vowels for the time being, only ED and BE show a significant difference for German and English back rounded vowels. ED is longer in German vowels and BE is shorter.

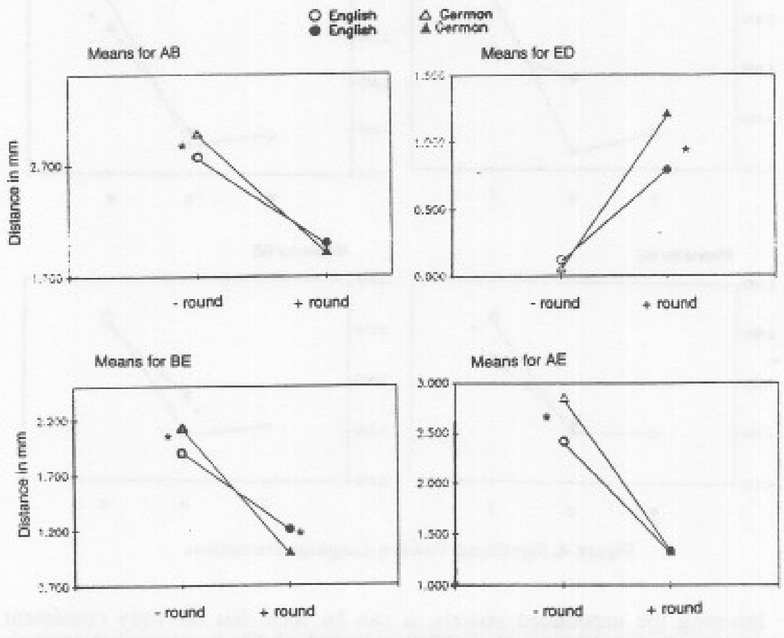


Figure 3. Significant Round x Language Interactions

2.- *vowel x language* This interaction was significant for 7 variables, 6 of which we will consider here. Graphs of the means for these are presented below in Figure 4. Again, differences between the same vowels in the two languages were determined by Newman-Keuls tests. All differences which are significant are significant at a $p < .05$ level and are marked with an asterisk.

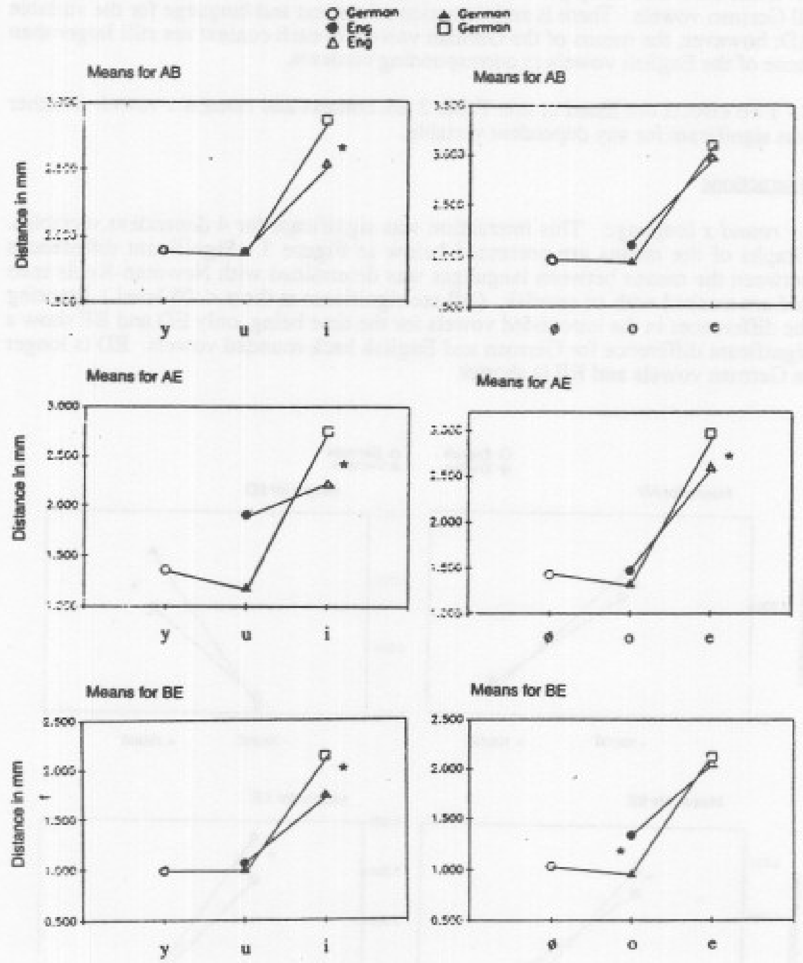


Figure 4. Significant Vowel x Language Interactions

Ignoring the unrounded vowels, it can be seen that the only consistent measures for distinguishing both the high and mid back German vowels from their English counterparts are ED and IAREA. German rounded vowels have larger

ED's and English vowels have greater IAREA. The only other significant result involving rounded vowels is that for the back mid vowel, BE is longer in English than in German.

3.- *context x language* This interaction was significant for two variables, HFI and AD. The means are presented below in the graphs in Figure 5. Apparently, HFI is shorter for German vowels in a bilabial context than the other context, and the reverse is true for English vowels. However, *none of these simple effects within language is significant (and none even approaches significance). The means for AD are shorter for the German vowels in a bilabial context as well, and the reverse is* true for English vowels. However, again none of these simple effects are significant. That of German bilabials versus other approaches significance at p < .08.

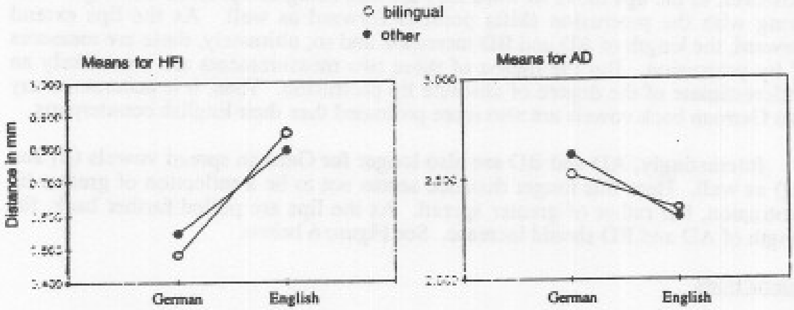


Figure 5. Significant Context x Language Interactions

German front rounded versus back rounded vowels

1. The same measures were not significant for the high and mid vowels. Only AE was significantly different for [y] and [u], and HDI and SQI were significantly different for [ø] and [o]. No other measures for either high or mid rounded vowels proved significant. A list of means for these significant measures is given in Table 3. Note that [y] has a significantly longer AE than [u] and that [ø] has significantly larger HDI and SQI. All were significant at the p < .05 level.

Table 3. Means for selected measures for German rounded vowels

AE		HDI		SQI	
[y]	1.36672176	[ø]	1.47731458	[ø]	18719131
[u]	1.19645537	[o]	1.22754611	[o]	15092589

Discussion

Main Effects

Returning to the first question posed in the introduction, it seems that German back vowels are indeed more rounded than their English counterparts. These

German vowels are firstly more closely approximated than in English, as they all involve significantly less vertical opening and horizontal opening. However, close approximation is only one component of rounding or one possible way of implementing this feature. It is possible that the these German vowels are more closely approximated but not more protruded. To determine this, we must consider the measurements taken from the side view and examine specifically measures of lip protrusion. Unfortunately, since the measures of protrusion measured from a fixed reference point proved to be not significantly different, i.e., AC and BC (probably because their variances were quite large), we must turn to other measures which perhaps do not provide this information as directly.

AD and BD were consistently longer for German vowels than for English vowels. Recall that these are distances between the outermost points on the upper and lower lip respectively and the corner of the mouth. If the lips are more protruded, A and B would be expected to extend more away from the face. However, as the lips move forward they are also being compressed sideways. This along with the protrusion shifts point D forward as well. As the lips extend forward, the length of AD and BD increases, and so, ultimately, these are measures of lip protrusion. But the means of these two measurements are most likely an underestimate of the degree of absolute lip protrusion. Thus, it is accurate to say that German back vowels are also more protruded than their English counterparts.

Interestingly, AD and BD are also longer for German spread vowels ([i] and [e]) as well. Here this longer distance seems not to be a reflection of greater lip protrusion, but rather of greater spread. As the lips are pulled farther back, the length of AD and BD should increase. See Figure 6 below.

<u>Interactions</u>

Some variables tested significantly for interactions. The results of the simple main effects within the round x language interaction for ED and BE support the conclusions in (A) above although it is not as easy to see how. BE is shorter in German back rounded vowels than in English back rounded vowels. While it seems as if BE is a similar measure to BD (and thus should be expected to lengthen if the lips were extending farther forward), it is really quite different. Recall that B is the point of forward-most contact of the lips. The longer the length of this contact, the less distance there is left over from the point where the lips separate (E) to the outermost points on the upper or lower lip. Note also that the length of the contact of the lips referred to above is the measure ED. When the lips are more protruded, it should be longer. (See Figure 7). Thus, one can conclude that German back vowels are more protruded than English back vowels, which supports the claims in (A).

The results of the *vowel x language* interaction don't tell much of a neat story about the way various identically transcribed vowels behave differently in the two languages. First, however, it can be said that the results of the tests of the simple effects within this interaction for the variables ED and IAREA lend further support to the claim that German back vowels are more rounded than English. ED is significantly longer for German [u] and [o], corresponding to greater upper and lower lip protrusion as discussed above. (See Figure 7). However, one wonders why there is no difference in the means of BE or AE. Could the lips be extended so far forward that a BE and AE distance is created equal to that of the English less protruded vowels?

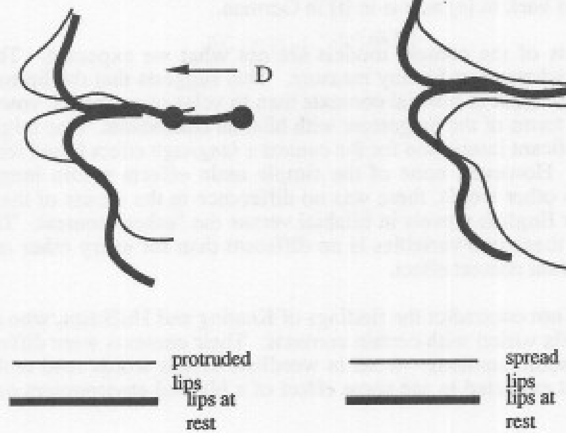


Figure 6. The Effect of AD and BD

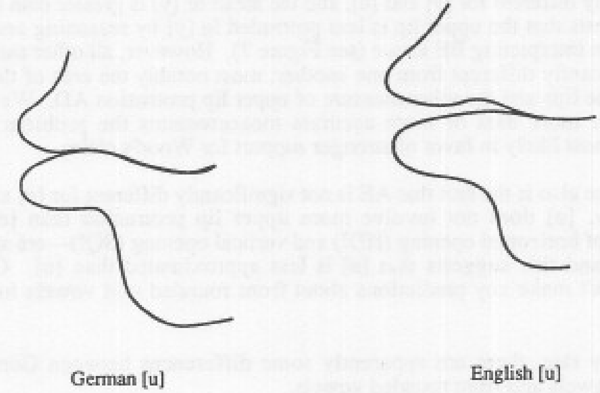


Figure 7. The Effect of BE and ED

Another possibility is that English back vowels are less protruded and more approximated at the same time. However, all measures of lip approximation show that English vowels are less approximated (except for AB, which is not significant). Means for the area of the opening between the lips (IAREA) are also greater for all English vowels. This also suggests that the lips are less approximated and less protruded in English back vowels. However, the spread English vowels also have greater area of opening than German spread vowels. Given that German non-round vowels seem to be more spread based on the interpretation of their greater AD and BD lengths, one wonders if increasing the spread of the lips decreases the area. It certainly does not seem as if it would. Furthermore, AE, BE and AB are all longer for German [i], suggesting that the upper and lower lips are more stretched out and the distance between the outermost points of the two lips are farther apart than in

English [i]. That is, it suggests that German [i] is more spread. Only AE is significantly longer for [e] in German than English, suggesting that the lower lip is not as stretched back in [e] as it is in [i] in German.

The results of the context models are not what we expected. There is no *context* effect whatsoever for any measure. This suggests that the lip rounding in vowels is no different in bilabial contexts than in velar ones, that is, vowels do not coarticulate in terms of the lip gesture with bilabial consonants. One might wonder about the significant interaction for the *context x language* effect found for variables HFI and AD. However, none of the simple main effects within language was significant. In other words, there was no difference in the means of the variables HFI or AD for English vowels in bilabial versus the "other" context. That is, the situation with these two variables is no different than for every other measure in terms of a relevant context effect.

This does not contradict the findings of Keating and Huffman, who noted that Japanese vowels varied with certain contexts. Their contexts were different; they compared prosodic contexts—word in wordlists versus words read embedded in prose. We just expected to see some effect of a bilabial environment on rounded vowels.

Turning now to the final question of whether German back rounded vowels are more rounded than front rounded vowels, the situation is also a bit murky. AE is significantly different for [y] and [u], and the mean of [y] is greater than that of [u]. This suggests that the upper lip is less protruded in [y] by reasoning analogous to that used in interpreting BE above (see Figure 7). However, all other measures are not significantly different from one another, most notably the area of the opening between the lips and the other measure of upper lip protrusion AD. We think that with either more data or more accurate measurements the problem might be resolved-most likely in favor of stronger support for Wood's claim.

Strange also is the fact that AE is not significantly different for [ø] and [o], so, apparently, [ø] does not involve more upper lip protrusion than [o]. Other measures-of horizontal opening (HDI) and vertical opening (SQI)—are significant, however, and this suggests that [ø] is less approximated than [o]. Of course, Wood didn't make any predictions about front rounded mid vowels in the 1986 paper.

At any rate, there are apparently some differences between German back rounded vowels and front rounded vowels.

Conclusion

We sought to answer three questions with this project. Are German back rounded vowels more rounded than English back rounded vowels? Do German vowels vary more with context than English vowels? Are German back rounded vowels more rounded than front rounded vowels?

We think it is very clear that the answer to the first question is a definite yes. The other two questions were answered less confidently. We could find no effect of context. Perhaps with more precise measurement techniques or more varied contexts, one would detect such an effect. We uncovered limited support of Wood's prediction. We would have expected other effects to be significant, but given that a consistent, although not unambiguous, measure of upper lip protrusion was significant for [u] and [y], cautious support of the claim that [y] is less rounded than [u] is extended. We suppose that with more precise measurements, stronger effects would be uncovered. In addition, something which has not been previously mentioned might have obscured some smaller effects. Measurements were not modified in any way to neutralize differences in shapes and sizes of speaker's lips. These differences certainly existed, the question is whether they were large enough to cover up effects. We think it would be a worthwhile question to pursue.

Finally, we suggest that some improvement of the method we used in this experiment in terms of making more accurate and precise measurements would make this approach far superior to those which involve making static records of a dynamic articulatory gesture.

*The original version of this paper was presented in the 119th meeting of the Acoustical Society of America, State College, Pennsylvania, 21-25 May 1990.

References

Disner, Sandra Ferrari. 1983. *Vowel Quality: the relation between universal and language-specific factors.* UCLA Working Papers in Phonetics, No. 58.

Keating, Patricia A. and Marie K. Huffman. 1984. "Vowel Variation in Japanese" *Phonetica*, 41:191-207.

Linker, Wendy. 1982. *Articulatory and Acoustic Correlates of Labial Activity in Vowels: A Cross-linguistic Study.* UCLA Working Papers in Phonetics, No. 56.

Wood, Sidney. 1986. "The Acoustical Significance of Tongue, Lip, and Larynx Maneuvers in Rounded Palatal Vowels," *Journal of the Acoustical Society of America,* 80 (2): 391-401.

Appendix A : Means for dependent variables by language

dependent variable	German	English
HDO	5.074465725	6.42247748
HFO	2.198813670	3.59766604
FQO	2.162718415	2.44257300
SQO	2.554385075	2.980625045
AC	10.8798038	10.60764903
BC	10.71210527	10.35472996
AB	2.272693475	2.37592131
DC	8.93014642	8.84531106
ED	0.83037591	0.457089875
AD	2.57639892	2.334414875
BD	1.91206837	1.71832825
XE	1.22180332	1.172476725
HFI	0.515213405	0.81974796
FQI	0.38512083	0.550455275
SQI	0.43835514	0.71371507
AE	1.84151771	1.874316315
BE	1.37503010	1.55356930
OAREA	7.00491681	9.09679273
IAREA	0.84705746	1.57053812

Labial Position and Acoustics of Korean and English High Vowels

Sun-Ah Jun

Department of Linguistics, The Ohio State University

Abstract

This paper examines and contrasts the labial configuration and formant frequencies of Korean and English high vowels. Korean has three high vowels, /i, ɨ, u/, and English has four, /i, I, U, u/. The lip gestures and formant frequencies were compared within each language and across languages to determine whether the idea of maximal dispersion (Liljencrants and Lindblom (1972)) can account for the distribution of formant frequencies and also be extended to account for the labial configurations. Four male speakers of each language produced each vowel in four different contexts with five repetitions. The production of each word was videotaped and the sound was simultaneously recorded. The lip configurations were assessed using measurements similar to those in Linker(1982) and MANOVA was performed on the measurement data. The formant values were converted to the Bark scale to better represent the perceptual distance. The results of formant frequencies show that the mean F2 values of English back vowels, /u, U/, are significantly higher than that of Korean back vowel, /u/, while the F2 means of /i/'s in both languages are almost the same. The fact that Korean high vowels take more spaces in F2 is contradictory to the prediction of dispersion theory. The same tendency is shown in the result of the labial measurements. That is, Korean /u/ is produced with more rounded and more protruded lips than English /u/, while /i/'s in both languages are produced with the same degree of lip opening even though Korean /i/ used more spread lips. However, if we consider vowels in the same height (similar values of mean F1), we can find the maximal dispersion theory could explain the result since English has only two high vowels, /i, u/, and Korean has three.

Introduction

Studies of vowel spaces across languages have suggested to phoneticians that there are at least three possible principles to account for the distribution. They are Lindblom's *principle of maximal dispersion* (Liljiencrants & Lindblom, 1972), a more generally proposed principle of sufficient contrast (Lindblom, 1975, 1979, Maddieson, 1977), and Stevens's quantal theory (Stevens, 1972, 1989).

By the principle of maximal dispersion, Liljiencrants and Lindblom proposed that vowels tend to be distributed so as to be maximally far from one another in the available phonetic space. And the point vowels, /i/, /u/, /a/, tend to be pheripheral, maximizing the difference from center of vowel space.

The principle of sufficient contrast is actually a revised version of *the dispersion theory*, suggested by Lindblom (1975, 1979), Terbeek (1977) and Maddieson (1977). This theory says that vowels need not be maximally separated, but only separated far enough to contrast perceptually with each other. The problem with this theory is that it is not clear what the "enough space" means. For this problem, Lindblom suggested that the degree of this contrast is invariant across languages and system size, but the phonetic values of vowel phonemes should exhibit more variation in small than in large systems.

The third theory of vowel systems, Stevens's(1972, 1989) *quantal theory* is completely independent of the first two. Stevens's model predicts fixed phonetic spaces for the point vowels, irrespective of the size of vowel systems. According to this theory, the relationship between articulatory and acoustic dimension is not linear, so that there are regions of the vocal tract where relatively great articulatory variations produce negligible changes in acoustic signal. Such an acoustically stable region is called a quantal region. Stevens proposed that the three point vowels are produced in such regions. Since point vowels do not require a great articulatory precision, most languages tend to use the same point vowels and they will occupy the same optimal positions in the acoustic space.

This paper studies the distribution of high vowels in two languages with different vowel systems. Korean has three phonologically high vowels, /i, ɨ, u/, and English has four, /i, ɪ, ʊ, u/. The distribution of high vowels in each language is examined acoustically and articulatorily and the observed distribution of measured values is evaluated against the predictions of the three principles mentioned above; *maximal dispersion*, *sufficient contrast*, and *quantal theory*. Most studies of dispersion or quantal theory have used only acoustic data (e.g., Formant values of vowels, voice onset time, tone spacing, etc.). However, in addition to an acoustic area, this paper tries to expand the application of the concept of dispersion or that of quantal area to an articulatory space of vowels. The distribution in the articulatory space is examined by measuring several aspects of labial position for each vowel. Since both languages have phonemically the same point vowels, /i/ and /u/, and each language has one or more vowels which the other language doesn't have, comparing the two languages' same or different vowels as well as their vowel systems as a whole allows us to differentiate the three theories as follows.

The *maximal dispersion theory* predicts that, in order to keep "the maximum contrast distance" between vowels, /i/ and /u/ in both languages would occupy extreme peripheral areas of the acoustic space and have extreme lip movements, while Korean /ɨ/ would occupy an intermediate position between English /ɪ/ and /ʊ/. Furthermore, to keep a maximal distance between vowels, vowels would not vary a lot.

The *sufficient contrast dispersion theory* predicts that the vowel space of Korean high vowels would take less space than English in the F2 dimension, which is believed to be an indicator of frontness versus backness. Also Korean vowels should have less extreme lip rounding for /u/ compared to English /u/, due to the fact that English has two round vowels, /u, ʊ/. In terms of variation, this theory would predict Korean vowels, in a smaller vowel system, would vary more than English ones.

Finally, *quantal theory* predicts that English and Korean /i/ and /u/, as quantal vowels, would take the same positions in acoustic space and use the same degree of labial movement. Also these vowels should vary a lot in labial positions but a little in acoustic space, since point vowels are supposedly very stable in the acoustic output compared with a wide range of articulation input. By contrast, the other non-point vowels would vary more in acoustic space and there should be only a corresponding variability in articulatory space.

To test these predictions, lip gestures and formant frequencies were compared within each language and across languages.

Method

Subjects

Four male speakers of each language participated in this experiment. The English subjects are in their early twenties (two OSU undergraduates and two OSU graduate students); the Korean subjects are in their late twenties (all OSU graduate students). The English speakers are all from a similar American Midwestern dialect area. The Three Korean speakers are from Seoul. One, originally from Kwangju (Chonnam province), had lived in Seoul for 8 years before coming to America. (These two dialects have no differences in vowel quality.)

Materials

Words with each target vowel in 4 different contexts - bilabial /b/ or /p/, alveolar /s/, velar /k/, glottal /h/ - were selected. For English words, coda consonants are not controlled, but they usually ended in /t/ and all of them are real words. Not all Korean words are real words, but the nonwords did not cause any problem for pronunciation. The test words are shown in Table I.

Table I. Korean and English test words

Korean

ki -"energy"	pi - "rain"	si - "an hour (unit)"	hi - NM
kɨ -"that"	pɨ - NM	sɨ - NM	hɨ - NM
ku -"a circle"	pu -"richness"	su -"excellence"	hu -"after"

* NM= no meaning

English

key	beat	seat	heat
kick	bit	sit	hit
cook	book	soot	hood
coo	boot	suit	who'd

Other words with other vowels were selected as foil data. There were 12 foil words for Korean and 16 foil words for English. (The foils are listed in an appendix.) The words were presented to the subjects on large cards (5 in. * 8 in.). 5 tokens of each word were pseudo-randomized so that the first and the last word in each card were from the foil words to avoid any list intonation effect. Each card contains seven words listed vertically.

Recording Procedures

Labial position was captured by video taping in a corner of a room. Each subject was sitting on a chair looking at the video camera, which was mounted on a tripod within 1 meter in front of the subject. A side view was caught by putting a mirror at a 45-degree angle to the right side of the subject's head. Behind and beside the subject's chair, there was calibration paper with horizontal and vertical gridlines 1 cm apart. These calibration grids were used to compare ratios of pictures between subjects and between tapings. One or more books were stacked behind subject's neck to prevent head nodding while reading. A subject's lips were outlined in a dark color for easier measurement later. An adhesive black triangle tape was attached to the subject's earlobe to be used as a reference point for side view measurements. Subjects read each word in citation form by looking at a card which was held by the experimentor standing beside the video camera. Sound was recorded simultaneously for acoustic analysis. Figure 1 shows pictures (printed with a video printer) of one frame of the video film for a production of /u/, by a Korean speaker (above) and an English speaker (below).

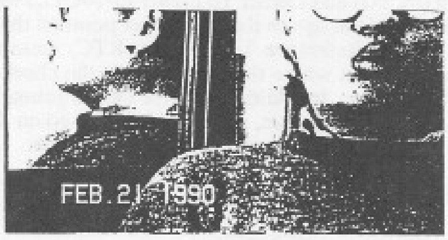


Figure 1. pictures of video film. (Hard copies of TV screen) The above one is a Korean subject and the below is an English subject saying [u].

<u>Measurements</u>

The tape was played on a VCR. A frame was chosen and paused when the lip movement reached a steady state, indicating the vowel reached its target position. Total eleven measurements of each lip shape were made on a TV monitor attached to the VCR. Four measurements were taken from the front view and seven measurements were taken from the side view. These are shown below :

<u>From the front view</u>

HOW : Horizontal outer width
HIW : Horizontal inner width
VCO : Vertical central opening
VSO : Vertical side opening

<u>From the side view</u>

VO : Vertical distance between upper and lower lips
CPP : Corner to protrusion plane

RTC : Reference to corner
CTU : Corner to upper lip
CTL : Corner to lower lip
VCU : Vertical distance from corner to upper lip
VCL : Vertical distance from corner to lower lip

Figure 2 defines these measurements. Most of these were adopted from measurements by W. Linker(1982). Among the side view lip measurements, Corner To Upper lip (=CTU) and Corner To Lower lip (=CTL) were defined as the distance from the corner of the lips to the outer most point on the upper and lower lips, respectively. For the Reference To Corner (=RTC) measure, the reference point was chosen as the point where the earlobe joins the cheek, since this place hardly moves when speaking. In addition to these eleven primary measurements, area of lip opening, a derived measure, was calculated based on the horizontal and vertical lip opening measurements using the following formula:

$$AREA = (VCO+VSO) * .25HIW + VSO * .25HIW.$$

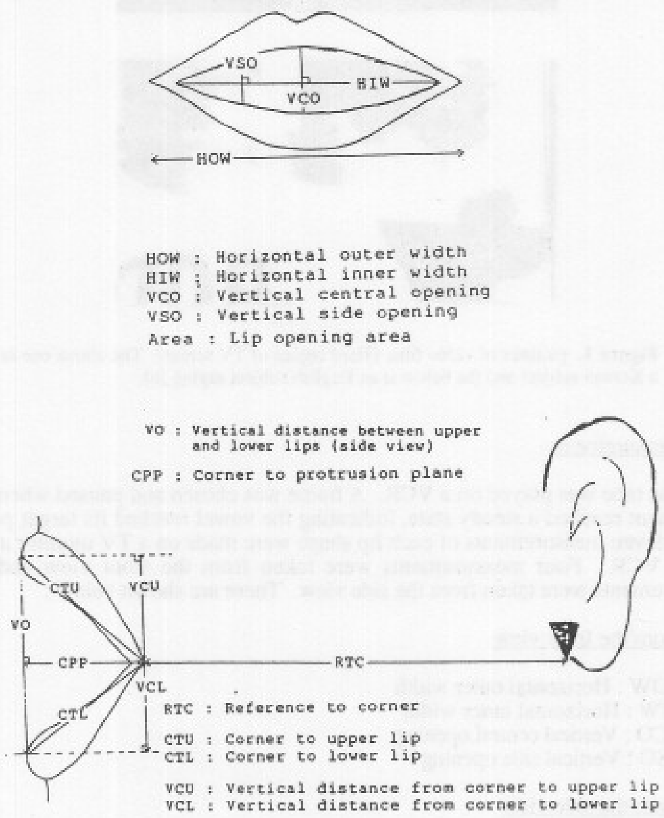


Figure 2. Eleven lip measurements, five from the front view and seven from the side view.

Also, since each subject's physical lip size is not the same, eight ratio values based on Horizontal Inner lip Width (HIW) measurement and four ratio values based on Vertical Central lip Opening (VCO) measurement were calculated as follows:

HIW/VCO,	HIW/RTC,	HIW/CTU,	HIW/CTL
HIW/VCU,	HIW/VCL,	HIW/VO,	HIW/CPP
VCO/RTC,	VCO/CTU,	VCO/CTL,	VCO/CPP

And two ratios of an upper lip opening relative to a lower lip opening:

CTU/CTL, VCU/VCL

were calculated. Thus, a total 25 measurements were collected from each vowel. Each measurement was made using a milimeter-scaled ruler. Whenever possible, a tenth of a milimeter unit was used. Since the side view was captured through a mirror set beside a subject's head, measurements from the side view were smaller than those from front view. To balance the side view with the front view measurements, each measurement of the side view was multiplied by the ratio of front to side view for each subject. These measurement data were analyzed using MANOVA & ANOVA to assess the effect of differences between languages, between vowels, and between contexts, and interactions among all these factors.

The sound was copied from the video tape to a cassette tape. The sound was digitized onto a PC and the formant values were measured using an LPC formant tracker with 14 coefficients, 200 window size and 100 window step. To obtain values that were closer to perceptual distances, formant measurements were converted from Hz to the Bark scale (Traunmuller, 1983).

Results and Discussion

Formant values

Figure 3 shows the formant values, F1 and F2, plotted in the Bark scale for each of the two languages. As the figure shows, English vowels are not maximally separated from one another. Also the vowel space for English is somewhat smaller in the F2 dimension than for Korean; this is mainly due to the low F2 of Korean /u/.

This result contradicts the predictions of all three theories discussed in the introduction. It is counter to the predictions of the Liljiencrants and Lindblom's maximal dispersion theory since the distance between two extreme vowels, /i/ and /u/, are not the same for the two languages. It is also counter to the predictions of Stevens's quantal theory since the same point vowel, /u/, occupies a different acoustic space in different languages. Also, the fact that English has a smaller range of F2 is counter to the predictions of Lindblom's Sufficient Contrast Theory, since English has phonemically more high vowels than Korean along the dimension of backness.

On the other hand, for the vowel /i/ the two languages are not significantly different in their mean values of F1 and F2 and their variation. This supports Stevens's quantal theory which discribes [i] as a quantal vowel, which should be relatively invariant across languages.

Figure 4 shows mean values for F1 and F2 for all of the vowels separately.

The error bars show the standard deviation(SD)s.

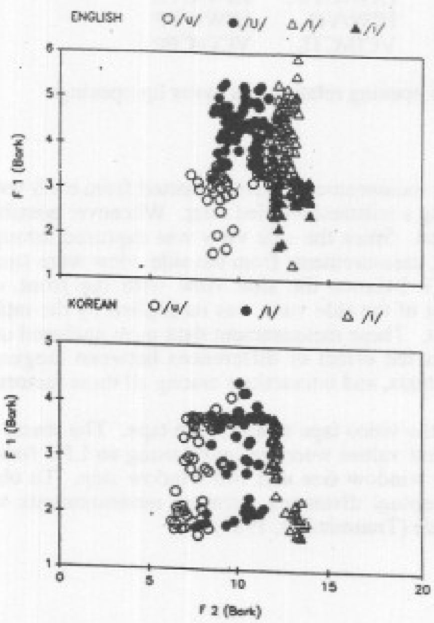


Figure 3. F1 and F2 of both languages in Barks

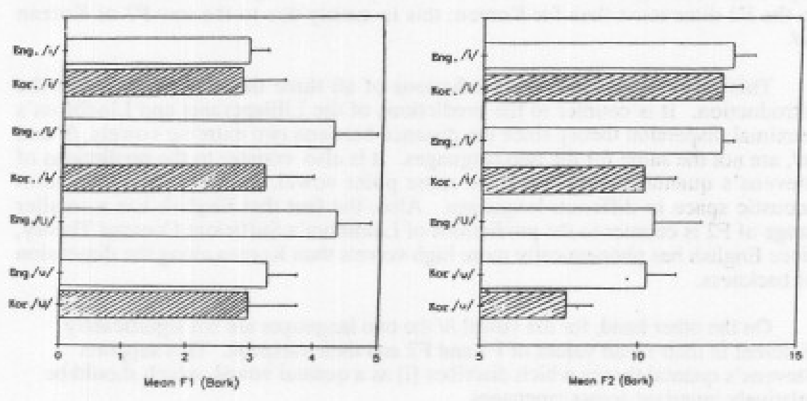


Figure 4. English/Korean means F1 & F2 in Barks

If the prediction of the sufficient contrast dispersion theory is right, Korean vowels should vary more than English vowels, because a smaller system is supposed to vary more. Since we are comparing all phonemically high vowels (meaning similar F1), we assume the variance would come from F2 dimension. For the vowel /i/, Korean does show more variance in F1 (=larger SD) than English, but a similar degree of variation in F2. For the vowel /u/, Korean also shows a larger SD in F1 but a smaller SD in F2 than English. Thus the fact that Korean /u/ varies less than English /u/ in F2 seems to be against the prediction of the sufficient contrast dispersion theory.

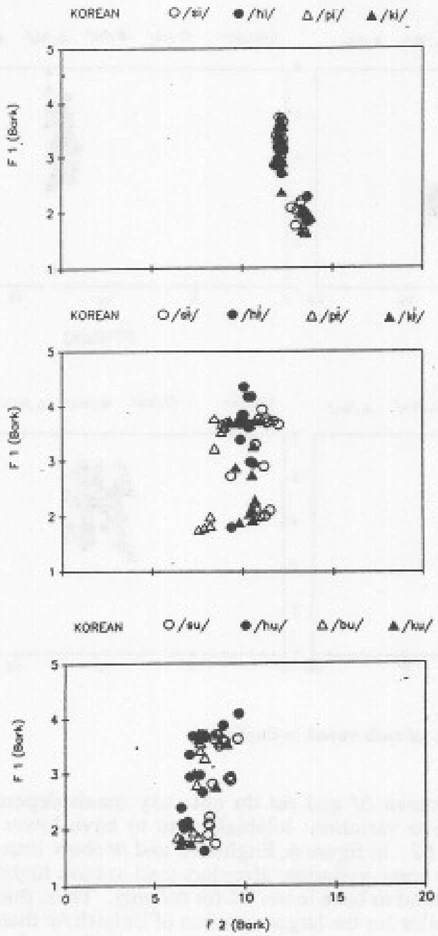


Figure 5. Context variations of each vowel in Korean

On the other hand, if we consider that English /I/ and /U/ are somewhat lower vowels(thus higher F1 values), we might reinterpret the result to be accounted for by the sufficient contrast theory. That is, English has phonetically only two very high vowels whereas Korean has three. Thus, English should vary more in F2. At the same time, this reinterpretation would also explain the larger variance of Korean in F1 dimension: Korean high vowels vary more than English /i/ or /u/ in F1 dimension since English has more vowels in this F1 dimension. Comparing only high tense vowels, we can say this result supports only the sufficient contrast theory, not the maximal dispersion or the quantal theory. Finally, the variations depending on contexts are shown in Figure 5 & 6.

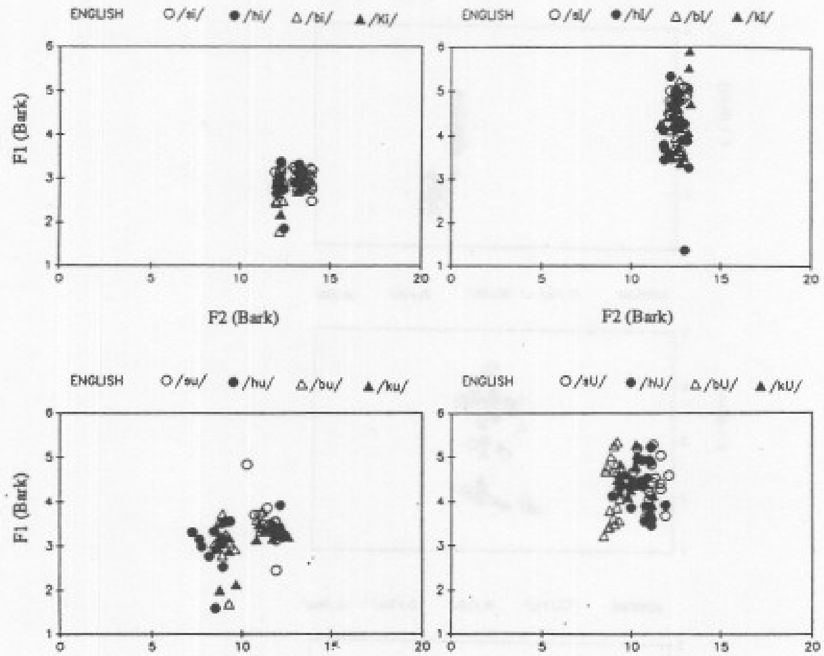


Figure 6. Context variations of each vowel in English

As seen in figure 5, Korean /i/ and /u/ do not vary much depending on contexts while /ɨ/ shows some variation: bilabials tend to have lower F2 and alveolars tend to have higher F2. In figure 6, English /i/ and /I/ show little context variation but /u/ and /U/ show some variation: alveolars tend to have higher F2 for both /u/ and /U/ and bilabials tend to have lower F2 for /U/ only. Thus, this context effect is only partially responsible for the larger variation of Enlgish /u/ than Korean /u/. That is, English /u/ varies more than Korean /u/ regardless of contexts. However, the reinterpreted sufficient dispersion theory cannot explain why front vowels vary very little in both languages.

Labial measurements

The ANOVAs show significant main effects of vowel and context in each language and a significant interactions of vowel*context and language*context. The measurement of RTC (Reference To Corner of mouth) was a good indicator distinguishing front vowels from back vowels in English. A few measurements were not good parameters for distinguishing vowels. They are HIW/VCL and CTU/CTL. VCU/VCL was also not a good parameter for some speakers. Most of the variations in the model of vowel, context, and vowel*context were best seen in five measurements: HIW, VCO, VSO, AREA, and VCO/RTC, in both languages. HIW/RTC also showed relevant variations, but in the Korean vowels only. Vowels in each language showed the same relationship in most measurements. For most lip measurements, a post hoc test of each language showed that English has three significantly different ($p < 0.01$) groups /i, ɪ/, /ʊ/ and /u/, while each of the three Korean vowels is significantly ($p < 0.01$) different from any other. Korean /ɨ/ usually has a measurement value intermediate between those of other two vowels, and English /ʊ/ usually has an intermediate value between /i, ɪ/ and /u/. Based on English grouping, we can tell the articulatory distribution, labial movements in general, of Englsih vowels is not accounted for either by the maximal dispersion theory or by the sufficient contrast theory. Figure 7 shows the relationships of vowels in English and Korean by lip measurements that showed significant effects of vowel.

English /i/ and /ɪ/ have very similar values for all measurements, while all the other vowels have significantly different values with one another. This corresponds to F2 related acoustic patterns described above. That is, English /i/ and /ɪ/ were overlapping more than any other vowel pairs in F2. Whereas all three Korean high vowels were well separated in F2.

However, the relationship between English vowels does not support the distinction between phonetically very high vowels and lower vowels we made before to reinterpret the sufficient contrast theory. Rather, the lower vowel /ʊ/ differs from other vowels in most measures instead of forming one group together with /u/. Next, variations in general and by contexts are considered to be compared with those of the acoustic results.

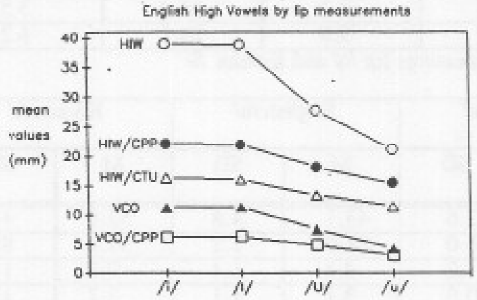


Figure 7. English and Korean high vowels by lip measurement values in means

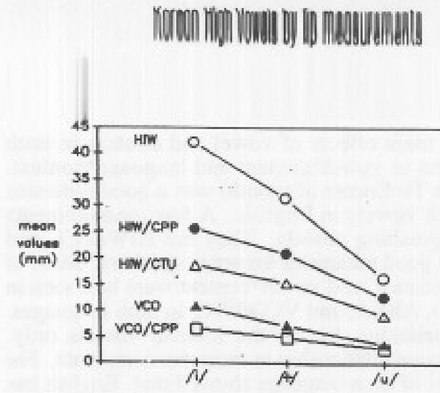


Figure 7. (Continue)

Variations

Variations in every measures for each vowel were considered. Table 2 shows means and SDs for /i/ and /u/ and Korean /ɨ/ (N=80 for each language). English /i/ varied a little more than Korean /i/, while both languages showed a similar degree of variation for every measures for /u/'s.

Table II

a. Mean and SD for each measures for /i/

Measurements	Korean		English	
	M	SD	M	SD
HOW	53.1	4.6	48.2	3.8
HIW	42.1	6.6	39.0	4.6
VCO	10.9	1.1	11.2	1.7
VSO	8.5	1.0	8.8	1.5
RTC	104.1	6.3	88.7	11.4
VO	24.3	3.2	21.0	2.6
CPP	16.8	2.2	18.2	3.5
AREA	29.4	5.9	28.2	6.2

b. Mean and SD for each measures for /u/ and Korean /ɨ/

Measure-ments	Korean /u/		English /u/		Korean /ɨ/	
	M	SD	M	SD	M	SD
HOW	48.5	3.6	43.1	4.8	52.6	4.6
HIW	16.4	6.0	21.0	2.2	31.4	8.0
VCO	3.7	1.2	4.3	1.4	7.3	1.8
VSO	2.7	0.9	3.1	1.1	5.7	1.3
RTC	113.3	4.9	95.6	12.4	106.1	6.0
VO	18.4	2.4	14.8	2.2	20.7	3.4
CPP	13.0	3.2	14.3	3.0	5.3	2.7
AREA	4.0	2.5	5.6	2.2	15.2	6.4

These results of articulatory variation does not always agree with that of the acoustic variation. First, the fact that F2 of English /u/ varied more than Korean /u/ corresponds to the larger variation of RTC in English, which is a measure of lip rounding and protrusion. On the other hand, the larger variance of CPP (a measure of protrusion) in Korean /u/ implies more variance in F2 dimension. Second, the larger variance of F1 in Korean /i/ and /u/ than English does not correspond to the similar variance of VCO, VSO or VO (measures of lip opening) in both languages.

In addition to these vowels, Korean /i/, however, varied in general more than any other vowel. If we apply our reinterpreted sufficient contrast theory to this result, we can find the reinterpretation doesn't hold for the labial movements of each vowel since Korean, having phonetically more number of high vowels, varies articulatorily more than English. Therefore, we may need more evidence to claim that the sufficient contrast dispersion theory works for the distribution of vowels in phonetically same dimension. Or, variations in labial movements may not reflect the variations of acoustic values very well. We may need to examine tongue shapes in addition to labial positions.

Context effects

Significant main effects of context and significant vowel*context and language*context interactions were found. More measurements showed signifi cant context effects in English than in Korean; 16 (HIW, VCO, VSO, VCU, VO, AREA, HIW/VCO, HIW/RTC, HIW/CTU, HIW/CTL, HIW/VCU, HIW/VO, HIW/CPP, VCO/RTC, VCO/CTL, and VCO/CPP) out of 26 measures showed a context effect in English, while only 7 (HIW, AREA, HIW/RTC, HIW/CTU, HIW/CTL, HIW/VO, and HIW/CPP) did in Korean. A post hoc test of context showed the bilabial context to cause significantly smaller values than those of other contexts for most measures, and this matches the acoustic results described above.

Next, context effects on each vowel in each language were considered. The following table shows measures which show context effects for each vowel.

Here, all English vowels except /u/ have context effects for many measures. This is almost the opposite of the acoustic result described above, since by formant values English /i/ showed little variation by contexts, while /u/ showed some variation. On the other hand, the context effect for Korean is very similar to the acoustic result: only / /i// varies a lot.

Table III.

Vowels	Measures showing context effects
English /i/	HIW, VCL, AREA, HIW/RTC, HIW/CTU, VCU/VCL, VCO/RTC, VCO/CTU, VCO
English /I/	HIW, VCO, VSO, VO, AREA, HIW/RTC, VCO/RTC, VCO/CTU, VCO/CTL, VCO/CPP
English /U/	HIW, VCO, VSO, AREA, HIW/VCO, HIW/CTU, HIW/VO, VCO/RTC
English /u/	HIW/CTU
Korean /i/	None
Korean / /i//	HIW, VCO, VSO, AREA, HIW/RTC, HIW/CTU, HIW/CTL,HIW/VCU, HIW/VO, HIW/CPP, VCO/RTC
Korean /u/	None

Since some of 26 measures are highly correlated with one another, multi variate ANOVA was performed for context effects on each vowel. There was a significant ($p < .01$) main effect of context on all English vowels. Korean /i/ had a significant ($p < .01$) main effect of context while /u/ had a slightly significant ($p < .05$) context effect. So, there was a language by context interaction.

In summary, this context effects show a similar tendency with the variance effects described above in that English /i/ varies more and shows more context effects than Korean /i/, but both languages have a small variance and small context effects for /u/. These tendencies are again the opposite of the acoustic result in terms of variation and context effects. Maybe some other factors such as the tongue shape are involved to compensate for the labial movements to form the acoustic patterns.

Individual vowels

Phonemically same vowels in each language were compared. For a vowel /u/, English showed a significantly ($p < .01$) larger value than that of Korean for HIW, meaning English /u/ has horizontally more spread lips. English also showed a significantly ($p < .05$) larger value than that of Korean for VCO, meaning English /u/ is produced with vertically more open lips. Since AREA value is highly correlated with HIW and VCO values, English /u/ has, there fore, a significantly ($p < .01$) larger AREA value than that of Korean. Finally, English /u/ showed a significantly ($p < .01$) larger value of CPP and smaller value of RTC, HIW/VCL, and VCO/CTL, meaning English /u/ is less round ed and protruded than Korean /u/. This result agrees to the acoustic outputs, since English /u/ was higher than Korean /u/ in F2 values (= less rounded). The lip measurement values for /u/ in each language are in Table IV.

Table IV. Mean lip measurements for /u/ in millimeters.

Lang.	HIW	VCO*	AREA*	CPP*	RTC*	HIW/ VCL*	CTU	CTL	VCO/ CTL*
English	21.0	4.3	5.6	14.3	95.6	54.4	18.9	14.9	3.1
Korean	16.4	3.7	4.0	13.0	113.2	46.0	17.5	15.5	2.4

* : significantly different ($p < 0.01$)

For a vowel /i/, Korean showed a significantly larger HIW, RTC and HIW/VCO values and a smaller CPP value, meaning Korean /i/ is spreader than English. But both languages were not significantly different with each other in VCO or VSO measures for /i/, meaning both languages use similar amount of vertical lip opening. And this is supported by the acoustic result since /i/ in both languages were very similar in their F1 values (see figure 4).

For both /u/ and /i/ vowels, English showed a significantly large value of CTU than Korean while both languages were not significantly different by the value of CTL. This may mean that English speakers use more upper lips than Koreans do. However, this assumption could be too strong since these measures were not a significant indicator of vowels for some speakers as mentioned above. Mean lip measurements for /i/ in each language are in Table V.

Table V. Mean lip measurements (in millimeter) for /i/

Lang.	HIW*	RTC	CPP*	HIW/ VCO*	AREA	VCO	CTU*	CTL
English	38.9	88.7	18.2	35.5	28.2	11.2	24.1	9.3
Korean	42.1	104.1	16.8	39.1	29.4	10.9	22.9	20.2

*: significantly different ($p < 0.01$)

In sum, since the same point vowels in the two languages showed different articulatory patterns for most measures as in the acoustic result, the predic tions of maximal dispersion and quantal theory are not working at least in these languages.

Conclusion

We have examined the result of labial position and acoustic values in terms of three principles of vowel systems and have compared the acoustic result with that of articulatory result within and across languages. The salient foundings are as follows.

First, some acoustic results matched the patterns seen in the labial gesture measurements: /i/'s in both languages have very similar formant values and variance and their values for vertical lip opening and area of lip opening are not significantly different from each other. Also Korean /u/ showed more lip rounding or protruding than did English /u/, both acoustically and articulatorily.

Second, some other acoustic result did not match the articulatory result very well. English /i/ showed larger variation and context effects than Korean /i/ in labial positions but showed less variation in F1 and similar variation in F2. /u/ in both languages showed a similar variance in labial position but English /u/ showed more variation in F2 than Korean /u/. Korean /i/ showed the largest variance in labial measurements but not the larger variance in acoustic values.

Finally, contrary to the predictions of maximal dispersion theory and quantal theory, the two languages' pheripheral vowels or point vowels, /i/ and /u/, did not take the same extreme spaces. Rather, Korean with three high vowels takes up more of the F2 space than does English with four high vowels. This difference results from a lower F2 value in Korean /u/. At first glance this result seems to be contrary to the sufficient contrast theory of disper sion also. However, if we consider vowels with phonetically the same height, the sufficient contrast theory would explain the smaller F2 range in English compared to Korean, as well as the larger variation of English /u/ compared to that of Korean /i/ or /i/At the same time, however, this reinterpretation cannot explain the three groupings of English vowels, /i, I/, /U/ and /u/, nor why Korean /i/ varies more than English /i/ or /u/ in the labial configuration measurements. That is, the reinterpretation cannot explain the results of labial configuration. To explain the dismatch between acoustic results and labial position measure ment result, we need to measure some other factors such as tongue gestures.

*The original version of this paper was presented in the 119th meeting of the Acoustical Society of America, State College, Pennsylvania, 21-25 May 1990.

References

Bladon, A. and Fant, G. 1978. "Speech analysis" in *STL-QPSR*

Bladon, A. 1983. "Two-formant models of vowel perception: short-comings and enhancements," *Speech Communication* 2: 305-313.

Disner, S.F. 1978. "Vowels in Germanic languges," *UCLA Working Papers in Phonetics* 40. UCLA Phonetics Laboratory, Los Angeles.

Disner, S.F. 1983. "Vowel Quality : the relation between universal and language-specific factors," *UCLA Working Papers in Phonetics* 58.

Liljencrants, J. and Lindblom, B. 1972. "Numerical simulation of vowel quality systems : the role of perceptual contrast," *Language* 48, No.4.

Lindblom, J. and Sundberg, J. 1971. "Acoustical consequences of lip, tongue, jaw, and larynx movement," *JASA* 50, No.4: 1166-1179.

Linker, W. 1982. "Articulatory and acoustic correlates of labial activity in vowels: a cross-linguistic study," *UCLA Working Papers in Phonetics* 56. UCLA Phonetics Lab., Los Angeles.

Maddieson, I. 1977. "Tone loans: a question concerning tone spacing and a method of answering it," *UCLA Working Papers in Phonetics* 36, 49-83. UCLA Phonetics Lab., Los Angeles.

Stevens, K.N. 1972. "The quantal nature of speech: evidence from articulatory-acoustic data," In David, E. & Denes, P., eds.,*Human Communication: A Unified View* 56-66.

Stevens, K.N. 1989. "On the quantal nature of speech," *Journal of Phonetics* 17: 3-45.

Traunmuller, H. 1983. *On Vowels: Perception of Spectrofeatures, related aspect of production and sociophonetic dimension.* University of Stocholm, Stoholm.

Weitzman, R. 1990. "The role of the critical band in learning vowel categorizations," presented at the 119th meeting of the Acoustical Society of America.

Zwicker, E. 1961. "Subdivision of the audible frequency range into critical bands," *JASA* 33

Appendix

English foil data

fox	pen	sun	vase
ring	boat	nun	cab
fog	bell	day	love
team	pay	girl	mud

Korean foil data

tʃʰə	'a wife'	tʃəl	'a temple'	sən	'a line'
he	'a sun'	tal	'a moon'	ton	'money'
pe	'a pear'	mos	'a pin'	san	'a mountain'
ke	'a dog'	tʃang	'a market'	non	'a field'

An articulatory study of the features ATR in Akan and emphasis in Arabic

Sook-hyang Lee

Department of Linguistics, the Ohio State University

Abstract

Akan contrasts two sets of vowels, one in which the tongue root is advanced and the larynx lowered ([+ATR]), and another in which the tongue root is retracted ([-ATR]) (Lindau, 1979). Arabic contrasts two sets of consonants, one in which the pharynx is constricted ([+emphasis]) and another in which it is not ([-emphasis]). Lindau suggests that the two phenomena can be combined as various settings along a single phonetic dimension of pharynx width, with [+ATR] as maximally expanded and [+emphasis] as maximally constricted, and that this dimension can be reduced to the binary phonological feature [±expanded], since no language contrasts more than two settings. This paper tests this hypothesis. Measurements of pharyngeal diameter were taken from X-ray tracings from productions by two Arabic speakers and three Akan speakers, and a multivariate analysis of variance was performed. Although emphasis is primarily a consonant feature in Arabic, it is legitimate to compare vowels in this cross-linguistic study, because as noted by Card (1983), [+emphasis] spreads to vowels and consonants within the same word. The results showed significant interaction between the more and less expanded feature values and the two languages, implying that emphasis in Arabic is controlled by a different mechanism from that used for [±ATR] in Akan.

Introduction

This paper compares two features in which pharyngeal width has been implicated -- [±ATR] in Akan and [±emphasis] in Arabic. Akan has a type of vowel harmony where two sets of vowels contrast; one in which the tongue root is advanced and the larynx is lowered, and as a result the pharynx is wide ([+ATR]), and another in which the tongue root is retracted and as a result the pharynx is narrow ([-ATR]) (Lindau, 1979). The vowels of Akan are as follows:

set 1			set 2	
i	u		ɪ	ʊ
e	o		ɛ	ɔ
		a		

The low vowel /a/ is neutral with respect to the vowel harmony. The tongue root mechanism involved with [ATR] is usually combined with vertical displacement of the larynx and sometimes with movements of the back pharyngeal wall and is also independent of the mechanism for controlling tongue height (Lindau, 1978).

Some consonant pairs in Arabic (/ṭ/ - /t/, /ḍ/ - /d/, /ṣ/ - /s/, and /ẓ/- /z/) contrast with respect to the feature "emphasis". [+emphasis] segments are produced by constricting the pharynx by mainly retracting the tongue. [-emphasis] involves no such constriction or retraction. For example, according to Lindau (1979), Ali & Daniloff (1970) note that 'the articulatory mechanism in the emphatic sounds is such that the tongue moves in a way which depresses the palatine dorsum and moves (mesially) the pharyngeal dorsum (tongue root) towards the pharyngeal wall simultaneously'.

Both of these features have thus been described in terms of using the tongue root to either expand or constrict the pharyngeal cavity. They also do not contrast phonologically in any language. These two observations led researchers to claim that only one feature is needed, [ATR] or [Constricted pharynx], etc., to describe them. For example, Lindau suggests that the two phenomena can be combined and regarded as various settings along the single phonetic dimension of pharynx width, with [+ATR] as maximally expanded and [+emphasis] as maximally constricted. This dimension can then be reduced to a binary phonological feature, [+expanded], under the assumption that one condition for combining two features from two different languages into a single cross-linguistic feature is that both features make use of the same phonetic mechanism. (Lindau prefers [expanded] to [ATR] because she found that the feature involves not only tongue root movement but also larynx movement as well. That is, for [+ATR] segments, the tongue root is advanced but the larynx is lowered, resulting in a larger pharyngeal cavity than that for [-ATR] segments.)

However, what Lindau means by the term 'phonetic mechanism' is not clear. Does it mean an articulator such as the tongue body or tongue root, or a muscle or a group of muscles such as the posterior genioglossus and hyoglossus which cause the tongue root to move forward and backward, resulting in variation in pharyngeal width? The following might clarify her idea about the relationship between articulatory mechanisms and features:

> "In both Swedish and Urhobo, the vowels and approximants differ by the use of two separate lip gestures, not by different degrees of the same gestures; So they should be characterized by separate features." (Lindau, 1978:550)

Lindau's statement about [tense] in English and [ATR] in Akan will also help us understand more accurately what she means by 'phonetic mechanism'. She examined the articulatory and acoustic aspects of the feature [tense] in English and concluded that [ATR] and [tense] cannot be combined into a single feature even though [tense] like [ATR] still seems to involve tongue root movement, because in the case of [tense] this movement might be just an artifact of tongue height adjustment for that feature (Perkell, 1971). Her notion of "articulatory mechanism" thus could be interpreted as follows: the same articulators, tongue root and tongue body, are involved in both features, [+ATR] and [tense], but they are controlled by different "phonetic mechanisms", i.e., by a different muscle or a set of muscles. [tense] in English is controlled by the set of muscles that act to raise the tongue body; here the tongue root is advanced as an artifact of this raising movement of the tongue body. By contrast, [ATR] is controlled by a different set of muscles that push the tongue root forward or pull it backward; here the raising and lowering movements of the tongue body are an artifact of the tongue root movement.

Using this same reasoning, we could say that [emphasis] and [ATR] are cross-linguistically different features if they are implemented with qualitatively different tongue root movements and not merely different degrees of the same movement or gesture. This paper tests Lindau's hypothesis, i.e., that [ATR] and [emphasis] are controlled by the same phonetic mechanism and therefore could be collapsed into the same feature.

To test this hypothesis, the pharynx shape of the vowels in these two languages (pharyngealized/nonpharyngealized vowels in case of Arabic, [+ATR], [-ATR] in the case of Akan) were quantified by measuring several distances between the pharyngeal wall and the tongue root in the pharyngeal cavity. Only /ṭ/ - /t/, /ḍ/ - /d/, /ṣ/ - /s/, and /ẓ/ - /z/ have formerly been recognized as consonants which contrast with respect to [emphasis]. However, [+emphasis] spreads so that all segments in the same word, including vowels, are pharyngealized (Card, 1983). Thus, it is possible to contrast vowels directly so that any difference can be attributed to a real difference of features between the two languages and not to artifacts of primary consonant constriction in Arabic. Arabic has a five vowel system like Akan but with a vowel length contrast. However, the low vowel /a/ in Arabic is affected by [emphatic] spreading to become pharyngealized, whereas the /a/ in Akan is neutral with respect to the [ATR] harmony.

Emphatic segments in Arabic are characterized acoustically by a lower second formant (Card, 1983) whereas a lower first formant characterizes [+ATR] segments in Akan (Lindau, 1979). In order to see if emphasized vowels also show lower F2 frequency compared to unemphasized vowels, I made acoustic measurements on a set of Damascan Arabic productions. As can be seen in Fig. 1, [emphasis] on vowels is also characterized by F2 frequency lowering.

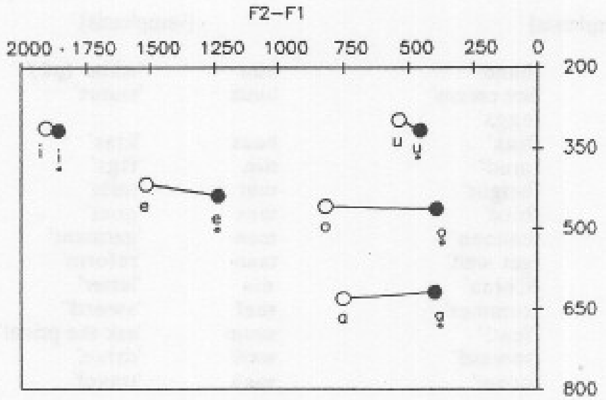


Figure 1. Formant values of emphasized and neutral vowels in Arabic.

Measurements were made from x-ray tracings of three Akan speakers and two Arabic speakers, and statistical analyses were performed on the measurement data.

Two basic claims about the descriptions of these features need to be examined first, however. Are the descriptions of these features in terms of using the tongue root to expand or constrict the pharyngeal cavity accurate? If they are, then the next question is: are the vowels of [+ATR] in Akan and [-emphasis] in Arabic the same in terms of degree of tongue root advancement and also are [-ATR] and [+emphasis] the same? If they are not the same, there would be four different phonetic events regardless of whether all of them can be described in terms of a single phonetic dimension. Finally, this paper tests whether the features are on the same phonetic dimension, or if they are controlled by different phonetic mechanisms.

Methods

The recordings

The study used an X-ray film in sagittal view of the vocal tracts of two Damascan Arabic speakers and existing tracings of the vocal tracts of three Akan speakers from X-ray films made by Lindau (For a more detailed description of the Akan data, see Lindau (1979)). The Arabic film was made by A. Abramson et al. at Haskins laboratory. The corpora used in the Akan and Arabic films are given below;

Akan

[+ATR]			[-ATR]		
fɪ	[fi]	'to leave'	fe	[fɪ]	'to vomit'
bu	[bu]	'to break'	bo	[bʊ]	'to be drunk'
hwie	[ʃɥe]	'to pour'	pɛ	[pɛ]	'to like'
mo	[mo]	'well done'	bɔ	[bɔ]	'to strike'

Arabic[2]

[+emphasis]		[-emphasis]	
ʕiiḍ	'hand'	biid	'white (pl.)'
buuẓ	'ice cream'	buuz	'snout'
beeḍ	'eggs'		
baaṣ	'bus'	baas	'kiss'
ṭiin	'mud'	tiin	'figs'
ṭuul	'height'	tuul	'tulle'
ṭeer	'bird'	tees	'goat'
ṭoob	'cannon'	toob	'garment'
ṭaab	'get well'	taab	'reform'
ṣiin	'China'	siin	'letter'
ṣeef	'summer'	seef	'sword'
ṣuum	'feat!'	suum	'ask the price!'
ṣoob	'toward'	sooʕ	'drive'
ṣaah	'crow'	saah	'travel'

Each verb stem in the Akan corpus was in the frame sentence ka...bio "say ... once more" while the Arabic data were read in citation form. Each word in Akan was repeated three times while no repetition was made in the case of the Arabic words.

There is some difference in the corpora of Akan and Arabic in terms of the consonantal contexts of vowels. Every target vowel in Akan is preceded by a labial consonant. However, in Arabic, X-ray tracings for minimal pairs for some vowels preceded by labial consonant were not available. Data for unpharyngealized [e] and for both pharyngealized [o] and unpharyngealized [o] in a labial consonantal context were not available. For these gaps I substituted measurements of each vowel preceded by the alveolar consonants /t/ and /s/, since data for minimal pairs for every vowel in an alveolar consonantal context were available. A statistical study was conducted to test whether the different preceding consonant could have affected the movement of the tongue for the following vowel, but no significant contextual effects of the consonant on the following vowel were found. The details of this particular test are presented later in this paper.

As mentioned above, the low vowel /a/ in Akan does not participate in vowel harmony while the low vowel /a/ in Arabic is affected by emphatic consonants in the same word. In addition, every vowel in Arabic (with the possible exception of /i/) contrasts with respect to the feature [emphatic]. For most statistical analyses done in the later sections in this paper, measures for /a/ in both languages were excluded from the data. One more difference in the two corpora is that the Arabic data contains only long vowels.

Measurement.

Tracings of the entire vocal tract were prepared from the X-ray video-tape of the two Arabic speakers in the following manner. The whole vocal tract shape on the TV monitor for each vowel was traced on the tracing paper while the tape was on pause. Slowing the tape down to frame-by-frame speed and alternately moving forward and backward on the tape, one frame was chosen as the most steady state for each vowel. Very often around the middle of a vowel there were several identical steady-state frames which occurred at least before the velum had started to move down to the resting position. In those cases, I selected the middle frame as the representative frame for the vowel. For the Akan data, this procedure was not necessary since Lindau had already made tracings using a similar procedure, and photocopies of these tracings were available.

To make measurements, I prepared a template for each speaker as given in Fig. 2, using the following procedure: First, I chose as fixed points some passive articulators - the back wall of the pharynx, the hard palate, Akanthion, and the upper teeth. I drew a vertical line through the back wall of the pharynx (AB) and two horizontal lines perpendicular to it on each tracing for each repetition of each word in both languages. One is from the Akanthion (AC) and another from the highest point of the hard palate (DE). The distance between those two lines for every tracing (AD) was measured and all distances from each speaker were averaged within that speaker. The average value from each speaker was used as a unit for the construction of the template for that speaker. I selected one tracing for each speaker for the representative vocal tract -- the tracing of the first token of the high front [-ATR] vowel /ɪ/ from each Akan speaker and that of high front [-emphasis] vowel /i/ from each Arabic speaker. These were chosen because the average value of the distance between Akanthion and the highest point of the hard palate was closest to that for those vowels. While some tracings deviated from these representative tracings, the deviation was very small in most cases.

The pharyngeal cavity was divided into four sections by drawing four lines perpendicular to the back wall of the pharynx: DF = FH = HJ = JL = LM. Since the Arabic tape does not show the lower part of the pharyngeal cavity around the larynx, lines could not be drawn there. The two lowest sections were divided into two subsections by adding one more line for each section: HN = NJ = JP = PL. These lines were also perpendicular to the pharyngeal wall. Putting on the template for each speaker on the tracing paper for each token, five distances, HI, NO, JK, PQ, LM, were measured with a ruler whose smallest unit was a millimeter. As mentioned above, some tracings deviated from the template, but the deviation was very small. However, there were two cases in Arabic where HI and NO were not available (all of them were pharyngealized [a]) because the tongue root was too far down in the pharynx. As a result, I could not use for the statistical analysis the measurement data on those two distances which might give us the information about the shape around the upper pharynx. As an alternative, I chose two more distances (DT and RU): DT bisects ∠EDF, and RU bisects ∠SRB. RS runs perpendicular from the upper teeth to the pharyngeal wall. In sum, five distances around the pharynx were measured in total: DT, RU, JK, PQ, and LM (I will call them Distance 1, Distance 2 , Distance 3, Distance 4, and Distance 5, respectively).

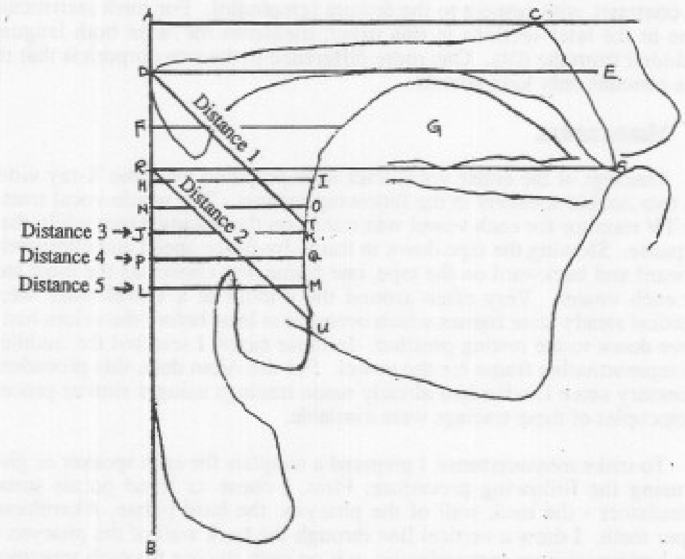


Figure 2. Construction of 5 distances: DT, Distance 1; RU, Distance 2; JK, Distance 3; PQ, Distance 4; LM, Distance 5.

Results

Consonantal contextual effect on vowels

As mentioned above, in Arabic, at least three different consonants precede the target vowel in the corpus, -- the bilabial stop /b/, the dental stop /t/, and the dental fricative /s/ -- while in Akan, labial consonants always precede the target vowel. Several researchers (e.g., Stevens & House, 1955), have reported a consonantal contextual effect on vowels. If this contextual effect is present in Arabic, a straightforward comparison of some Akan and Arabic vowels would not be possible. If there is an effect, I need to deduce the effect of the labial consonant on the vowel /e/ or /i/ for which the labial context is missing by comparing the effect of labial consonant and dental consonant on the other vowels and limit the corpus to the words or phrases which start with the labial consonant.

In order to test whether different consonantal contexts affect the following vowels significantly with respect to [emphasis], I performed an Analysis of

Variance and also a Multivariate Analysis of Variance on the measures for the three vowels other than /e/ and /o/. The five distances were the dependent variables and the three different consonants were the independent variables. If the contextual effects of consonants on the following vowels are significant, there should be significant main effect of consonantal context. ANOVA and also MANOVA showed no significant main consonantal effect on any distances ($\alpha = 0.01$). Based on this result, I conclude that different consonantal context would not affect the results of the statistical analyses I will perform later in this study.

The vowel /i/

Based on formant measurements, Card (1983) reports that /ii/ and some consonants in Palestinian Arabic are not affected by the emphatic consonants and block the spreading of emphasis; the second formant of /ii/ is not lowered and the second formants of the preceding or the following segments in the same word are not lowered even in the emphatic environment. To see whether /ii/ in Damascan Arabic is affected by the emphatic consonant, I only took data for Arabic /ii/'s and performed an ANOVA and MANOVA. There were no significant differences between /ii/ in the non-emphatic condition and /ii/ in the emphatic condition, demonstrating that /ii/ in Arabic is not affected by emphatic consonants. (i.e. it does block the spread of emphasis) The image tracings of two vowels given in Appendix show almost no difference in vocal tract shape, especially in the pharyngeal cavity. Therefore, measures for the high front vowel /i/ and /ii/ in both languages were excluded from the data for the statistical analyses done in the later sections.

Characterizing the features in qualitative terms

As mentioned above, measures for the low vowel /a/ and the high front vowel /i/ in both languages were excluded from the data for the statistical analyses done in this and later sections in this paper because of neutrality of /a/ with respect to vowel harmony in Akan and neutrality of /i/ with respect to [emphasis] in Arabic. To verify whether previous descriptions of [ATR] as an expansion of the pharyngeal cavity and [emphasis] as a constriction of the pharyngeal cavity are accurate, separate ANOVA and MANOVA were performed for each language on the same measurements as before with the five distances as the dependent variables and the two values of the feature for the language as the independent variables. If the descriptions of these features are accurate, there should be a significant main effect of feature and the mean should be smaller for the more constricted value.

As can be seen in Fig. 3, the main effect of feature for all distances was significant in Akan. In Arabic, the main effects of feature for all distances except for the 2nd distance were significant and the main effect in the MANOVA was also significant ($P < 0.0001$).

Note, however, that the means for distance 1 in Arabic do not differ in the direction we would expect. Given the description that pharyngealized vowels show a greater constriction in the pharynx, they should have smaller means for all relevant distances. Instead, the mean for this distance for [+emphasis] was larger than that for [-emphasis]. This may be because the tongue root moves backwards and simultaneously lowers farther down for [+emphasis] in Arabic. Even though the measurement data for /a/ in both languages are not included in this study because of the neutrality of /a/ with respect to the ATR harmony in Akan, there were two cases in Arabic where the tongue root lowers so much that I could not get the value of distances HI and NO which are right above the distance 3 in Fig.2. On the other hand, all five distances in Akan differ in the direction we would expect, i.e. means of distances for [+ATR] are larger than those for [-ATR].

From these results we could conclude that the descriptions of [ATR] and [emphasis] are generally accurate: the tongue root is involved in these features. However, there seems to be some difference with respect to which part of the pharynx is constricted by these features: [+emphasis] in Arabic involves constriction in the lower pharynx whereas [-ATR] in Akan involves constriction of the whole pharynx. This difference seems to imply that these two features are controlled by different phonetic mechanisms, which we test in section 3.5. in this paper.

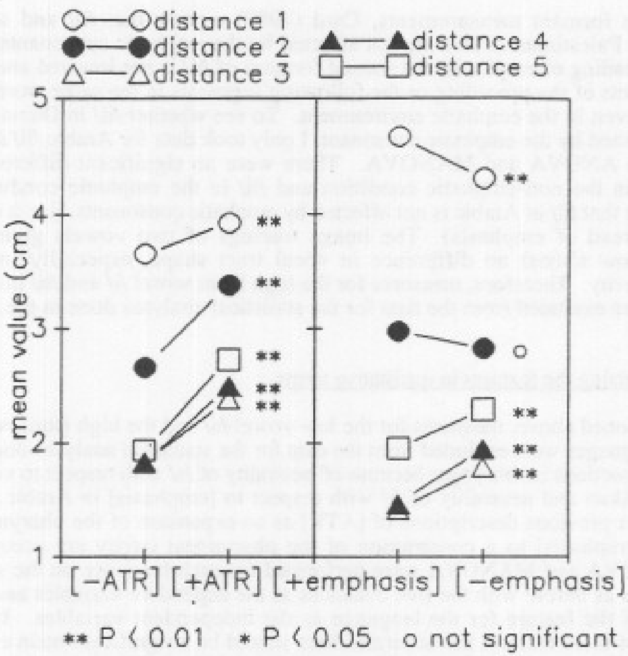


Figure 3. Effects of features, [ATR] in Akan and [emphasis] in Arabic.

<u>Comparing wide and narrow pharynx values across the languages.</u>

The next question to be raised is whether [ATR] and [emphasis] are ultimately the same feature or not: Are [+ATR] and [-emphasis], both characterized by a wide pharynx, the same? Are [-ATR] and [+emphasis], both characterized by a narrow pharynx, the same? If the members of the pairs above are implemented differently, then there could be four levels of pharyngeal width or shape, regardless of whether they are on the same phonetic dimension or not. In order to test this, an ANOVA and MANOVA were performed on the same measurement data above but it is organized differently. The five distances were dependent variables and the two languages within each value of two (wide pharynx and narrow pharynx) were independent variables. If [+ATR] and [-emphasis] or [-ATR] and [+emphasis] are the same, there should no main effect for either the wide-pharynx feature or the narrow-pharynx feature.

As shown in Fig.4, the effect of language within the wide-pharynx vowels was significant for all distances except for the 5th, and the effect of language in the MANOVA was also significant. The effect of language within the narrow-pharynx vowels for all distances were significant. The effects of language in the MANOVA was also significant (P < 0.0001).

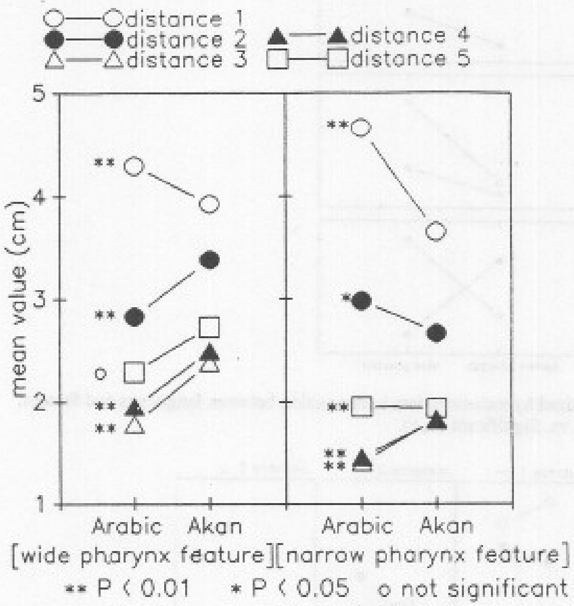


Figure 4. Wide and narrow pharynx values across Akan and Arabic.

Based on these results, we could conclude that [+ATR] and [-emphasis] are different from each other with respect to the five distances around the pharynx constructed for this study. [-ATR] and [+emphasis] are different as well. In other words, there are four levels of pharyngeal width. Once four different levels of pharyngeal width are established, the next question to be raised is whether those four levels are in the same phonetic dimension or not, as proposed by Lindau.

<u>Comparing the contrasts across the languages</u>

In order to test the hypothesis that [ATR] and [emphasis] are controlled by the same phonetic mechanism such that they are phonetically the same or they are in the same phonetic dimension, an ANOVA was performed on the same measurement data with the five distances as the dependent variables and language and feature as independent variable. If [ATR] in Akan and [emphasis] in Arabic contrast phonologically in each language, and these features are controlled by the same phonetic mechanism, we could expect a statistically significant main effect of language but an insignificant interaction between language and feature as given hypothetically in Fig. 5a. On the other hand, if they are controlled by a different phonetic mechanism, we should expect both a significant main effect of language and a significant interaction between language and feature as exemplified in Fig. 5b and 5c.

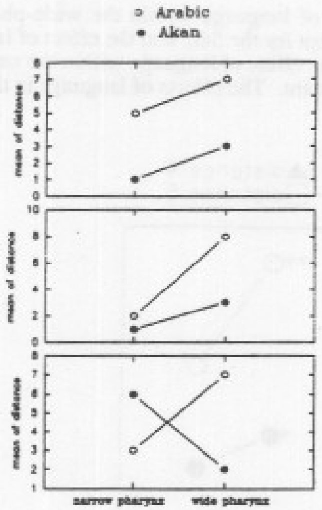


Figure 5. Schematized hypothetical data in interaction between languages and feature: Not significant (a) vs. Significant (b, c).

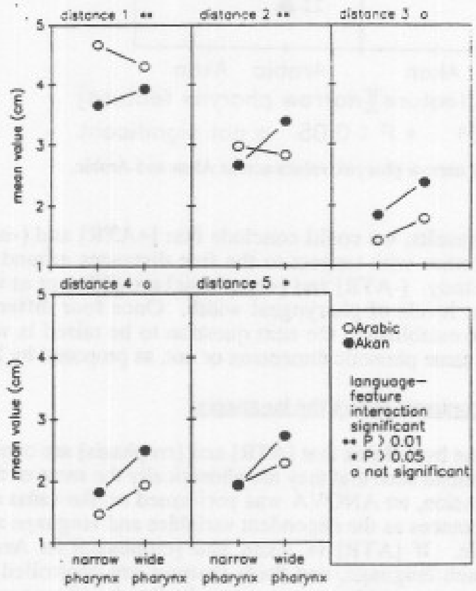


Figure 6. Interaction between language (Akan and Arabic) and feature ([ATR] and [emphasis]).

As in section 3.3, there was a significant main effect of language. As shown in Fig. 6, the 1st, 2nd and 5th distances gave a significant interaction between language and feature. The MANOVA also yielded a significant interaction between language and feature ($P < 0.0001$).

These results seem to demonstrate that Lindau's hypothesis that [ATR] and [emphasis] involve the same phonetic mechanism is not correct. Rather, these features are controlled by some different phonetic mechanisms. The patterns described in section 3.3 suggest that they might involve different parts of the pharynx.

Discussion and Conclusion

In this paper, I have tested Lindau's (1979) hypothesis that [ATR] in Akan and [emphasis] in Arabic are controlled by the same phonetic mechanism by taking measurements from x-ray tracings of some distances between the tongue root and the back wall of the pharynx in sagittal view.

Before I tested this hypothesis, I first tested whether the previous descriptions of these features in terms of using the tongue root to expand or constrict the pharyngeal cavity were correct. Statistical analysis of the articulatory measurement data done in this study confirmed the description. I then examined whether [+ATR] and [-emphasis] are phonetically the same value, and likewise for [-ATR] and [+emphasis]. Statistical analysis of the measurement data shows that they are different from each other, and that there are four levels of pharyngeal width involved.

Finally, Statistical analysis of the measurement data strongly suggests that [ATR] in Akan is controlled by a different phonetic mechanism from [emphasis] in Arabic. [ATR] and [emphasis], therefore, cannot be combined and regarded as various settings along a single phonetic dimension of pharynx width as Lindau proposed.

In order to figure out what mechanisms control [ATR] in Akan and [emphasis] in Arabic, we need further study. Investigating which muscle or muscles are active with these features using EMG, for example, might illuminate the mechanism.

Notes

*The original version of this paper was presented in the 119th meeting of the Acoustical Society of America, State College, Pennsylvania, May 1990.

1. According to Card (1983), the four emphatic obstruents are : /ṭ, ḍ,ṣ ,ð̣ / In Classic Arabic. /ẓ/ and /z/ are used in many urban colloquial dialects in place of /ð̣/ and /ð/ in Classic Arabic.

2. Dots under the consonants in Arabic data indicate that they are primary emphatic consonants.

References

Al-Ani, S. H. 1970. *Arabic Phonology: An Acoustical and Physiological Investigation*, Mouton, Hague.

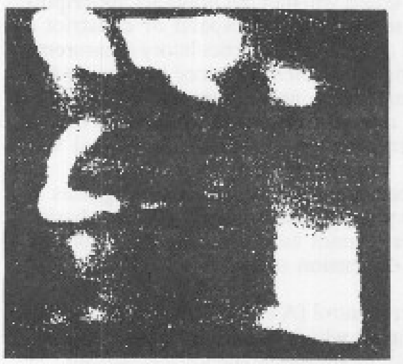
Ali, L. & Daniloff, R. 1970. 'Emphatic Arabic sounds: a mechanism and evidence of co-articulation,' *J. Acoust. Soc.Am.* 49, Supp. 9(1).

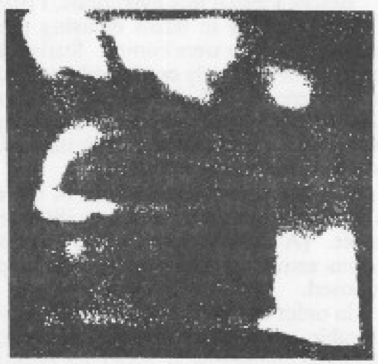
Card, E.A. 1983. *A Phonetic and Phonological Study of Arabic Emphasis*, Ph.D. dissertation, Cornell University.

Halle, M. & Stevens, K.N. 1969. 'On the feature Advanced Tongue Root,' *Quarterly Progress Report* 94, 209-215. Research Laboratory of Electronics, Massachussetts Institute of Technology.
Lindau, M. 1978. 'Vowel features,' *Language* 54: 541-563.
Lindau, M. 1979. 'The feature expanded,' *Journal of Phonetics* 7: 163-176.
Perkell, J. 1971. 'Physiology of speech production: a preliminary study of two suggested revisions of the features specifying vowels,' *Quarterly Progress Report* 102: 123-139. Research Laboratory of Electronics, Massachusetts Institute of Technology.
Stevens, K. N. & House, A. S. 1955. 'Development of a quantitative description of vowel articulation,' *J. Acoust.Soc. Am.*, 27: 484-493.

Appendix Image scans of /ii/ of /ʕiid/ and /biid/ in Arabic.

/ʕiid/ (neutral condition) /biiḍ/ (emphatic condition)





Lip rounding in Amoy and Mandarin high vowels: maximum dispersion, or adequate separation.

Ho-hsien Pan

Department of Linguistics, The Ohio State University

Abstract

There are two hypotheses about the relationship between phonological contrasts and phonetic feature scales. Some phoneticians propose that values are chosen so that contrasting phonemes are maximally separated, e.g., Lilijencrants and Lindblom, 1972, while others claim that they need only to be adequately separated, e.g., Maddieson, 1977. This paper tests the competing hypotheses by comparing lip position in Mandarin [i], [y], [u] with that of Amoy [i] and [u]. According to adequate separation, the lip spreadness/roundness of Mandarin will be more extreme than that of Amoy, since there are three high vowels in Mandarin but only two in Amoy. According to the maximum dispersion hypothesis, the degree of roundness should be the same in both languages. Amoy and Mandarin data were collected from three bilingual speakers. The results support the *Adequate Separation Theory*. This paper also tests Wood's (1986) claim that in a language with two high rounded vowels, /u/ and /y/, /u/ is more rounded than /y/. The result shows that this claim is not necessarily true.

Introduction

There are two competing theories for predicting the positions of phonological entities within phonetic space. According to Liljencrants and Lindblom's (1972) *Maximal Dispersion Theory*, phonological entities are maximally separated. Thus in a language with two high vowel phonemes, /i/ and /u/, maximal dispersion predicts that the vowels will spread maximally apart to occupy opposite corners of the vowel space. If there is a third high vowel, say /y/, the first two vowel will still be maximally separated, occupying the same peripheral positions as shown in (1):

(1)	2 vowel system:	i		u
	3 vowel system:	i	y	u

Adequate Separation Theory (Maddieson, 1977), by contrast, predicts that there should be some fixed interval between adjacent vowels, as phonological entities are "as separated as they need to be", therefore /i/ and /u/ in a two-vowel system can be closer together than in a three-vowel system as shown in (2):

(2)	2 vowel system:	i		u
	3 vowel system:	i	y	u

In this study, lip rounding of high vowels is used to test both theories. The degree of lip rounding influences F2; the more rounded a vowel is, the lower the

F2, and the more spread the lips are, the higher the F2. Thus, the degree of lip rounding can be used as a crude measure of the spacing of vowels in the F2 dimension of the vowel space. Acoustic measurement of F2 will be carried out in follow up study to compare with the articulatory data here. In addition to testing the two hypotheses, Wood's (1986) claim about the degree of lip rounding of /u/ and /y/ in a language is also tested. In his study, Wood claims that in a language with two high rounded vowels, the back rounded vowel will be more rounded than front rounded vowel. Thus Mandarin /u/ would be more rounded than Mandarin /y/.

Experiment

Mandarin is a language with three high vowels /i/, /y/, and /u/. Amoy is a language with two high vowels /i/ and /u/. The study used three subjects, JRG, CK (male), and CYT (female), all bilingual speakers of both languages. (Bilingual speakers were used to control for speaker variability in lip size while testing for inter-language differences.) The corpus is given in Table I. The Mandarin words are all in fourth tone which is a falling tone. The Amoy words are all in second tone, which is also a high falling tone. Tokens from each language are repeated five times. For each repetition, the tokens are randomized to avoid context effect.

Table I. Corpus

Amoy		Mandarin	
/i/	'chair'	/i/	'easy'
/li/	'you'	/li/	'stand'
		/y/	'jade'
		/ly/	'discipline'
/u/	'small island'	/u/	'object'
/lu/	'female'	/lu/	'road'

Method

A video recording of a speaker's face was made simultaneously with the audio recording. Subjects were asked to sit with their heads against a wall covered with a paper marked with 1 cm squares. This background grid was for detecting camera distortion, and for scaling the measurements from the TV into a constant proportion for the three subjects. A mirror was placed at a 45 degree angle at the right side of the subject to capture the side view, as shown in Figure 1. There are four reference points drawn on the subjects' faces; one on the nose, one on the center of the chin, one on the right side of the jaw, and one beneath the right ear.

Measurements from the frontal view are made of the following distances as shown in Figure 2. Measurements preceded by triple asterisks are statistically significant ($p< 0.05$):

***Width between outer corners (measurement BC)
***Width between inner corners (IJ)
Distance from nose reference to lip line (perpendicular from A to BC)
***Vertical compression -- distance from vermillion border of upper lip to inter--lip line (perpendicular from D to BC)
Vertical compression -- distance from inner edge of upper lip to inter--lip line (perpendicular from E to BC)

***Vertical compression -- distance from inner edge of lower lip to inter--lip line (perpendicular from F to BC)
***Vertical compression -- distance from vermillion border of lower lip to inter--lip line (perpendicular from G to BC)
Vertical compression -- distance from front jaw reference to inter--lip line (perpendicular from H to BC)
***Distance from vermillion border of upper lip to right outer corner (DC)
***Distance from inner border of upper lip to right outer corner (EC)
***Distance from inner border of lower lip to right outer corner (FC)
***Distance from vermillion border of lower lip to right outer corner (GC)
***Distance from front jaw reference to right outer corner (HC)
***Distance from vermillion border of upper lip to left outer corner (DB)
***Distance from inner border of upper lip to left outer corner (EB)
***Distance from inner border of lower lip to left outer corner (FB)
***Distance from vermillion border of lower lip to left outer corner (GB)
***Distance from front jaw reference to left outer corner (HB)
***Distance from vermillion border of upper lip to right inner corner (DJ)
***Distance from inner border of upper lip to right inner corner (EJ)
***Distance from inner border of lower lip to right inner corner (FJ)
***Distance from vermillion border of lower lip to right inner corner (GJ)
***Distance from front jaw reference to right inner corner (HJ)
***Distance from vermillion border of upper lip to left inner corner (DI)
***Distance from inner border of upper lip to left inner corner (EI)
***Distance from inner border of lower lip to left inner corner (FI)
***Distance from vermillion border of lower lip to left inner corner (GI)
***Distance from front jaw reference to left inner corner (HI)

Figure 1. Setting of the experiment: subjects with marks on the face, sits in front of grid papers with a 45 degree mirror showing the side view.

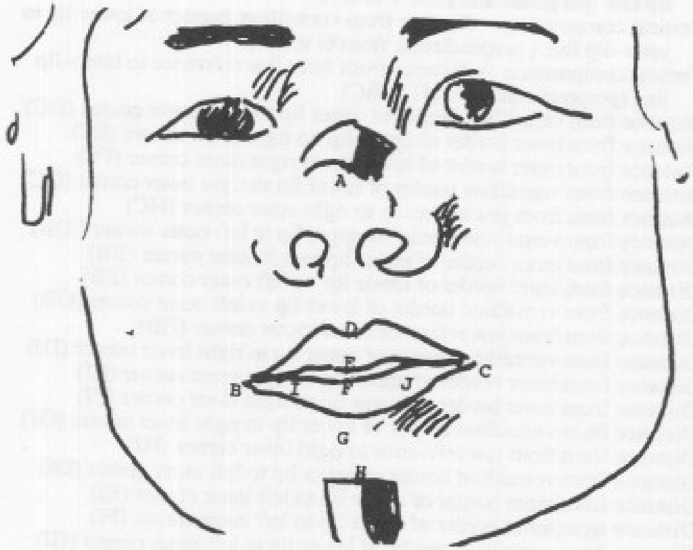


Figure 2. 10 points for the measurement in the front view.

Following are measurements from the sagittal view as shown in Figure 3.:
For lip compression, following measurements were taken:

> ***Distance between vermillion border of upper lip to vermillion border of lower lip

For lip protrusion, following measurements were taken.

> Distance from the reference under the ear to the vermillion border of upper lip (12)
> ***Distance from reference under the ear to the outer corner of lip (13)
> Distance from reference under the ear to the vermillion border of lower lip (14)
> ***Distance from reference under the ear to inner corner of lip (17)
> ***Distance from inner corner of lip to the inter--lip line (perpendicular from 3 to 24)
> Distance from outer corner of lip to the inter--lip line (perpendicular from 7 to 24)
> Distance from reference under ear to the inter--lip line (perpendicular from 1 to 24)

Other measurements taken were:

> Jaw position --- distance from reference under the ear to the side jaw reference

Vertical position of jaw -- distance from nose reference to lateral jaw reference (65)
Vertical position of jaw -- distance from 5 to horizontal line passing through nose reference (perpendicular from 5 to horizontal line 6x -> 56x)
Vertical position of lower lip -- distance from nose reference to vermillion border of lower lip (64)
Vertical position of lower lip -- distance from vermillion border of lower lip to horizontal line passing through nose reference (perpendicular from 4 to horizontal line 6x -> 46x)

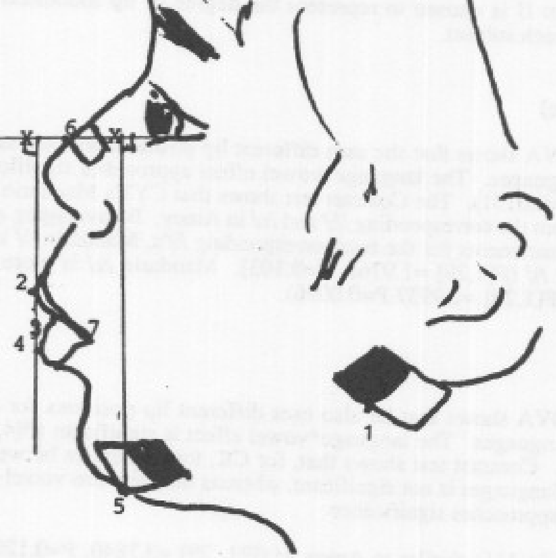


Figure 3. 7 points and a reference line for the measurement in the side view.

Data Analysis

A two-way ANOVA (language * vowel) was done on each measurement separately. Only those measurements which showed a significant ($p<0.01$) or nearly significant ($p<0.05$) language by vowel interaction is discussed in detail below. They are BC, IJ, DBC, FBC, GBC, DC, EC, FC, GC, DB, EB, FB, GB, HB, DJ, EJ, FJ, GJ, HJ, DI, EI, FI, GI, HI, 13, 17, 24, and 324.

A MANOVA was done for each subject to determine whether the measures are correlated with difference in lip configuration for the corresponding vowels in two languages. A Contrast test was also done to test the difference between the corresponding vowels in two languages.

Results and Discussion:

Even though for one out of three subjects the difference between corresponding vowels in two languages is not significant, generally the trend for all three subjects is that Mandarin /i/ to be more spread than Amoy /i/ and Mandarin /u/ is more rounded than Amoy /u/. In other words the *Adequate Separation Theory* is supported to be more correct in this experiment. Sidney Wood (1986) claimed that in a language with two high rounded vowel, /u/ is more rounded than /y/. Thus, Mandarin /u/ should be more rounded than /y/. By comparing all the measurements, it was found that this is not necessarily true. Again Mandarin IJ (width between inner corners of lips) is used here to represent the degree of roundness of different vowels and this is shown in Figure 4. In this figure, measurement IJ is chosen to represent the degree of lip roundness for different vowels for each subject.

CYT (female)

MANOVA shows that she uses different lip positions for the "same" vowel in the two languages. The language*vowel effect approaches significance ($F[4,87] =1.4704$, $P=0.0131$). The Contrast test shows that CYT's Mandarin /i/ and /u/ are different from the corresponding //i/ and /u/ in Amoy. By averaging and comparing all the measurements for the two corresponding /i/'s, Mandarin /i/ is more spread than Amoy /i/ ($F[1,29] =1.9767$, $P=0.103$). Mandarin /u/ is more rounded than Amoy /u/ ($F[1,29] =1.9937$ $P=0.0096$).

CK (male)

MANOVA shows that he also uses different lip positions for each vowel in different languages. The language*vowel effect is significant ($F[4, 116]=1.7198$, $P=0.0001$). Contrast test shows that, for CK, the difference between vowel /i/'s in the two languages is not significant, whereas that between vowel /u/'s in the two languages approaches significance.

Mandarin /i/ is similar to Amoy /i/ ($F[1, 29] =3.7840$, $P=0.1208$). However, by adding and comparing all the measurements, Amoy /i/ is less spread than Mandarin /i/. As for /u/, Mandarin /u/ is more rounded than Amoy /u/ ($F[1, 29] = 1.8373$, $P=0.0190$).

JRG (male)

MANOVA shows that he seems to use the same vowel in both languages. There is no significant language* vowel effect ($F [4, 145]=1.455$, $P=0.1575$). Contrast test shows that the difference between /i/'s in both languages is not significant, while, that between /u/'s is significant at a 0.05 level. Mandarin /i/ is similar to Amoy /i/ ($F[1, 29]=0.9922$, $P=0.4923$). Mandarin /u/ is less rounded than Amoy /u/ ($F [1, 29]=1.6983$, $P=0.0364$).

The comparison between Mandarin and Amoy in terms of lip roundness is described for each of three subjects.

In Sum, even though for one out of three subjects the difference between corresponding vowels in two languages is not significant, generally the trend for all three subjects is that Mandarin /i/ is more spread than Amoy /i/ and Mandarin /u/ is more rounded than Amoy /u/. In other words, the *adequate separation theory* is supported to be more correct in this experiment.

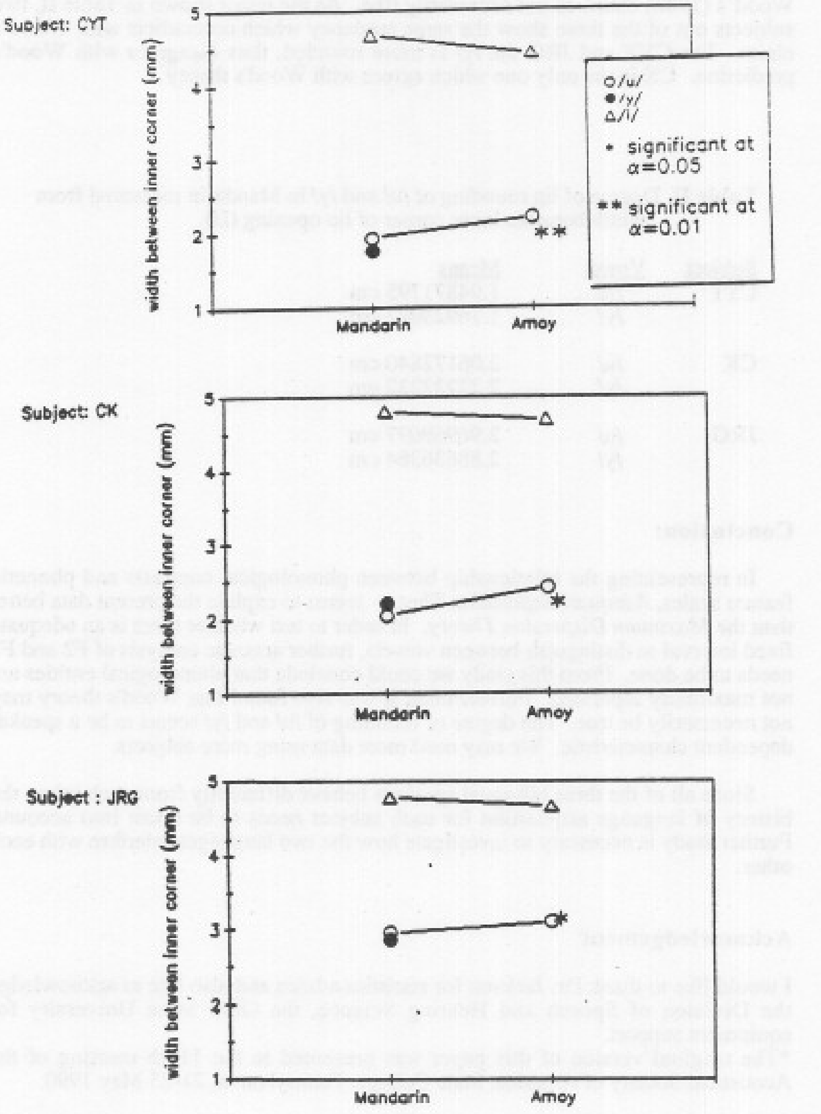


Figure 4. Measurement IJ for three vowels, /u,y,i/, plotted for Mandarin and Amoy for three subjects. Here * indicates significant at $\alpha = 0.05$ and ** at $\alpha = 0.01$.

Next, Sidney Wood's (1986) claim that, in a language with two high rounded vowel, /u/ is more rounded than /y/ is tested using Mandarin data. By comparing all the measurements for /u/ and /y/ in Mandarin, it was found that Wood's (1986) claim is not necessarily true. As the result shown in Table II, two subjects out of the three show the same tendency which contradicts with Wood's claim. For CYT and JRG the /y/ is more rounded, thus disagrees with Wood's prediction. CK is the only one which agrees with Wood's theory.

Table II. Degree of lip rounding of /u/ and /y/ in Mandarin measured from width between inner corner of lip opening (IJ).

Subject	Vowel	Means
CYT	/u/	1.94871795 cm
	/y/	1.76923077 cm
CK	/u/	2.06172840 cm
	/y/	2.22222222 cm
JRG	/u/	2.96969697 cm
	/y/	2.88636364 cm

Conclusion:

In representing the relationship between phonological contrasts and phonetic feature scales, *Adequate Separation Theory* seems to explain the present data better than the *Maximum Dispersion Theory*. In order to test whether there is an adequate fixed interval to distinguish between vowels, further acoustic analysis of F2 and F3 needs to be done. From this study we could conclude that phonological entities are not maximally separated. Furthermore, it was also found that Wood's theory may not necessarily be true. The degree of rounding of /u/ and /y/ seems to be a speaker dependent characteristic. We may need more data using more subjects.

Since all of the three bilingual speakers behave differently from each other, the history of language acquisition for each subject needs to be taken into account. Further study is necessary to investigate how the two languages interfere with each other.

Acknowledgement

I would like to thank Dr. Jackson for statistics advice and also like to acknowledge the Division of Speech and Hearing Science, the Ohio State University for equipment support.
*The original version of this paper was presented in the 119th meeting of the Acoustical Society of America, State College, Pennsylvania, 21-25 May 1990.

Bibliography

Liljencrants, J. & Lindblom, B. 1972. "Numerical simulation of vowel quality systems: the role of perceptual contrast", *Language* 48: 840--862.

Lindblom B. & Sunberg, J. 1971. "Acoustical consequences of lip, tongue, jaw and larynx movement", *JASA* 50: 1166-1179.
Maddieson, I. 1977. "Tone loans: a question concerning tone spacing and a method of answering it", *UCLA Working Paper in Phonetics* 36: 49-69.
Wood, S. 1986. "The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels", *JASA* 80 (2): 391-401.

The duration and perception of English epenthetic and underlying stops.

Sook-hyang Lee

Department of Linguistics, The Ohio State University

Abstract

In American English, an intrusive stop occurs before the fricative in words such as *tense* and *false*, making them very much like words with underlying stops, such as *tents* and *faults*. Ohala (1975) treats the inserted stop as an artifact of universal physiological or aerodynamic constraints. But this approach can't account for the fact that South African English speakers don't insert the stop between sonorant and fricative clusters (Fourakis and Port, 1986). Another approach posits a language- or dialect-specific phonological rule which inserts a phonological segment (Zwicky, 1972). Fourakis and Port (1986), argue against this approach on the ground that in some pairs the intrusive stop is significantly shorter than the underlying one (although the difference is always very small). This paper presents perception data and duration measurements supporting something like Zwicky's approach. Phrases with intrusive and underlying stops (*intense* and *in tents* , respectively) in citation forms produced by three speakers of Mid-Western dialects were presented over earphones in random order for subjects to identify. Identification was very poor, just at chance level. Also, duration measurements of the silence gap between the /n/ and /s/ in these words show no significant difference, contrary to Fourakis and Port's findings. Moreover, token judgments in the perception experiment show very poor correlation with the durations except for one speaker, implying that whatever duration differences there might not be a crucial cue that listeners exploit for labeling the words with epenthetic and underlying stops.

Introduction

The mapping between discrete phonological representations and the continuous speech event has been a controversial subject in phonetics. The occurrence of epenthetic stops in American English is one phenomenon in which there may be a discrepancy between the two representations. As shown in Fig. 1, in American English, an epenthetic stop occurs before the fricative in words such as dense and false, making them very much like words with underlying stops, such as dents and faults. Fourakis and Port (1986) measured the durations of these silent gaps in a cross-dialect study, and found, first, that in South African English, the epenthetic stops do not occur, and second, that in American English their durations are very slightly shorter than those of underlying stops.

Zwicky treats the epenthetic stop as the output of a dialect-specific phonological insertion rule. For example, in the case of *dense*, the silent gap between /n/ and /s/ is produced by inserting /t/ between them in the syllable coda position. On the other hand, Ohala treats the epenthetic stop as an artifact of universal physiological or aerodynamic constraints. According to him, the velum moves quickly compared to the tongue. Therefore, for *dense*, the velum raises for the following fricative /s/ before tongue tip closure for the nasal /n/ is released.

However, these two approaches have problems accounting for Fourakis & Port's findings: Zwicky's phonological approach cannot account for their finding that the duration of the epenthetic stop closure is significantly shorter than that of an underlying stop. Ohala's universal phonetic approach cannot account for their finding that in South African English, epenthetic stops do not occur.

In order to explain the significantly shorter duration of epenthetic stops, what they call "incomplete neutralization", Fourakis & Port suggest that the epenthetic stops are products of language specific phonetic rules, or phase rules, which have access to phonological structure. Ali et al. in an earlier study suggested three different ways in which the timing of gestures involved in /ns/ sequences could result in the appearance of an epenthetic stop.

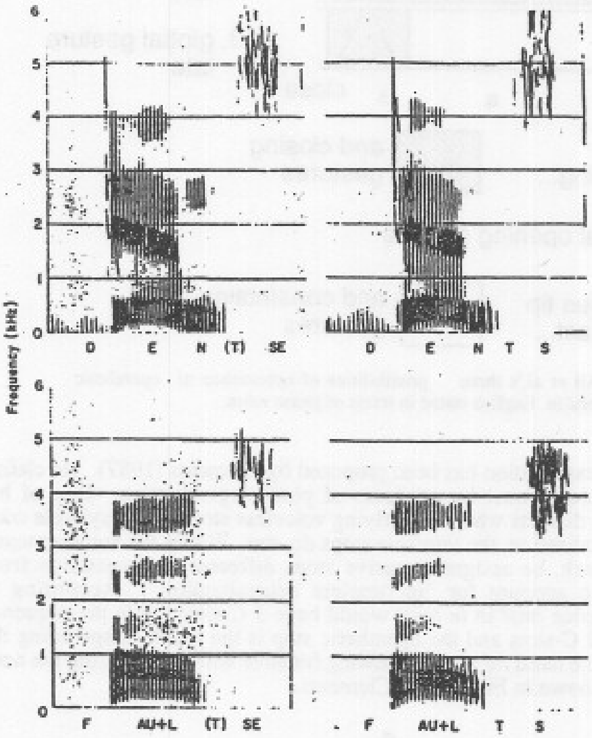


Figure 1. Spectrograms of American productions of *dense-dents* and *false-faults* from Fourakis and Port (1986).

If we think of the articulation of an /ns/ sequence as an arrangement of gestures as in Fig. 2, we can state Ali et al.'s three possibilities in terms of phase rules which affect one or another gestures. First, the tongue tip gesture might be phased late with respect to the other relevant gestures. This would result in a short oral stop articulation. Second, the velic aperture gesture might be phased late. The venting of the air through the nose might prevent the buildup of air pressure behind oral constriction necessary for the onset of turbulence. Third, the glottal gesture might be phased late. The lesser flow of air through the still adducted vocal folds might cause a similar delay in the buildup of air pressure necessary for [s] frication.

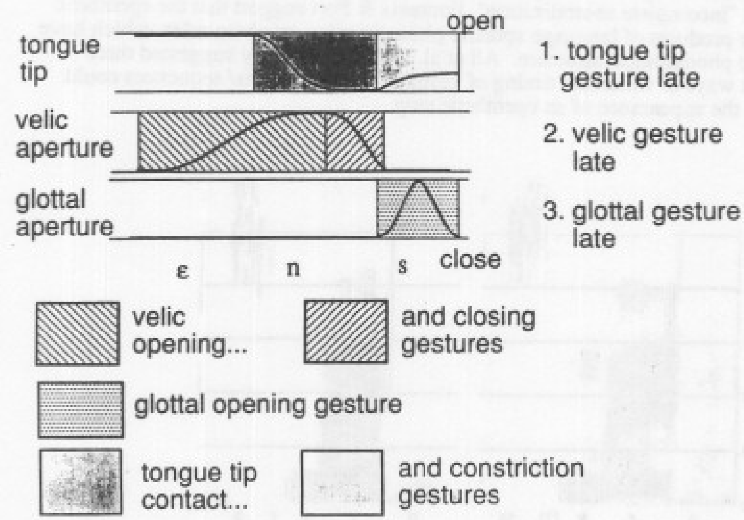


Figure 2. Ali et al.'s three possibilities of occurrence of epenthetic stops in American English stated in terms of phase rules.

An alternative explanation has been proposed by Clements (1987). He claims that the intrusive stops must be products of phonological rules. One of his arguments is that in dialects where underlying voiceless stops in the syllable coda trigger rules of glottalization, the intrusive stops do, too. Within the autosegmental phonology framework, he assigns intrusive stops different representation from underlying stops to account for "incomplete neutralization". According to Clements, the sequence /nts/ in *in tents* would have 3 C-slots while the sequence /ns/ in *intense* has 2 C-slots and the epenthetic stop is the result of spreading the oral cavity node of the nasal /n/ to the following fricative without delinking the node for the fricative as shown in Fig. 3 from Clements.

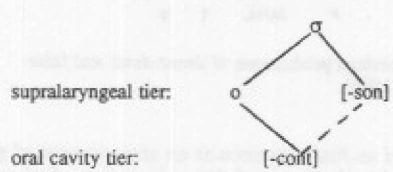


Figure 3. Oral cavity node spreading rule which derives epenthetic stops in /-ns/ sequences at syllable coda position in American English. [From Clements, 1987)

As a result, duration differences between underlying stops and epenthetic stops are automatically explained by their different surface forms. The underlying stop occupies its own C-slot as in Fig. 4a, whereas the epenthetic stop shares in a sense the first half of the /s'/ s C-slot as in Fig. 4b.

This paper reports on perceptual and durational evidence which might shed light on the phonetic and phonological status of epenthetic stops. If the underlying and epenthetic stops show a statistically significant difference and listeners can tell tokens of intense and in tents apart, the two stops are phonetically different. If there is no significant duration difference but listeners can distinguish them, then duration is not the salient perceptual cue and some other perceptual cue(s) might play a crucial role. If the duration difference between two types of stops is significant but listeners cannot tell the two tokens apart, then listeners are insensitive to the stop duration difference even though they are phonetically different. If there is no significant duration difference and listeners cannot tell them apart, we can say that they are phonetically the same. Finally, the correlation between stop duration and perception of epenthetic vs. underlying stops was calculated to see how much listeners depend on the stops' duration during the perception experiment.

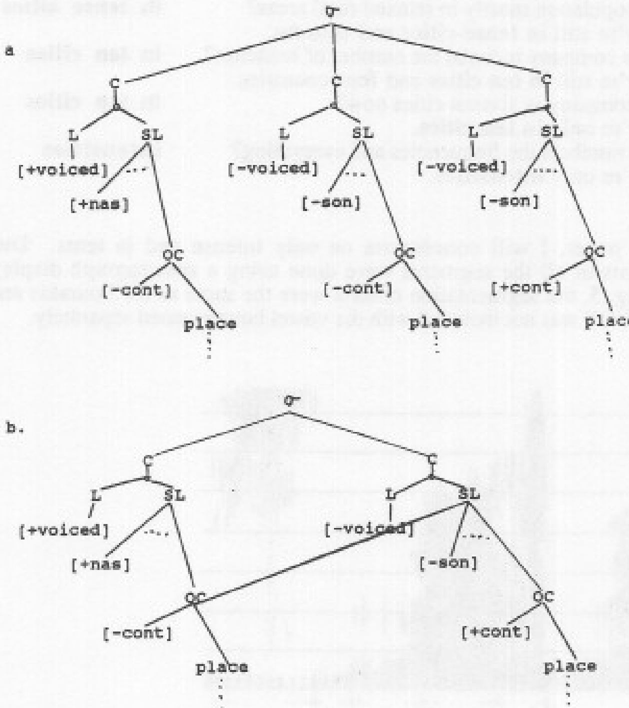


Figure 4. Differences in surface forms between /-nts/ and /-ns/ sequences at syllable coda position in American English.

Experiment 1: Acoustic duration of stops

Method

In the first experiment, duration measurements of segments, including epenthetic and underlying stops, were made. The corpus consisted of the 7 short dialogues shown in Table 1. In a sound-proof booth, one male and two female speakers of Mid-Western dialects read the corpus which had been prepared so that there were 6 repetitions of each type at 3 loudness levels in a randomized order. The subjects read each second sentence in a dialogue and then repeated the target word or phrase in citation form. For each type, this yielded 18 tokens of citation forms and 18 of noncitation forms per speaker.

Table I. Corpus

Noncitation forms	Citation forms
1. Isn't the light too glaring here? Yes, it's quite **intense** and bright.	**intense**
2. Aren't Mongolians nomads? Yes, they live and yurts.	**in tents**
3. Aren't the refuges in houses now? No, they're still in tent cities and shacks.	**in tent cities**
4. Isn't the population mostly in relaxed rural areas? No, they're still **in tense cities** and suburbs.	**in tense cities**
5. Hasn't the company reduced the number of branches? No, they're still **in ten cities** and four countries.	**in ten cities**
6. Isn't the company in sixteen cities now? No, they're only **in ten cities**.	**in ten cities**
7. Are these numbers the frequencies and everything? No, they're only **intensities**.	**intensities**

In this paper, I will concentrate on only intense and in tents. Duration measurements of all the segments were done using a spectrograph display. As shown in Fig. 5, the segmentation criteria were the same as for Fourakis and Port except that VOT was not included with the vowel but measured separately.

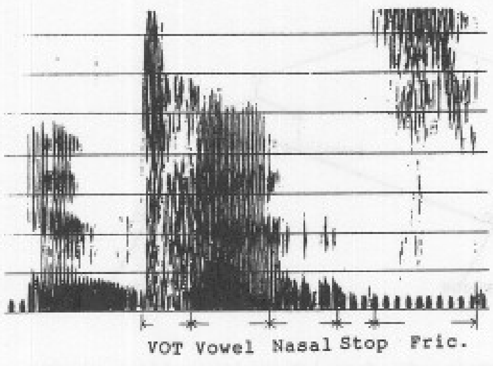


Figure 5. Segmentation of *intense* from spectrogram.

Results

As shown in Fig. 6, there is variation among speakers with respect to whether a silent gap appears: the two female speakers, MS and MC, always produced stops in both words but the other speaker ES did not. Sometimes there was no silent gap between /n/ and /s/ in *in tents* and the silent gap did not always appear in *intense*. This was true of citation forms as well.

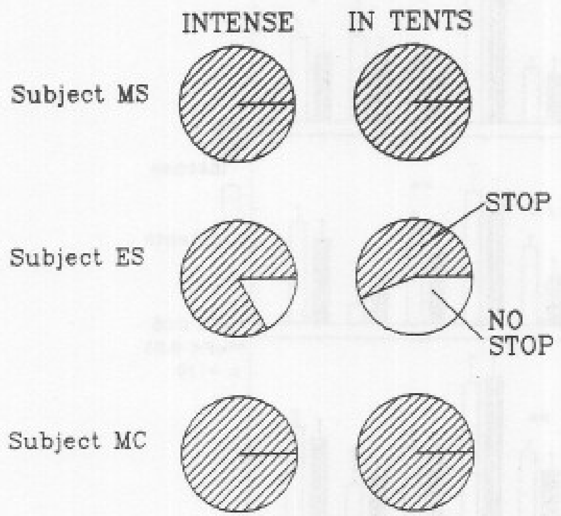


Figure 6. Percentage of occurrence of acoustic stops in the speech of each speaker in noncitation forms.

Looking at all of the segment durations for noncitation forms given in Fig. 7, speaker ES produced short epenthetic and short underlying stops compared to the other two speakers. Fig. 7 includes the token with no stop. However, even when these are excluded, speaker ES's stop duration is still much shorter.

An analysis of variance shows that in noncitation forms, the nasal, stop, and fricative of speaker MS in *in tents* were significantly longer than those in *intense*. The nasal of speaker ES in *in tents* was significantly longer than its counterpart in *intense*. (This result is counter to what would be expected if *tents* has underlying voiceless stop and *tense* does not, since the duration of vowel is negatively correlated with the number of consonants in the syllable coda.) The VOT of speaker ES and MC in in tents was significantly longer than that in *intense*. It may be because of the presence of a word boundary. That is, word initial /t/ in *in tents* may have longer aspiration as a signal of the word boundary.

Figure 8 compares citation and noncitation forms for subject MS. Citation and noncitation forms differed in two ways: first, in both words, all segments in citation forms are much longer than those in noncitation forms. (This was also true of the other speakers.) Second, in citation forms, none of the segmental duration in *intense* is significantly different from its counterpart in *in tents*. That is, although unlike the other speakers, speaker MS had a significant difference in stop duration in noncitation forms, this difference was reduced to the point of disappearing in citation forms, perhaps because of a different prosodic structure.

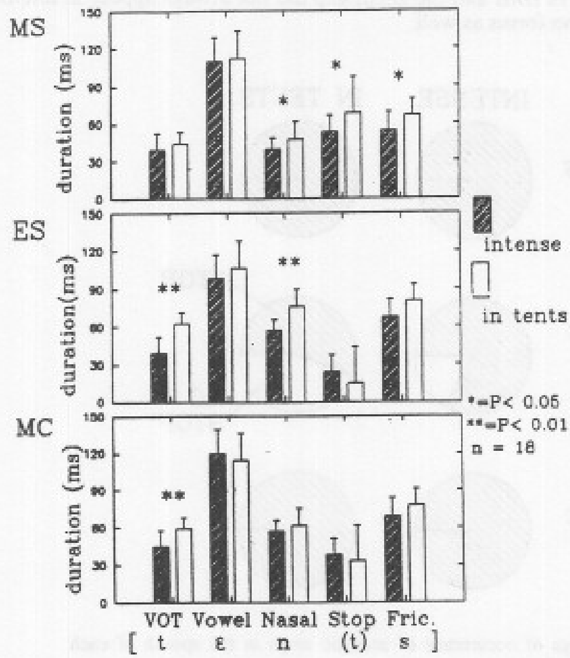


Figure 7. Duration measurements for each speaker in noncitation forms

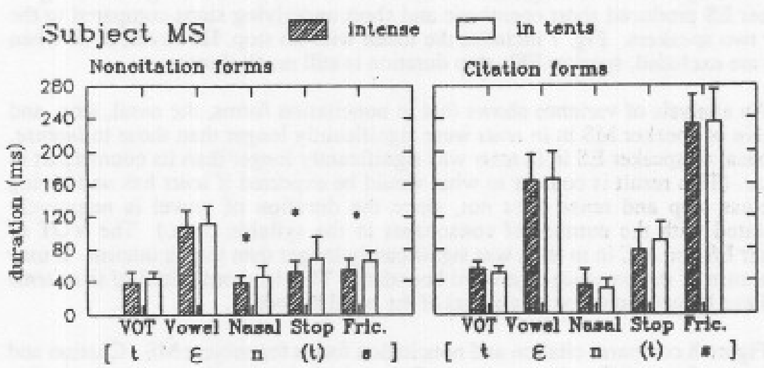


Figure 8.. Duration measurements for speaker MS.
*= P < 0.05, ** = P < 0.01, n = 18.

Discussion

The results of this experiment seem to suggest that *intense* and *in tents* might be the same phonetically for the speakers ES and MC, unless there is some factor other than duration which is used by listeners to distinguish them from one another. For speaker MS, on the other hand, there seems to be a small difference but only in noncitation forms.

Experiment 2: Perception Task

Method

In order to see whether listeners perceive a difference between *intense* and *in tents,* I ran a perceptual identification experiment. I only used citation forms in this experiment. I made a stimulus tape using the 18 tokens for each word type produced by each speaker for the first experiment. All the tokens for each speaker were repeated twice and two sets were constructed. In each set, all tokens were separated into two blocks and their order was randomized keeping the number of each word type constant within a block. Fifteen adult native speakers, three of whom were the subjects who produced the tokens, performed a forced answer identification test. For each speaker, the number of correct responses was averaged across all subjects.

Results

As shown in Fig. 9, listeners performed very poorly, essentially at chance level for all three speakers. Speakers even performed poorly for tokens which they produced themselves. Furthermore, even for the same token, listeners often responded differently to the first repetition and the second repetition when tokens were in different blocks.

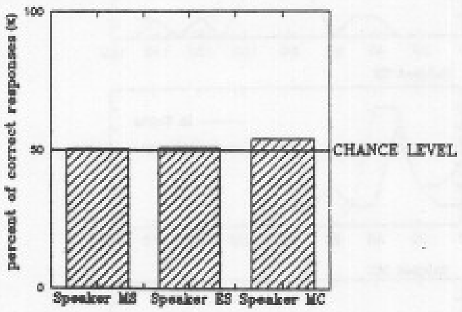


Figure 9. Percent of correct responses for identification task.

Discussion

The poor identification of tokens in this perception study demonstrates that these listeners could not tell *intense* and *in tents* apart and that they guessed when they were asked to label the tokens. These results and the acoustic duration measurement data together suggest that there are no crucial perceptual cues available for listeners to distinguish underlying and epenthetic stops from each other, and that the two types of stops in citation forms are not phonetically different. Even potential phonetic cues to the number of words constituting the tokens (one word for *intense* vs. two words for *in tents*) didn't help listeners distinguish the tokens from each other.

Experiment 3: Correlation of token judgments with stop durations

Method

Closer examination of the duration of stops in the citation tokens shows that the durations of epenthetic and underlying stops lie along two continua which overlap with one another, as shown in Fig. 10. Despite the overlap, speaker MS tended to have a longer stop duration in underlying stops, but this difference is not apparent in the other speakers.

It might be possible that listeners make use of the stop duration differences for speaker MS in the perception experiment even though they were not significant statistically. That is, generalizing from the small duration differences for this one speaker, listeners might tend to respond to shorter silent gaps as epenthetic and to longer gaps as underlying.

In order to test this possibility, I correlated perceptual responses with stop durations. If the stop duration difference is used as a perceptual cue, there should be a correlation between longer duration and more responses toward the underlying stop.

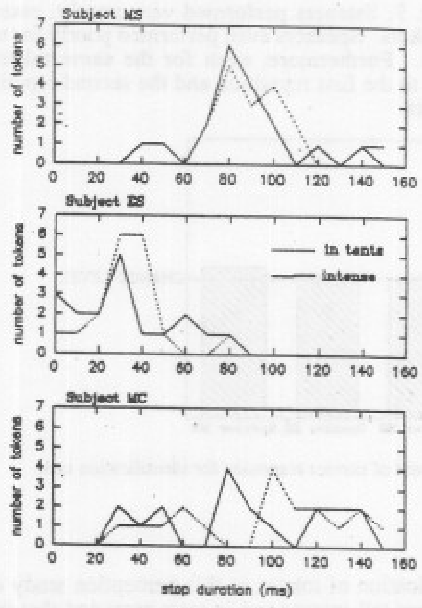


Figure 10. Range of duration of epenthetic and underlying /t/ produced by three speakers.

Results

As shown in Table 2, the correlation between the duration of stops in the tokens and subjects' responses to them is very poor except for one speaker, MS. The other two speakers, ES and MC, show even slightly negative correlations.

Table II. Correlation of token judgments in perception experiment with the duration of stops.

Speaker MS	0.55 ($P < 0.0005$)
Speaker ES	-0.095
Speaker MC	-0.13

Discussion

The poor correlation in the cases of speaker ES and speaker MC would be another indication that the two tokens are phonetically the same for these speakers. For speaker MS, on the other hand, even though the correlation was not very high, it still explains some of the perceptual variance, 25%. This result seems to suggest that there were some cases where duration differences were salient enough for listeners to use as a perceptual cue for speaker MS.

Conclusions

In this paper, I tried to determine the phonetic and phonological status of the so-called epenthetic stops in English. Acoustic duration measurements and data from the perception, and correlation studies together demonstrate that epenthetic and underlying stops are phonetically the same for some subjects. The duration of the stops were not different and listeners could not tell the two tokens apart. Fourakis and Port's study also showed that for some lexical items, the two types of stops were phonetically the same. They found the significant difference for only one minimal pair, *dense* and *dents* but not other pairs such as *tense* and *tents*. The subject who did show a difference in this study showed it more clearly in noncitation forms: therefore, the difference is also context sensitive. In future studies, I will try to figure out with more subjects and more lexical items, what about the context produces the difference .

* The original version of this paper was presented at the 121st meeting of the Acoustical Society of America, Baltimore, Maryland, 1991.

References

Ali, L. H. & R. Hammarberg. 1979. 'Intrusive stops in nasal-fricative clusters: An aerodynamic and acoustic investigation', *Phonetica* 3: 85-97.

Beckman, M. 1984. 'Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese', *Phonetica* 41: 61-71.

Browman, C. P. & L. Goldstein. 1987. 'Tiers in articulatory phonology, with some implications for casual speech', *Status Report on Speech Research* 92. Haskins Lab.

Browman, C. P. & L. Goldstein. 1989. 'Articulatory gestures as phonological units', *Phonology* 6: 201-251.

Clements, G. N. 1985. 'The geometry of phonological features', *Phonology*

Yearbook 2.

Clements, G. N. 1987. 'Phonological feature representation and the description of intrusive stops', in A. Bosch et al. (eds.) *Autosegmental and Metrical Phonology*. Parasession of the Annual Meeting of the Chicago Linguistics Society, Chicago, 1987.

Davis, S. 1989. 'The location of the feature [continuant] in Feature Geometry', *Lingua* 78: 1-22.

Fourakis, M & R. Port. 1986. 'Stop Epenthesis in English', *Journal of Phonetics* 14: 197-221.

Fowler, C.A. 1980. 'Coarticulation and theories of extrinsic timing', *Journal of Phonetics* 8: 113-133.

Hayes, B. 1986. 'Assimilation as spreading in Toba Batak,' *Linguistic Inquiry*, 17: 467-499.

Kim, H. S. 1990. *The Manner Features in Phonological Representations*, Ph.D. dissertation, University of Indiana.

Lee, S-H. 1990. 'Evidence from glottal articulation for the analysis of intrusive stops in English,' *Journal of the Acoustical Society of America* 87:Suppl.1, p.S121.

Nolan, F. 1990. 'The Descriptive Role of Segments: Evidence from Assimilation' in D. R. Ladd and G. Docherty ed., *Laboratory Phonology* 2 (papers from the second conference on laboratory phonology). Cambridge University Press.

O'Dell, M. & R. Port. 1963. 'Discrimination of word-final voicing in German', *Journal of the Acoustical Society of America* 73: suppl.1, p.S31.

Ohala, J. J. 1975. 'Phonetic explanation for nasal sound patterns', in Ferguson, Hyman, Ohala ed., *Nasalfest: papers from a symposium on nasals and nasalization*, 289-316, Language Universals Project, Stanford.

Ohala, J. J. 1981. 'Speech timing as a tool in phonology', *Phonetica*, 38: 204-212.

Ohman, S. E. 1966. 'Coarticulation in VCV utterances: Spectrographic measurements', *Journal of the Acoustical Society of America*, 39: 151-168.

Zwicky, A. M. 1972. 'Note on a phonological hierarchy in English,' in R.P Stockwell & R. K. S. Macauley, ed., *Linguistic Change and Generative theory*, 275-301.

Interference for 'new' versus 'similar' vowels in Korean speakers of English

Sun-Ah Jun and Islay Cowie

Department of Linguistics, The Ohio State University

Abstract

This paper tests Flege's (1987) Speech Learning Model and Bohn and Flege's (in press) hypothesis about the 'deflected' realization rule of a 'similar' L2 vowel. It is shown that Korean-English bilinguals' production of new English vowels, /I, U/, conforms to Flege's prediction. However, their production of similar English vowels, /i, u, U/, conformed to neither Flege's model nor to Bohn and Flege's hypothesis. We especially examined the interference between Korean and English high vowels, /i, I, U, u/, and /U/ based on 8 Korean-English bilinguals with different years of residnece in the States, 4 English monolinguals and 3 Korean monolinguals. Formant values of English vowels produced by Korean-English bilinguals with different years of residence in America were compared with those of English monolinguals. For the vowel /i/, Flege's notion of 'similar' L2 vowels needs be redefined to distinguish similar and identical vowels. He may need either some continuous measures or more systematic criteria to categorize whether a phone in L2 is new or similar to phones in L1.

Introduction

Languages differ in their ways of phonetic realization as well as their phonemic inventories. Weinreich (1953) states that learners of a second language (henceforth L2) tend to identify L2 phones in terms of their native language (henceforth L1) categories and, as a result, use articulatory patterns established during L1 acquisition to produce those L2 phones. Therefore, it is easily noticeable whether an utterance is produced by a native speaker or not.

In general, the degree to which a new phonetic category in L2 is established is proportional to the degree of experience in the L2 and the age at which L2 is acquired (Williams, 1980). Other factors also seem to play a role, such as the degree of similarity between L1 and L2. For example, Flege's (1987) Speech Learning Model showed that equivalence classification prevents adult L2 learners from establishing a phonetic category for 'similar' L2 phones, as opposed to 'new' phones. A 'new' phone is defined as an L2 phone which has no counterpart in L1 and therefore differs acoustically from all phones found in L1. A 'similar' L2 phone, on the other hand, is defined as an L2 phone which is realized in an acoustically different manner from an easily identifiable counterpart in L1. Flege (1987) found that the production of French /y/, a 'new' phone, by English speakers who were highly experienced in French was more authentic than that of French /u/, a 'similar' phone, which was close to their English /u/. That is, after long exposure and practice, adults could eventually produce an L2 vowel authentically if it were a 'new' vowel, but could not produce it authentically if it were a 'similar' vowel.

This speech model was refined somewhat when Bohn and Flege (in press) found that German-English bilinguals' production of a 'new' English vowel /æ/

affected the production of a 'similar' English vowel /ɛ/, as long as a category for the new vowel /æ/ had *not* been established. That is, the production of the new English vowel /æ/ by the less experienced German-English bilinguals was closer to the English norm than that by the more experienced bilinguals. It was claimed that the realization rules used to produce a 'similar' L2 vowel are "deflected" by the neighboring 'new' vowel, for which L2 learners had not yet established a phonetic category. Therefore, according to the refined model, if a 'similar' vowel and a 'new' vowel are neighboring vowels, the more experienced bilinguals would produce the 'new' L2 phone more authentically than the 'similar' phone due to the equivalence classification but the less experienced bilinguals would produce the 'similar' phone more authentically due to the 'deflected' realization rule until the neighboring 'new' phone is acquired.

This paper tests Flege's Speech Learning Model and the refined version of it, using Korean and English data. We especially examined the interference between Korean and English high vowels, /i, ɪ, ʊ, u/, and /ʌ/. Formant values of English vowels produced by Korean-English bilinguals with different years of residence in America were compared with those of English monolinguals.

A pilot study was conducted to decide what are 'new' and what are 'similar' vowels between Korean and English high vowels and /ʌ/. Formant values of /i/ and /u/ were overlapping very much between Korean and English, with English /u/ having somewhat higher F1 and F2 values than Korean /u/. On the other hand, English lax vowels /ɪ, ʊ/ showed separate vowel spaces from Korean high vowels /i, ɨ, u/, even though the edges of each vowel space were overlapped; In the F1 dimension, English /ɪ, ʊ/ have higher F1 values than those for Korean high vowels, and in the F2 dimension, English /ɪ/ is in between Korean /i/ and /ɨ/ while that of English /ʊ/ is in between Korean /ɨ/ and /u/. These results were also found in Jun (1990). Based on formant values, we categorized English /i/ and /u/ as 'similar' vowels and English /ɪ, ʊ/ as 'new' vowels to Korean bilinguals.

The pilot study also showed that Korean and English /ʌ/ have mean formant values which differ more than those of English /i/ and /ɪ/; Korean /ʌ/ tends to have lower F1 values and higher F2 values than English /ʌ/. We also found individual variations for the formant values of /ʌ/ among English monolinguals. One speaker showed formant values for a very high back vowel which are closer to those of Korean /ʌ/, while the other showed formant values for a more central vowel. However, since the monolinguals who participated in our present experiment produced values similar to the formant values for English /ʌ/ observed in Peterson and Barney (1952): F1: 640 Hz, F2: 1190 Hz for 33 male speaker, which have been assumed to be the standard values, we did not consider the possible individual deviation for /ʌ/ in our discussion. Even though /ʌ/ in both languages differs acoustically, Korean learners of English tend to recognize English /ʌ/ as Korean /ʌ/ since there is no other unrounded close-to-mid vowel in Korean. Therefore, we categorized English /ʌ/ as a 'similar' vowel together with English /i/ and /u/ but English /ɪ, ʊ/ as 'new' vowels to Koreans. Figure 1 shows a schematic vowel space for each language based on data from our pilot study and Jun (1990). The same tendency was also found in the present experiment.

In Flege (1987), an intermediate value for VOT norms in L1 and L2 was observed for experienced L2 speakers. He suggested that the phonetic category may be restructured or modified as experienced L2 speakers judge the L2 and L1 phone to be equivalent. Therefore, to examine the effect of L2 on L1, we compared formant values of Korean vowels produced by Korean-English bilinguals with those of Korean monolinguals. The vowel space for each bilingual was examined to see whether, and if so how, acquiring or recognizing a new vowel influences the production of other vowels.

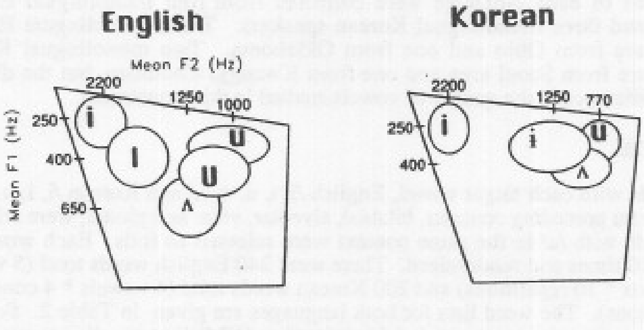


Figure 1. Schematic vowel spaces for English /i, I, u, U, ʌ/ and Korean /i, ɨ, u, ʌ/, based on data from Jun (1990).

The results support Flege's model to some extent in that more experienced Korean-English bilinguals produced 'new' English vowels more authentically than 'similar' English vowels. However, another aspect of the results contradicts the predictions of both Flege's (1987) original hypothesis and Bohn and Flege's (in press) refinement of it, namely that less experienced Korean-English bilinguals produce 'similar' English vowels better, i.e., closer to the English norm and distinct from Korean counterpart, than more experienced bilinguals do.

Method

Subjects

Eight Korean-English bilingual speakers with different durations of residence in the U.S.A. participated in the experiment. Four of them resided in the U.S. from 1.2 to 5.3 years and their average age was 29 (Less Advanced Group). The other four resided in the U.S. from 26 to 31 years and their average age was 56 (Advanced Group). The years of residence (YrR in a table) for each speaker is shown in Table 1.

Table 1. Years of Residence (YrR) for Advanced (subj.1-4) and Less Advanced(subj 5-8) group.

Levels	Advanced				Less Advanced			
Subjects	1	2	3	4	5	6	7	8
YrR	31	30	28	26	5.3	3.3	1.3	1.2

All of the subjects came to the U.S. in their late twenties to attend graduate school and have practiced spoken English since that time. All of the members of the less advanced group had been in Columbus since they came to this country and have been graduate students in various areas of humanities. Members of the advanced group had been in Columbus an average of 10 years prior to the experiment, and they all studied Chemistry or Chemical Engineering. Among those in the advanced group, only subject 2 is married to an American. Except for him, all the bilinguals speak English at work or school but Korean at home. All of them speak the same dialect of Korean, i.e., Seoul Korean. As norms for each language, productions of each language were collected from four monolingual English speakers and three monolingual Korean speakers. Three monolingual English speakers are from Ohio and one from Oklahoma. Two monolingual Korean speakers are from Seoul area and one from Kwangju Chonnam, but the dialects have no difference in the quality of vowels studied in this experiment.

Material

Words with each target vowel, English /i, ɪ, u, ʊ, ʌ/ and Korean /i, ɨ, u, ʌ/, in four different preceding contexts, bilabial, alveolar, velar and glottal, were selected. Four words with /a/ in the same context were selected as foils. Each word was repeated 10 times and randomized. There were 240 English words total (5 vowels * 4 contexts * 10 repetitions) and 200 Korean words total (4 vowels * 4 contexts * 10 repetitions). The word lists for both languages are given in Table 2. Some of Korean forms are nonsense words (abbreviated as 'NM') but are all pronounceable syllables.

Table 2

Korean test data

pi 'rain'	ti NM	ki 'a flag'	hi NM
pɨ NM	tɨ NM	kɨ 'that'	hɨ NM
pu 'richness'	tu 'two'	ku 'nine'	hu 'after'
pəl 'punishment'	tət 'a trap'	kəp 'a scare'	hət 'vain'

Korean foil data

pal 'a foot'	tal 'a moon'	kat 'a hat'	hap 'a sum'

English test data

bead	tea	key	heed
bid	tip	kid	hid
book,	took	could	hood
booed	two	coo	who'd
but	dug	cut	hut

English foil data

pot	dot	cot	hot

Recording

Recording was made in a pseudo-anachoic chamber for all English monolinguals and subjects 1 and 7. For others, a portable cassette recorder (SONY Walkman, model WM D6C) was used to record their readings in a quiet room of a house or an office. Each monolingual speaker read the appropriate word list and bilingual speakers read both word lists. For the bilinguals, the order of language was counterbalanced between subjects: half of each group read English first, and the other half Korean first. To prime the bilinguals' productions of each language, the first author explained the procedures in the relevant language just before recording the corpus for that language. In addition to this, each subject was asked to read an English passage (a paragraph from either Lawrence's novel, "Rocking

Horse Winner", or Fry's (195?) article) aloud just before reading the English corpus. One month later, recordings were made again for an accent judgment test. The bilinguals read five sentences excerpted from *The Lantern* (the Ohio State University student newspaper) and a tape is made by editing the recordings; for each sentence, the order of readings for each speaker was randomized and rerecorded with a two second pause between each speaker and a four second pause between each sentence. The five sentences are given in the Appendix.

Accent Judgements

The degree of accent in English of the bilinguals was rated by 15 native English speakers who listened to the tape and rated them, based on their general impression, on a scale of 1 to 10, where 1 was 'native-like' (or 'no foreign accent') and 10 was 'very strong foreign accent'. The score of the degree of accent was correlated with the years of experience.

Formant Measurements

Each of the test words was digitized and a vowel with some portions of surrounding consonants was extracted using a waveform editing program. The formant values of the vowel were measured around a center of the vowel showing a steady formant value using an LPC formant tracker with 14 coefficients, 200 window size and 100 window step.

Results

The F1 and F2 values of similar and new vowels in both languages were compared in their absolute values and in relation to other vowels within a vowel space for each bilingual. Then, a correlation between the accent judgment score and years of residence for each bilingual is shown.

Comparison for similar vowels

The results show that the formant values of /i/ were more than similar between English and Korean. This was also found in Jun (1990)[1] and in our pilot study. Formant values for /i, u, ʌ/ in both languages by monolinguals and Korean bilinguals are shown in Figure 2.[2] In figure 2, the leftmost column within each graph represents mean formant values for each monolingual group and two horizontal lines in the middle of each graph indicate the value of the leftmost column, i.e. the means of each vowel in each language.

As shown in Fig. 1(a), since the formant values of /i/ in both languages are almost identical, we cannot test Flege's (1987) Speech Learning Model which is about 'new' vs. 'similar' vowel. Mean formant values are represented with one standard error in both directions. The results for another similar vowel, /u/, showed inconsistency with the hypothesis that the production of similar vowels is roughly the same for all bilinguals regardless of L2 experience. This is shown in Fig.1 (b). That is, less experienced bilingual subjects produced /u/ with more distinction in the two languages than the more experienced subjects did, i.e, they

---

[1] Jun(1990) found that the labial positions for /i/ were not significantly different in English and Korean.

[2] Since we did not normalize speaker variation, a relative, rather than the absolute, formant value was compared between two languages in relation to that of monolinguals.

produced English /u/ lower and fronter (higher F1 and F2) than Korean /u/, thus closer to the English norm.

Lastly, monolinguals' production of the supposedly similar vowel, /ʌ/, shows that formant values are quite different for each language. However, as seen in Fig. 2 (c), almost all bilinguals produced this vowel with similar formant values for each language. Moreover, most of the formant values fall between norms for each language. This may indicate bidirectional interference.

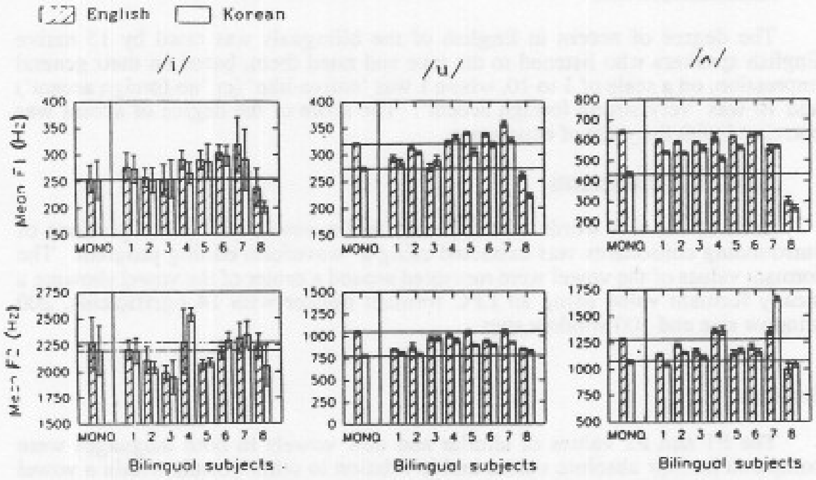


Figure 2. Mean formant values of English and Korean (a) /i/, (b) /u/, and (c) /ʌ/ by English and Korean monolinguals (MONO) and eight Korean bilinguals. SD is shown around the mean.

<u>Comparison for new vowels</u>

Since the important aspect of /ɪ/ and /ʊ/ is less the absolute formant value than the relationship to the tense counterpart, a relative formant values of /i/ versus /ɪ/ or /u/ versus /ʊ/ for each subject were compared with the monolinguals' norms. That is, to see whether bilinguals acquired the 'new' vowels, /ɪ, ʊ/, or not, the relative formant values between tense /i, u/ and lax /ɪ, ʊ/ counterparts are shown in Figure 3. Shaded portion and the white portion of the first bar graph (MONO) represent formant values of English monolinguals' tense and lax vowels, while the other bar graphs represent values of tense/lax vowels by Korean-English bilinguals.

Figure 3 (a) and (b) show that the differences in F1 and F2 values between tense and lax vowels by subjects with the most experience are very close to those by the English monolinguals, while these differences are very small for the subjects with less experience. However, some less experienced subjects, subject 7 and 6 (1.3 and 3.4 years of residence, respectively) showed a clear F2 distinction between /i/ and /ɪ/ only (Figure 3 (a)). At the same time, subject 4 (26 years of residence) does not seem to show any sign of acquiring the 'new' vowels with small difference in formant values between tense and lax vowels. This seems to indicate that years of residence does not correlate well with the degree of acquisition of a new vowel. This discrepancy will be discussed more in the next section where we correlate the years of residence with the degree of accent in general.

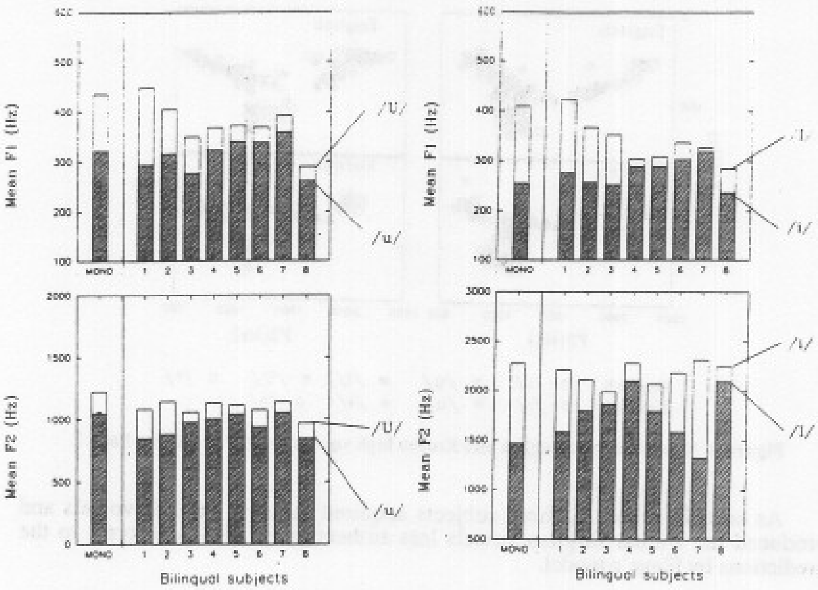


Figure 3. Mean F1 and F2 values of (a) English /i/ and /I/ (b) English /u/ and /U/ by monolinguals of English and Korean and by eight Korean-English bilinguals

Except for these 3 subjects, the general pattern seems to support Flege's Speech Learning Model. That is, L2 learners with more experience will produce a new L2 phone more authentically, while those with less experience will substitute the L1 phone which is acoustically and/or articulatorily close to the new L2 phone for the new L2 phone.

Since it looks as though the less experienced bilinguals produce 'similar' vowels better (meaning closer to the English norm and distinguishing from Korean counterpart) than the more experienced bilinguals but 'new' vowels worse than the more experienced bilinguals, we examined the vowel space for each bilingual to see whether, and if so how, acquiring or recognizing a new vowel influences the production of other neighboring vowels.

Vowel Space for each subject

When we look at each bilingual's vowel spaces for each language, three types of vowel spaces were observed. The first type shows a separate space for each vowel within each language, especially between English /i/ and /ɪ/ or /u/ and /ʊ/, showing that the new L2 vowels are acquired. It also has overlapping spaces for similar vowels between languages. This type of vowel space is shown by subject 1 and 2 and possibly by subject 3. We will refer to this group as the 'separate-vowel-space' group. Figure 4 shows the vowel spaces for subject 1 and 2.

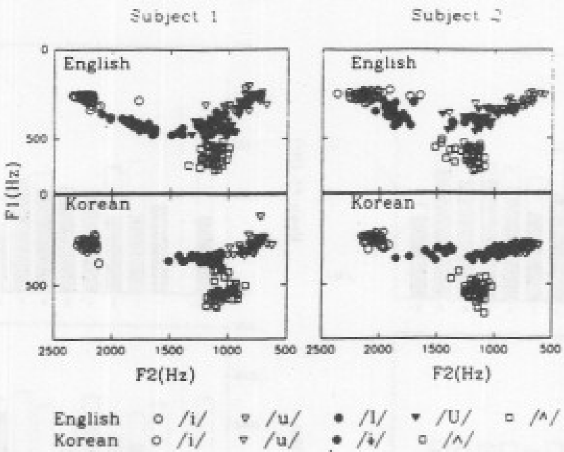


Figure 4. Vowel space of English and Korean high vowels and /ʌ/ by subject 1 and 2

As seen in Figure 4, these subjects acquired the new English vowels and produced the similar English vowels less authentically. This conforms to the predictions by Flege's model.

The second type is the opposite of the first type. The vowel space for English tense (i.e. 'similar') and lax (i.e. 'new') vowels are completely overlapping and the spaces covered by both vowels are the same as the space of Korean /i/ or /u/. This type of vowel space tells us that Korean /i/ and /u/ are substituted for English /i, ɪ/ and /u, ʊ/, respectively. This type is what we can expect from the bilinguals with little experience in L2. We will refer to this group as the 'overlapping-vowel-space' group. Two subjects, subject 5 and 6 showed this type of vowel spaces even though they have been in the States for at least four years. Figure 5 shows their vowel spaces.

The third type shows a stage inbetween the first and the second type. As in the second type, it has overlapping spaces for English tense and lax vowels covering spaces for both vowels. However, unlike the second type, the space for English /i/ is lower and more back (i.e., more central) than the norm and thus close to the space for /ɪ/ of the English norm. Also, the space for English /u/ is lower and fronter (i.e., more central) than the norm and thus close to the space for /ʊ/ of the English norm. This type of vowel space is shown by the rest of bilinguals with some variations. These bilinguals may be in the process of acquiring new English vowels, and we will refer to this group as the 'intermediate' group. The vowel spaces of subjects 4 and 7 represent this type of vowel space and are shown in Figure 6. The vowel space also indicates that there is interference from English to Korean; Korean /i/ or /u/ has more variation in F2 values than the Korean norms.

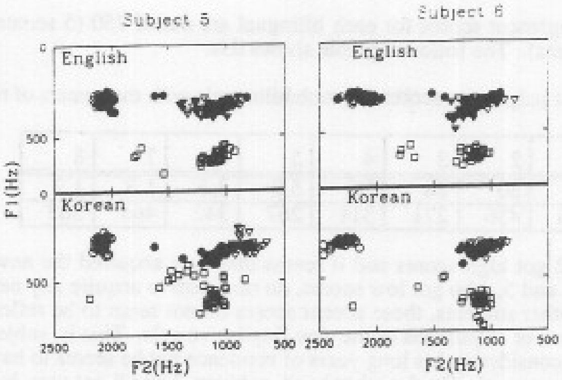


Figure 5. Vowel space of English and Korean high vowels and /ʌ/ by subject 5 and 6

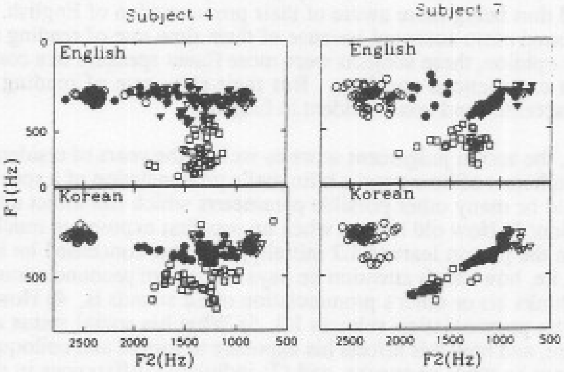


Figure 6. Vowel space of English and Korean high vowels and /ʌ/ by subject 4 and 7

Since some of the bilinguals with fewer years of residence show signs of acquiring a new vowel while some of the bilinguals with longer years of residence show no signs of acquiring a new vowel, it seems that the years of residence alone is not a good criterion for dividing bilinguals into two groups. Accent judgement scores given to each bilingual were compared in terms of the degree of acquisition of the new vowel. The following section shows the results of the correlation between years of residence and accent judgement scores.

<u>The correlation between years of residence and accent scores</u>

Accent scores for each subject were correlated with the years of residence. The resulting correlation was very low, $r = 0.1323$. This seems to be due to the low score (i.e. more accented) for experienced speakers 3 and 4 and the high score (less accented) for less experienced speakers 7 and 8.

Accent judgement scores for each bilingual are out of 750 (5 sentences * 15 judges * 10 points). The following table shows this.

Table 3. Accent judgement scores for each bilinguals with their years of residence.

Subject	1	2	3	4	5	6	7	8
YrR	31	30	28	26	5.3	3.3	1.3	1.2
Scores	468	456	271	314	267	342	465	362

Subject 1 and 2 got high scores and it seems that they acquired the new vowels, while subject 4 and 5, who got low scores, do not seem to acquire any new vowel. However, for other subjects, these accent scores do not seem to be reflected very well in the degree of acquisition of the new English vowels. That is, subject 3 got a very low score considering his long years of residence but he seems to have almost acquired the new vowel. On the other hand, subjects 7 and 8 got very high scores (2nd and 4th out of 8 bilinguals) with less than 2 years of residence. However, neither of them have completely acquired the new English vowels as has been shown in Figure 3. Maybe they have a better intonation pattern, timing or better pronunciation of consonants. This may possibly be due to their majoring in linguistics, and thus being more aware of their pronunciation of English. Subjects 3 and 4 may sound more accented because of their slow rate of reading and voice quality. In our opinion, these subjects were more fluent speakers in a conversation than most less experienced speakers. But their slow rate of reading passages sounded more accented and less confident in English.

Therefore, the accent judgement score as well as the years of residence cannot be used as an indicater of how good a bilingual's pronunciation of a specific sound is. There could be many other possible parameters which can affect a bilingual's L2 pronunciation: 1) How old he was when he was first exposed to much English, 2) From whom the person learned L2 initially, 3) How concerned he is with his pronunciation, i.e. how much attention he pays to his own pronunciation, and how important he thinks his or other's pronunciation of L2 sounds is, 4) How much he knows about the pronunciation rules in L2, 5) What his social status and social environment are, and how this affects his exposure to formal and colloquial L2, 6) How fast or slow he reads or speaks, and (7) individual differences in mimicking skills, etc.

Discussion

Productions of new English vowels, /ɪ/ and /ʊ/, by Korean bilinguals were consistent with Flege's hypothesis: the 'separate-vowel-space' group produced them authentically while less experienced ones did not. However, productions of similar English vowels, /i/, /u/, and /ʌ/, showed characteristics which were not predictable from Flege's model. First, productions of a similar English vowel, /u/, were not consistent with the hypothesis: the 'intermediate' group produced /u/ more authentically than the 'separate-vowel-space' group. The 'intermediate' group bilinguals showed a category merge effect between the similar, /u/, and the new vowel, /ʊ/, thus, making the similar vowel /u/ much closer to the English norm than to the Korean /u/, while making no clear distinction between English /u/ and /ʊ/. This may happen because they recognize the new L2 vowel but have not quite acquired it yet.

On the other hand, the 'separate-vowel-space' group showed formant values of English /u/ very close to those of their Korean /u/. This may be due to their acquisition of the new English vowel, /ʊ/. That is, they may try to avoid invading the vowel space for the newly acquired L2 phone.

A similar tendency was shown for the other 'similar' vowel /i/. Even though formant values of /i/ in both languages are almost identical, the the 'intermediate' group bilinguals produced English /i/ closer to English /ɪ/, thus making a distinction from Korean /i/. On the other hand, the 'separate-vowel-space' group produced English /i/ close to Korean /i/, thus making no distinction from Korean /i/ but clear distinction from English /ɪ/.

A similar reversed phonomenon is also shown in Bohn and Flege (in press) where inexperienced Germans produced /ɛ/ with durations that were English-like, whereas the experienced Germans produced /ɛ/ with shorter durations than both the native English and the inexperienced German group. They claimed that this phenomenon was attributed to the influence of new neighboring L2 phones on similar L2 phones only for the less experienced bilinguals, and that the realization rules used to produce a similar L2 vowel are "deflected" by the neighboring new vowel, for which L2 learners had not yet established a phonetic category.

This deflection by the neighboring new vowels for less experienced bilinguals doesn't work for our data since only the 'intermediate' group, not the 'overlapping-vowel-space' group showed this so-called 'deflection' phenomenon. This reversed phenomena can be explained better if we consider the vowel dispersion theory proposed by Liljiencrants and Lindblom (1973), Lindblom (1975, 1979) and Maddieson (1977). By the dispersion theory, each vowel in a language is maximally dispersed. So, if there are more vowels in the same height or backness dimension, there will be less variation for each vowel and the vowels will be evenly dispersed in the same dimension.

The same principle would work for the vowel space of an individual. That is, if someone acquired a new L2 vowel (as in the case for the 'separate-vowel-space' group), he or she might try not to invade the new vowel's space when producing a vowel adjacent to the new L2 vowel. However, if there is no established space for a new L2 vowel, nor is there any recognition for the acoustic difference of the new vowel (as in the case for the 'overlapping-vowel-space' group), the new vowel and a vowel adjacent to the new vowel can both be produced in the same space which is originally for the vowel adjacent to the new vowel. And finally, if one recognizes the acoustic difference of the new vowel and at least tries to produce the new vowel (as in the case for the 'intermediate' group), one might overshoot the target (i.e. neighboring 'similar') vowel, thus, close to the new vowel, and creating a merge effect.

Therefore, the fact that the 'intermediate' but not the 'overlapping-vowel-space' group produced 'similar' phones better than the 'separate-vowel-space' group might suggest that bilinguals are reorganizing their vowel space and the degree of reorganization depends on the degree of bilingualism. Bohn and Flege's explanation that the realization rule for the similar vowel is 'deflected' due to the new neighboring vowel as far as the new L2 vowel is acquired does not work for the results of our experiment since the 'overlapping-vowel-space' group did not show any better performance as far as the new L2 vowel is concerned. Rather, they substituted their L1 for the new L2 phone as in the case of the 'separate-vowel-space' group.

And finally, bilinguals did not produce the 'similar' vowel /ʌ/ close to the norm of any language. Rather, they produced /ʌ/ in both languages with formant values intermediate to the two languages; although, the norms of each language differed significantly. The less experienced bilinguals did not show any deflection phenomena, as predicted by Bohn and Flege (in press). We assume that this may be because there is no vowel space established for a neighboring 'new' vowel. This seems to indicate bidirectional interference.

Conclusion

This paper tested Flege's (1987) Speech Learning Model and Bohn and Flege's (in press) hypothesis about the 'deflected' realization rule of a 'similar' L2 vowel. It is shown that Korean-English bilinguals' production of new English vowels, /ɪ, ʊ/, conforms to Flege's prediction. However, their production of similar English vowels, /i, u, ʌ/, conformed to neither Flege's model nor to Bohn and Flege's hypothesis.

For the vowel /i/, Flege's notion of 'similar' L2 vowels needs be redefined to distinguish similar and identical vowels. He may need either some continuous measures or more systematic criteria to categorize whether a phone in L2 is new or similar to phones in L1.

For the vowels /u/ and /ʌ/, contrary to Flege's model, English /u/ was produced closer to the English norm by the 'intermediate' group bilinguals than by the 'separate-vowel-space' group. This reversed phenomenon was claimed by Bohn and Flege to be due to the 'deflected' realization rule of a 'similar' vowel next to a 'new' vowel until the acquisition of the new L2 vowel. But their hypothesis cannot explain the fact that the 'overlapping-vowel-space' Korean-English bilinguals did not produce English /u/ better than other bilingual groups. Rather, this could be explained better if we consider the combined vowel space for the two languages for each individual and apply the dispersion theory. Consideration of the vowel space can explain both why the reversed phenomenon was found also for English /i/, an almost identical vowel, and why the phenomenon was not found for English /ʌ/. The vowel space observed for English /ʌ/ by bilinguals also indicates that there is bidirectional interference.

Finally, we found that the years of residence in L2 is very poorly correlated with the accent judgement score of each bilingual and that, for adult learners, the accent judgement score as well as the length of residence in L2 was not a good indicator by itself of how good the production of a specific L2 phone is.

Appendix

1. If you are currently enrolled and do not receive instructions in the mail, by May 16, contact your college office immediately.
2. A Franklin County grand jury found that grades on state bar exams were fixed for nine people in 1986.
3. Your letter to Aunt Ruth may be scattered physically all over the disk.
4. A number of these students took jobs just to pay for the big event.
5. At present we have an extremely limited number of positions for this summer because of the mandated budget cuts.

Acknowledgement

The authors appreciate all the subjects who participated in the experiment; C. Chung, I. Choi, C.H. Kim, H.Y. Kim, C.L. Lee, T.Y. Moon, H.G. Shin and S.M. Yoo as bilingual subjects, T.K. Cho, J.H. Han and B.C. Kim as Korean monolingual subjects, and T. Bassam, K. deJong, R. Stout and C.Merhar as English monolingual subjects. We also thank those native speakers of English who judged the degree of accent in English.

* The original version of this paper was presented at the 121'st meeting of the Acoustical Society of America at Baltimore, Maryland.

References

Bohn O.-S. and Flege, J.E. in press. "The Production of New and Similar Vowels by Adult German Learners of English," paper submitted for publication, *Journal of 2ng Language Acquisition.*

Caramazza, A., Yeni-Komshian, G.H., Zurif, E.B. and Carbone, E. 1973. "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals," *Journal of Acoustical Society of America*, 54: 421-428.

Elman, J., Diehl, R. and Buchwald, S. 1977. "Perceptual switching in bilinguals," *Journal of Acoustical Society of America*, 62: 971-974.

Flege, J.E. 1987. "The Production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification," *Journal of Phonetics*, 15: 47-65.

Flege, J.E. 1988. "Factors affecting degree of perceived foreign accent in English sentences," *Journal of Acoustical Society of America*, 84(1): 70-79.

Flege, J.E. and Davidian, R.D. 1984. "Transfer and developmental processes in adult foreign language speech production," *Applied Psycholinguistics*, 5: 323-347.

Flege, J.E. and Eefting, W. 1987. "Cross-language switching in stop consonant perception and production by Dutch speakers of English," *Speech Communication* 6: 185-202.

Jun, S.-A. 1990. "Labial position and acoustics of Korean and English high vowels," a paper presented at the 119th meeting of the Acoustical Society of America, Penn State Univ. Apeared in this volume.

Liljencrants, J. and Lindblom, B. 1972. "Numerical simulation of vowel quality systems: the role of perceptual contrast," *Language* 48: No.4.

Mack, M. 1982. "Voicing-dependent vowel duration in English and French: Monolingual and bilingual production," *Journal of Acoustical Society of America*, 7(1): 173-178.

Maddieson, I. 1977. "Tone loans: a question concerning tone spacing and a method of answering it," *UCLA Working Papers in Phonetics* 36: 49-83. UCLA Phonetics Lab., Los Angeles.

Peterson, G.E. and Barney, H.L. 1952. "Control Methods Used in a Study of the Vowels," *Journal of Acoustical Society of America*, 24(2): 175-184.

Weinreich, U. 1953. *Languages in Contact, Findings and Problems*. The Hague: Mouton.

Williams, L. 1980. "Phonetic Variation as a function of second-language learning," *Child Phonology - Perception* Vol. 2. eds. by Yeni-Komshian, J. Kavanagh & C. Ferguson, 185-216, New York: Academic Press.

Zimmermann, G.N., Price, P.J. and Ayusawa, T. 1984. "The production of English /r/ and /l/ by two Japanese speakers differing in experience with English," *Journal of Phonetics*, 12: 187-193.