The Ohio State University

Working Papers in Linguistics # 36

Papers from the Linguistics Laboratory 1985–1987

Edited by
Mary Beckman
and
Gina Lee

The Ohio State University
Department of Linguistics

204 Cunz Hall
1841 Millikin Road
Columbus, Ohio 43210-1229
U. S. A.

September 1987

Information about Ohio State University WORKING PAPERS IN LINGUISTICS (OSU WPL)

Working Papers in Linguistics is an occasional publication of the Department of Linguistics of Ohio State University and usually contains articles written by students and faculty in the department. There are one to three issues per year. Information about available issues appears below. Nos. 1, 5, 10, & 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. To order Working Papers, please send a check made payable to "Ohio State University" to: OSU WPL, Department of Linguistics, Ohio State University, 1841 Millikin Rd., Columbus, OH 43210-1229.

OSU WPL #18, $2.00. 183 pp. (June 1975): Papers by Michael Geis, Sheila Geoghegan, Jeanette Gundel, G. K. Pullum, and Arnold Zwicky.

OSU WPL #19, $2.00. 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 Ohio State University on April 11, 1975.

OSU WPL #20, $2.00. 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.

OSU WPL #21, $2.50. 252 pp. (May 1976), Edited by Arnold Zwicky: **Papers on Nonphonology.** Papers by Steven Boer and William Lycan, Marian Johnson, Robert Kantor, Patricia Lee, and Jay Pollack.

OSU WPL #22, $3.50. 151 pp. (February 1977), Edited by Olga Garnica: **Papers in Psycholinguistics and Sociolinguistics.** Papers by Sara Garnes, Olga Garnica, Mary Louise Edwards, Roy Major, and John Perkins.

OSU WPL #24, $3.50. 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.

OSU WPL #25, $3.50. 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.

OSU WPL #26, $4.50. 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.

ISSN 0473-9604

OSU WPL #27, $4.50. 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.

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

OSU WPL #29, $5.00. 207 pp. (May 1984), Edited by Arnold Zwicky & Rex Wallace: **Papers on Morphology**. Papers by Belinda Brodie, Donald Churma, Erhard Hinrichs, Brian Joseph, Joel Nevis, Anne Stewart, Rex Wallace, and Arnold Zwicky.

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

OSU WPL #31, $6.00. 194 pages. (July 1985), Edited by Michael Geis. **Studies in Generalized Phrase Structure Grammar**. Papers by Belinda Brodie, Annette Bissantz, Erhard Hinrichs, Michael Geis and Arnold Zwicky.

OSU WPL #32, $6.00. 162 pages. (July 1986), **Interfaces**. 14 articles by Arnold Zwicky concerning the interfaces between various components of grammar.

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

OSU WPL #34, $6.00. 164 pages. (December 1986), Edited by Brian Joseph. **Studies on Language Change**. Papers by Riitta Blum, Mary Clark, Richard Janda, Keith Johnson, Christopher Kupec, Brian Joseph, Gina Lee, Ann Miller, Joel Nevis, and Debra Stollenwerk.

OSU WPL #35, $10.00. 214 pages. (May 1987), Edited by Brian Joseph and Arnold Zwicky. **A Festschrift for Ilse Lehiste.** Papers by colleagues of Ilse Lehiste at Ohio State University.

OSU WPL #36, $10.00. 140 pages. (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 Riitta Välimaa-Blum.

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 Serv. (ED) Ctr. for Applied Linguistics, 1611 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, 225 pp. ED 077 268.
Parts of #14. April 1973, 126 pp. ED
#15. April 1973, 221 pp. ED 082 566.
Parts of #16, December 1973, 119 pp. ED

OSU WPL # 36

Introduction

Papers from the Linguistics Laboratory 1985-1987

Edited by Mary Beckman and Gina Lee

This volume reports on some of the research completed in the Ohio State University Linguistics Laboratory during the period between late 1985 and the early part of 1987. The Linguistics Laboratory is part of the Linguistics Department at the Ohio State Unviersity and receives regular funding from the College of Humanities and fron the university. The directorship of the lab is a committee consisting of those Linguistics faculty who are primary users and one student. Support for individual research projects is augmented by outside grants to faculty members and by small grants to students from the Department of Linguistics. There is also provision for linguists in other departments within Ohio State University to apply to use the laboratory for their research, and for scholars from other institutions to apply to visit our department and conduct research in the laboratory. The papers in this volume are reports of research conducted by four students in the Linguistics department (Keith Johnson, Gina Lee, Riitta Välimaa-Blum, and Ann Miller), and by two scholars from outside of the department (Dr. Christiane Laeufer, who teaches French Linguistics in the Department of Romance Languages and Literatures here at Ohio State University, and Dr. Shiro Kori, who visited us for an extended period in 1986 while on sabbatical leave from the Osaka University of Foreign Studies). This collection of papers inaugurates a regular subseries of the Ohio State University Working Papers in Lingustics, the primary purpose of which will be to provide a periodic report of ongoing work in the Linguistics Laboratory.

A second purpose of this subseries within the OSU WPL is to attempt to partially fill the gap left by the retirement of Ilse Lehiste in June 1987. While none of us alone can hope to fill the place she made for herself in the field of experimental linguistics, perhaps by pooling our resources, we can maintain a bit of the reputation that she established for Ohio State University as a place where experimental work is done and where experimental approaches are encouraged in the pursuit of linguistic knowledge. With this purpose in mind, we dedicate this volume:

to our esteemed colleague and beloved teacher

Dr. Ilse Lehiste

OSU WPL # 36

Introduction

Papers from the Linguistics Laboratory 1985-1987

Edited by Mary Beckman and Gina Lee

This volume reports on some of the research conducted in the Ohio State University Linguistics Laboratory during the period between late 1985 and the early part of 1987. The Linguistics Laboratory is part of the Linguistics Department at the Ohio State University and receives regular funding from the College of Humanities and from the university. The directorship of the lab is a committee composed of three Linguistics faculty who are primary users and the Assistant. Support for individual research projects is supported by outside grants to faculty members and by small grants to graduate students from the Department of Linguistics. There is also provision for linguists in other departments within Ohio State University to apply to use the laboratory for their research, and for scholars from other institutions to apply to visit and conduct research in the laboratory. The papers in this volume are reports of research conducted by four students in the Linguistics department ([illegible]), and by two scholars from outside of the department: Dr. Christiane Laeufer, who teaches French linguistics in the Department of Romance Languages and Literatures here at Ohio State University, and Dr. Mieko Han, who visited during an extended period in 1986 while on sabbatical leave from the Osaka University of Foreign Studies. This collection of papers inaugurates a regular subseries of the Ohio State University Working Papers in Linguistics, the primary purpose of which will be to provide a periodic report of ongoing work in the Linguistics Laboratory.

A second purpose of this subseries within the OSU WPL is an attempt to partially fill the gap left by the retirement of Ilse Lehiste in June 1987. While none of us alone can hope to fill the place she made for herself in the field of experimental linguistics, perhaps by uniting our endeavors, we can maintain a bit of the reputation that she established for Ohio State University as a place where experimental work is done and where rigorous empirical approaches are encouraged in the pursuit of linguistic knowledge. With this purpose in mind, we dedicate this volume:

To our esteemed colleague and beloved teacher

Dr. Ilse Lehiste

OSU WPL # 36

Table of Contents

Papers from the Linguistics Laboratory 1985-1987

Edited by Mary Beckman and Gina Lee

OSU #36: Papers from the Linguistics Laboratory 1985-1987, Sept. 1987.

Intonation in Cantonese

Keith Johnson
Linguistics, Ohio State University

Abstract: The experiment in this paper explores the nature of intonation in a language which has lexical tone. In a pilot study it was found that a method of accounting for tone preservation (the identifiability of lexical tones in sentence contexts) which included a declining tone space was better suited to the task than one which assumed a level tone space. The main experiment attempted to separate and observe the contributions to this general downtrend made by boundary effects, tonal interaction and declination. There appears to be evidence for one type of boundary effect (initial raising) and declination. The data of this experiment were not consistent with the presence of the other type of boundary effect (final lowering) or tonal interaction factors. Two important variables were manipulated in this experiment. First, the length of a test sentence was manipulated on the assumption that longer sentences would show a greater decline of F0 if there was a declination effect. Second, the discourse position of test sentences was varied (from discourse medial to discourse final) as a test for the effect of discourse final lowering.

1. Introduction

Observation of tone and intonation patterns in many languages has often revealed a tendency for fundamental frequency to decline over the course of an utterance (Ladd 1983, Cohen, Collier and 't Hart 1982). In this paper, I will call this general phenomenon by the theory-neutral term 'downtrend'. Pierrehumbert and her colleagues, in their research on intonation patterns in English and Japanese, have identified several potential contributing factors to downtrend (Pierrehumbert 1980, Liberman and Pierrehumbert 1984, Beckman and Pierrehumbert 1986).

One class of factors, boundary effects, has to do with the intonational marking of prosodic units (prosodic phrases or possibly larger units of prosodic organization). Just as prosodic boundaries can be marked by a durational process like final lengthening in many languages, they are also signaled in English by intonational processes. The two types of intonational processes most often mentioned are final lowering and initial raising.[1] These terms refer to the tendency for F0 to fall at the ends of prosodic units and for units to start with a relatively high F0. The combined result of these intonational boundary effects is an overall drop in F0 within prosodic units.

---

[1] In non-tone languages boundary tones are often used in the description of intonation (Pierrehumbert 1980 for example). This does not seem to be a possibility in Cantonese because tonal distinctions are preserved before boundaries, as elsewhere.

Tonal interaction is another source of F0 patterns which exhibit downtrend. The proto-typical case in which the interaction between tones results in downtrend is known as downstep. This phenomenon, often found in African languages, is probably best characterized as a lowering of pitch range following the occurrence of a low tone. Thus, the second high tone in a sequence HLH is lower than the first high tone in the sequence. Downstep and other kinds of tonal interaction are called tone sandhi when they occur in East Asian languages, although it should be noted that not all types of tone sandhi produce downtrend.

One other type of factor which may result in downtrend is declination. Ladd (1983) defines declination as 'a gradually changing backdrop to local F0 events' (p. 54). Unlike boundary effects and tonal interaction, declination is an effect which operates over the entire duration of a phonological unit (exactly what constitutes the domain for declination is not yet clear).

In this paper the presence of a declination factor (as a backdrop to other F0 events) in speech production is tested. Robert Ladd (1983) suggests two methods for testing a declination-as-frame-of-reference hypothesis. The first of these two suggestions derives a test from the fact that in a declination-as-frame-of-reference model, declination is a function of time. Thus, 'the difference in F0 between two accent peaks should correlate with the length of the interval between them' (p. 67). Second, Ladd suggests the 'detailed phonetic study' of tone languages as a method of testing for declination. The advantage of studying a tone language is that 'the phonological identity or non-identity of points in contours could be much more rigorously controlled, and mean values could be obtained from phonologically comparable cases' (p. 69).

The language under study in this paper is Cantonese. It has been described as a language 'rich in the number of tones' (six in the standard descriptions) and 'rather poor in tone sandhi' (Hashimoto 1972, p. 112). The six lexical tones found in Cantonese are: high level (55), high rising (35), high mid (33), mid rising (23), low mid (22) and low level (11). (These are the same tone numbers as were assigned by Wong, 1982, with one minor change - 11 for 21.) This description of the inventory of tones is based on F0 analysis made of isolated pronunciations of a minimal set with the segments [fʌn] as produced by one speaker. The F0 patterns of the six different lexical tones of Cantonese are shown in figure 1. A description in terms of tone numbers is also indicated on this figure. The figure shows that the tones are more different from each other toward the end of the syllable than they are at the beginning. For this reason in all F0 measurements in sentence context in this paper I measured the F0 of a tone at the end of the syllable upon which it occurs.

In figure 2 the data of figure 1 are presented as percentages of the F0 envelope defined by the highest and lowest tones. In this figure the highest tone was defined as 100% of the possible range for tones and the lowest one was given the value 0%. The other tones were computed as percentages of this tone space.

The lack of tone sandhi in Cantonese seems to imply that tonal interaction phenomena can be eliminated from consideration prima facie. Casual observation indicates that even without tonal interaction downtrend

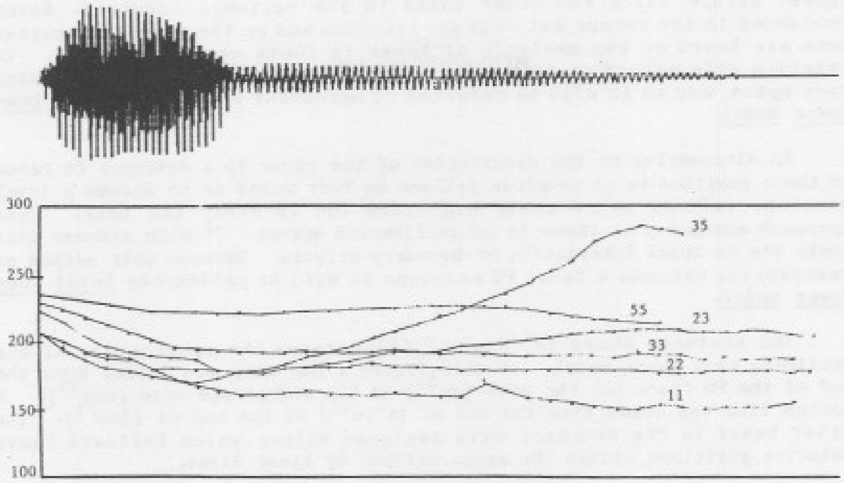


Figure 1: F0 patterns of the six different lexical tones of Cantonese.

is present in Cantonese. For instance, in sentences in which all of the words have high tones the later occurring high tones have lower F0 than earlier ones. It is also the case that the identities of tones are preserved in sentence context. Hashimoto (1972) describes this tone preservation: 'Except for unstressed syllables, almost every syllable is pronounced with the same tone in isolation that it bears in sequence' (p. 112).

In a pilot study I examined the possibility of accounting for tone preservation by the inclusion of a declination component in the description of Cantonese intonation. The study was an investigation of the F0 patterns of a portion of a corpus of sentences which were originally recorded during a field methods class at Ohio State. As such, the corpus was not designed to meet any particular qualifications but rather was a random collection of sentences. The sentences included in this study all exhibited the following pattern: there was a high-level or high-rising tone early in the sentence and one near the end of the sentence. Also, there was a low tone early in the sentence and one late in the sentence. These four words in each

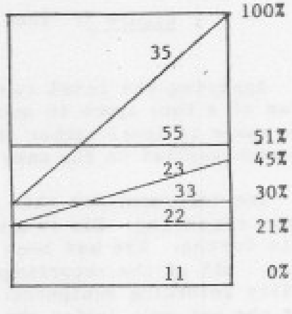
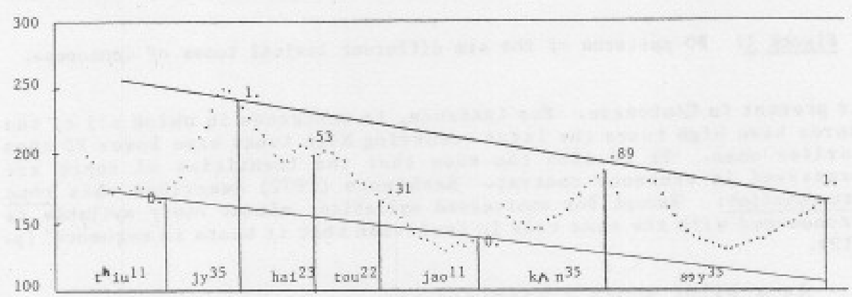


Figure 2: Tones in isolation as percent of the tone space.

sentence served to define an F0 envelope (which I will also call a tone space) within which the other tones in the sentence occurred. Seven sentences in the corpus met this description and so the results reported here are based on the analysis of tones in these seven sentences. In practice this method of defining an F0 envelope resulted in a declining tone space, and so it will be referred to hereafter as the <u>declining tone space model</u>.

An alternative to the description of the tones in a sentence in terms of their position in an envelope defined by four words is to assume a level envelope defined by an early high tone and an early low tone. This approach assumes that there is no declination effect. It also assumes that there are no tonal interaction or boundary effects. Because this method of description assumes a level F0 envelope it will be called the <u>level tone space model</u>.

The sentence shown in figure 3 illustrates the application of the declining tone space model. In this figure a top line was fitted from the end of the F0 trace for the word [jy^{35}] to the end of the word [sɔy^{35}]. A bottom line was drawn from the end of [t^{h}iu^{11}] to the end of [jao^{11}]. The other tones in the sentence were assigned values which indicate their relative positions within the space defined by these lines.



<u>Figure 3</u>: Tone values in a declining tone space.

Applying the level tone space model we could analyze the tones in terms of a tone space in much the same way - the difference being that the tone space is level rather than declining. Figure 4 is an example of this approach applied to the same sentence.

The informant for this study was a student at Ohio State at the time of the recording. She is a native of Hong Kong and her parents both grew up in Canton. She had been living in the United States for about three years. All of the recordings were made in an anechoic chamber using high quality recording equipment. The informant read the sentences from notes that she had made during the course of the field work class sessions. The fundamental frequency analysis was performed using the Sift algorithm in the ILS software package operating on a DEC PDP 11/23 at the Linguistics Lab at Ohio State.

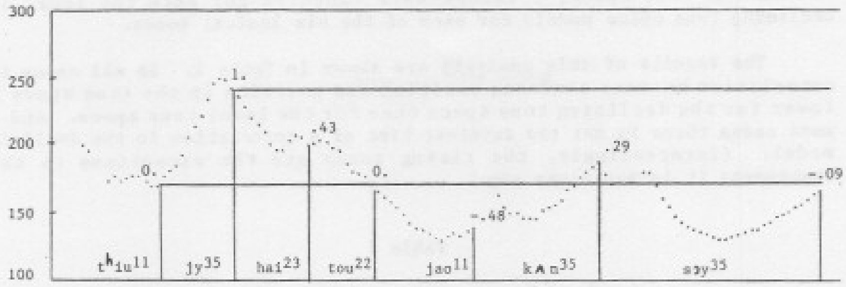


Figure 4: Tone values in a level tone space.

The results illustrated in figure 5 indicate that both the declining tone space and the level tone space models succeed in keeping the lexical tones separate from each other when averages are considered. Thus, for example, when all occurrences of the level tones 33 and 22 are considered their relative average positions within the tone space are distinct. There is, however, a difference between the two models which is hidden by this presentation of the data. The difference is that in the level tone space model there was a good deal more variation in the position of a tone within the tone space. This is illustrated in figure 4 where the first high rising tone [jy^{35}] is at the top of the range, the next high rising tone [$k\Lambda n^{35}$] is 29% up in the range and the last one [$s\text{ɔ}y^{35}$] is at the bottom of the range.

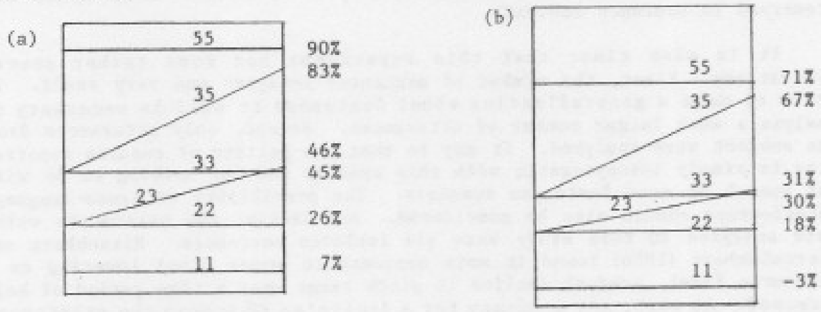


Figure 5: Tone values (a) as percent of declining tone space and (b) as percent of level tone space.

This figure also presents the possibility that the extra variation found in the level tone space model was not random but rather was correlated with the position of the word within a sentence - so that high tones early in a sentence would be higher in the range than high tones later in the sentence. To test for such a correlation between position in a sentence (as indicated by counting words from the beginning) and position within the tone space, r values were computed for both the level and declining tone space models for each of the six lexical tones.

The results of this analysis are shown in Table I. In all cases the correlation between sentence position and position in the tone space is lower for the declining tone space than for the level tone space. And in most cases there is not the faintest hint of a correlation in the declining model. (Interestingly, the rising tones are the exceptions to this statement; it is not clear why.)

Table I

r values for tones by sentence position.

Tone number:	55	35	33	11	23	22
Tone space model:						
Declining	.062	.323	.008	.009	.239	.095
Level	.615	.793	.377	.492	.820	.725

The results of this pilot study indicate that the phenomenon which I described earlier as tone preservation can be modeled (for these sentences) much more straightforwardly and simply by the incorporation of a declining tone space in the description of Cantonese tones in sentences. The level tone space model suffered from the limitation that the value of a tone within the tone space was dependent upon its location in the sentence. This runs counter to the native speaker's intuition that tones are preserved in sentence contexts.

It is also clear that this experiment has some rather severe limitations. First, the number of sentences analyzed was very small. In order to make a generalization about Cantonese it will be necessary to analyze a much larger number of utterances. Second, only utterances from one subject were analyzed. It may be that the pattern of results reported here is simply idiosyncratic with this speaker and has nothing to do with the speech of most Cantonese speakers. The possibility of cross-language interference should also be considered. And lastly, the utterances which were analyzed in this study were all isolated sentences. Hirschberg and Pierrehumbert (1986) found it more accurate to model final lowering as a discourse final, gradual decline in pitch range over a time period of half a second. As such, any tendency for a declining F0 across the utterances in this study may have been due to final lowering rather than declination. And so the experiment does not really constitute a test of the declination hypothesis. Almost all of these considerations were dealt with in the main experiment. (The possibility of cross-language interference could not be adequately addressed because of the nature of the subject pool available in Columbus.)

2. Main Experiment

The limitations of the pilot study were taken into consideration and an experiment was designed which avoided most of the problems which limited the interpretability of the pilot. The main difference between the two experiments was that a corpus of sentences was designed for this experiment with which the declination hypothesis could be tested.

The materials employed in this experiment incorporated three variables which tested various factors which may contribute to downtrend. The test sentences contained two test words which were segmentally similar and had the same lexical tone. The factors which were tested were length of test sentence, medial or final discourse position, and the number of sentence initial syllables (i.e. the number of syllables before the first test word).

The predictions of a model which includes declination versus a model in which there is no declination component are illustrated in Table II. In this table, an x in a given cell means that a factor could or should have a significant effect, and 0 means that it should not be significant.

Table II

Predictions for experiment 2 from two kinds of models

model	with declination	without declination
factor:		
long vs. short	x	0
discourse position	0	x (final lowering)
number of sentence-initial syllables	0	x (initial raising)

If speakers employ some sort of declination component the length factor should prove to be significant. It is, of course, possible that the length factor may be confounded with some other aspect of the test sentences in this experiment, and indeed in at least one test sentence this appears to have happened (it will be discussed later). When other factors which contribute to the F0 of a sentence are eliminated, the 'no declination' model predicts that length will not be a significant factor in F0 analysis.

If there is a final lowering factor in Cantonese we would predict that the discourse position of the test sentence will be a significant factor. This involves the assumption that final lowering is an indication of the end of a discourse and not a feature of every sentence within a discourse (see Hirschberg and Pierrehumbert 1986).

The presence or absence of initial raising should be indicated by an interaction of the sentence initial syllables variable with the length variable and by an overall difference in rate of decline depending on the number of sentence initial syllables. The interaction is expected because one of the test paragraphs confounds the length condition with the initial syllables variable.

2.1 Methods

The materials were designed to test for the effects of sentence length and discourse position on F0. Sentences were constructed in sets of three which constituted a short discourse, and in which the second and third sentences were interchangeable, thus allowing for a test of discourse position (whether the test sentence was medial or final). Each test sentence contained two target words (nouns with high level tone), one early in the sentence and another later. In one version of each test sentence the interval between the target words was short (from 2 to 5 syllables); in a second version, the interval between target words was long (from 5 to 8 syllables). In this way the effect of sentence length was included in the experiment. (The three sentence paragraphs which were used in this experiment are listed in the Appendix. There were six of these; they will be referred to as paragraph 1, paragraph 2, and so on.) The structure of each three sentence paragraph was thus:

1. First sentence establishes the discourse topic.
2a. First version of the test sentence has a short interval between the target words.
2b. Second version of the test sentence has a long interval between the target words.
3. A filler sentence which can be interchanged with the second sentence.

These paragraphs were constructed by a native speaker of Cantonese and fellow linguist (Zheng Sheng Zhang) using target words which were matched in vowel quality and contained sonorants. Subjects read four versions of each paragraph (2 intervals (long vs short) X 2 discourse positions (medial vs final)). The corpus was thus composed of 24 paragraphs. They were written in traditional characters on separate sheets of paper.

Two speakers read the corpus a total of five times each. Each time the order of the items was re-randomized. Recordings were made in an anechoic chamber using high quality recording equipment. The fundamental frequency analysis was performed on a DEC PDP 11/23 using the Sift algorithm (or a modified cepstral processing technique) in the ILS signal processing software package. The two F0 measurements for each test sentence (measured at the end of each test word as in the first experiment) were converted to a ratio - TW1/TW2. Thus, for each of the 120 X 2(subjects) productions recorded there was a ratio indicating the relative difference between an early and late high tone under the various conditions being tested.

2.2 Results

In a repeated measures ANOVA performed on the data from this experiment the interval between test words proved to be a significant factor ($F(1,1)=59.39$, $p<.1$). The main effect for paragraph number was marginally significant ($F(5,5)=2.23$, $p=.2002$), and the interaction of the interval and paragraph factors was marginally significant ($F(5,5)=2.85$, $p=.1378$). No other main effects or interactions approached significance.

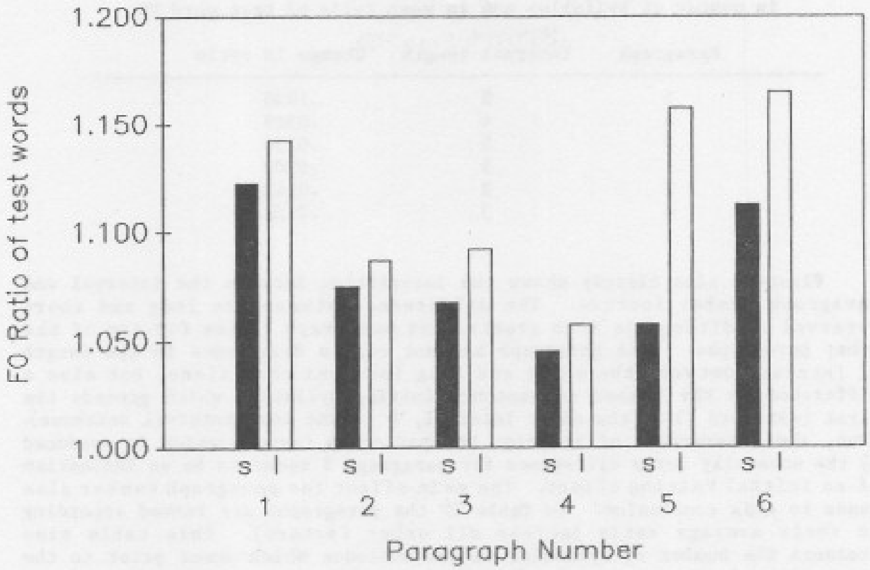


Figure 6: Mean F0 ratio of test words, divided by interval length and paragraph. Clear bars are for long intervals, shaded bars for short.

Figure 6 presents the results of this experiment. The vertical axis is the mean ratios of the test words, averaged over the two speakers. In all cases the ratio is greater than 1.00. This indicates that under all combinations of treatments the F0 of the first test word was on average higher than the F0 of the second test word. In this graph the length of a bar corresponds to the amount of F0 decline over the interval between test words. The twelve bars drawn in this graph represent the long and short interval conditions for each of the six paragraphs. The interval main effect which was reported above can be seen in this graph by comparing the short vs long condition for each paragraph. In all cases the short interval condition has a smaller amount of decline than does the long interval condition. As is clear in the paragraphs listed in the appendix the difference in terms of number of syllables between the long and short interval conditions is not the same for each paragraph.

Table III lists the differences between long and short interval conditions in terms of (1) number of syllables and (2) change in average ratio. There is a rough correspondence among these measurements such that greater change in length corresponds to greater change in ratio. This is consistent with the hypothesis that a declination component is involved in the production of intonation in Cantonese.

Table III

Differences between short and long interval
in number of syllables and in mean ratio of test word F0

Paragraph	Interval length	Change in ratio
5	5	.1038
6	4	.0519
3	3	.0236
1	5	.0203
2	2	.0143
4	3	.0126

Figure 6 also clearly shows the interaction between the interval and paragraph number factors. The difference between the long and short interval conditions is much greater for paragraph 5 than for any of the other paragraphs. This paragraph has not only a difference in the length of interval between the short and long interval conditions, but also a difference in the number of sentence initial syllables which precede the first test word (3 in the short interval, 0 in the long interval sentence). Thus, the interaction of interval and paragraph factors which is produced by the unusually large difference for paragraph 5 seems to be an indication of an initial raising effect. The main effect for paragraph number also leads to this conclusion. In Table IV the paragraphs are ranked according to their average ratio (across all other factors). This table also presents the number of syllables in the sentence which occur prior to the occurrence of the test word. The correlation between overall decline and position of the first test word relative to the beginning of the sentence is striking evidence for initial raising.

Table IV
Relationship between number of initial
syllables and F0 ratio

Paragraph	Ratio	Initial syllables
6	1.1383	0
1	1.1316	1
5	1.1076	1.5*
2	1.0808	1
3	1.0806	2
4	1.0501	6

*3 in short version and 0 in long version thus, 1.5 average.

3. Discussion

This experiment has identified some of the contributing factors to the downtrend found in the pilot study. Of the <u>boundary effects</u>, <u>initial raising</u> seemed to play a part in the realization of F0 in these utterances, while final lowering did not seem to be present. The present experiment offers no evidence in favor of positing any tonal interaction effects (if low level assimilation is not included in this category). Because of the reports to the effect that tone sandhi is limited in Cantonese, I chose not to include a test for tonal interaction factors. The possibility of tonal interaction should, however, be tested in an experiment which manipulates as an independent variable the identity of the tones which occur between two test words. Until such an experiment is conducted the question is an open one.

Finally, this experiment does provide evidence for a declination effect in Cantonese. It was the explicit goal of the experiment to test the declination hypothesis in an experiment which took Ladd's (1983) suggestions into account. The results indicate clearly the presence of a declination component in the production of F0 in Cantonese.

The declination component has a couple of properties which make it worthy of further study. First, unlike boundary effects and tonal interaction, declination is a global factor. As such it presents different challenges for explicit descriptive systems: for instance, whether declination should be modeled as a linear or logarithmic function; whether the rate of decline should be modeled as constant or varying; and how much preplanning should be included in a model of declination (there is a tendency for both steeply falling and slowly falling utterances to end at the same F0). It should be noted that the relative difficulty involved in modeling declination or the extra power that such a global effect adds to the formal power of the descriptive system is not a valid argument to the effect that declination is not really a factor in the production of speech. It is rather only an argument to the effect that declination is relatively difficult to model or that a powerful formalism is required to describe this aspect of speech. The presence or absence of declination is an empirical issue which must be decided empirically.

The second interesting property of declination is that it has a physiological motivation. Unlike boundary effects and tonal interaction the declination effect seems to have a physiological cause. Lieberman's (1967) proposal that virtually all F0 downtrend in English could be attributed to declining subglottal pressure has been discredited (see Ohala 1978). However, there is a tendency for a small decline in subglottal pressure over the course of utterances (Lieberman 1967, Ohala 1978). The correlation between subglottal pressure and F0 demonstrated by van den Berg (1958) leads to the conclusion that if subglottal pressure does decline during an utterance then the potential range for F0 will also decline. If declination is an automatic consequence of speaking while the other intonational effects discussed here are not, then there is the interesting possibility that the nonautomatic effects are derived from the more natural (physiologically motivated) one. This could be an explanation for the fact that final lowering and initial raising are common while final raising and initial lowering are not - that downstep is common while upstep isn't.

Acknowledgements

The results of the pilot experiment were presented at the LSA 1985 Winter meeting at Seattle. Preliminary results of the main experiment were presented at the ASA 1986 Spring meeting at Cleveland. An earlier version of this paper was also presented at the Conference of Chinese Linguistics at Ohio State (March, 1986). I appreciate the helpful comments from several people who attended these meetings: in particular, Dana Bourgerie, Osamu Fujimura, Tim Light and John Ohala. I started on this path under the direction of Cathy Callaghan and Ilse Lehiste. Mary Beckman has lately taken on the thankless task of trying to get me to work. Thanks for your patience. (There, now it's not thankless.) Special thanks to Zheng Sheng Zhang for his indispensible help with the materials for the second experiment. Thanks also for musical assistance to Vince Andrews.

References

Beckman, M. and J. Pierrehumbert. 1986. Intonational structure in Japanese and English. Phonology Yearbook. 3:255-309.

Cohen, A., R. Collier. and J. 't Hart. 1982. Declination: construct or intrinsic feature of speech pitch? Phonetica. 39:254-273.

Hashimoto, Oi-Kon Yue. 1972. Phonology of Cantonese. London: Cambridge University Press.

Hirschberg, J. and J. Pierrehumbert. 1986. The intonational structuring of discourse. From Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics. 10-13 June 1986, pp. 136-144.

Ladd, D. Robert. 1983. Declination: A review and some hypotheses. Phonology Yearbook. 1:53-74.

Liberman, Mark and Janet Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In Mark Aronoff and Richard Oehrle (eds.) Language Sound Structure. Cambridge: MIT Press.

Lieberman, Philip. 1967. Intonation, Perception and Language. Cambridge: MIT Press.

Ohala, John. 1978. Production of tone. In V. Fromkin (ed.) Tone: A Linguistic Survey. New York: Academic Press. pp. 5-39.

Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. PhD. dissertation, MIT.

van den Berg, Janwillem. 1958. Myoelastic-aereodynamic theory of voice production. Journal of Speech and Hearing Research. 1:227-244.

Wong, Maurice Kuen-Shing. 1982. Tone Change in Cantonese. PhD. dissertation, University of Illinois, Urbana-Champaign.

Appendix: Sentence Sets

I. 1. sɔ·^{35} jɑʊ̯^{23} kɛ·^{33} tɕɔ·ɪ̯^{21} li·ʊ̯^{35} toʊ̯^{55} hɔ·^{35} ji·^{23} tɕoʊ̯^{11} tɕɔ·ŋ^{53}
all material in all cases can do pack

søy^{35} kə jɛ·^{23}
water thing

2. (a) pɔ·^{55} $\boxed{\text{leɪ̯}^{55}}$ tɕøn^{55} hɑɪ̯^{11} pɔ·^{55} $\boxed{\text{leɪ̯}^{55}}$ tɕoʊ̯^{11} kɛ·^{33}
bottle be glass made

(b) pɔ·^{55} $\boxed{\text{leɪ̯}^{55}}$ tɕøn^{55} tʰoŋ^{21} mɑ·ɪ̯^{21} pɔ·^{55} leɪ̯^{55} pu·ɪ̯^{55}
bottle and drinking glass

hɑɪ̯^{11} pɔ·^{55} $\boxed{\text{leɪ̯}^{55}}$ tɕoʊ̯^{11} kɛ·^{33}
be glass made

3. søy^{35} tʰoŋ^{33} hɑɪ̯^{11} kəm^{53} sok^{11} wɑk^{11} tɕɛ·^{35} mɑk^{11} tʰɑʊ̯^{21} tɕoʊ̯^{11} kɛ·^{33}
water bucket be metal or wood end made

II. 1. hɑ^{11} tʰi·n^{55} tʰɕoŋ^{21} tɕi·^{35} tɔ·^{55} kwɔ·^{33} toŋ^{55} tʰi·n^{55}
summer insect suff. much more winter

2. (a) ji·ʊ̯^{33} $\boxed{\text{mən}^{55}}$ hœ·ŋ^{55} jɑʊ̯^{11} jiʊ̯^{33} $\boxed{\text{mən}^{55}}$ tɕœ·ŋ^{33} si·n^{53}
must have mosquito incense and must have mosquito net only

tɕi·^{33} fʌn^{21} tək^{33} tɕœ·k^{11} kɑ·ʊ̯^{33}
sleep can

(b) ji·ʊ̯^{33} $\boxed{\text{mən}^{55}}$ hœ·ŋ^{55} mən^{55} jɑʊ̯^{21} tʰoŋ^{21} mɑ·ɪ̯^{21} $\boxed{\text{mən}^{55}}$
must have mosquito incense mosquito oil and mosquito

tɕœ·ŋ^{33} si·n^{53} tɕi·^{33} fʌn^{21} tək^{33} tɕœ·k^{11} kɑʊ̯^{33}
net only sleep can

3. tøy^{35} mən^{55} tɕən^{53} hɑɪ̯^{11} moʊ̯^{23} mat^{55} jɛ·^{23} hoʊ̯^{33} pɑ·n^{11} fɑ·t^{33}
as for mosquito really there is nothing good way of managing

III. 1. ŋɔ·^{23} tɕɛ·k^{33} mɑ·ʊ̯^{55} hoʊ̯^{33} pɑ·k^{33} ji·m^{33}
meas. cat very naughty

2. (a) nɛ·^{55} tɕɛ·k^{33} mɑ·ʊ̯^{55} tɕɑɪ̯^{35} hoʊ̯^{33} tɕoŋ^{53} ji·^{33} tɕøy^{53}
part. meas. kitten much likes to chase

koŋ^{55} kɑɪ̯^{55} t^{h}oŋ^{21} mɑ·ɪ̯^{21} wɑ·n^{35} t^{h}ɛ·k^{33} p^{h}eɪ̯^{21} k^{h}ɑʊ̯^{21}
rooster and play kick leather ball

(b) nɛ·^{55} tɕɛ·k^{33} mɑ·ʊ̯^{55} tɕɑɪ̯^{35} hoʊ̯^{33} tɕoŋ^{53} ji·^{33} li·ʊ̯^{23}
part. meas. kitten much likes to tease

kɑʊ̯^{35} tɕɑɪ̯^{35} tɕøy^{53} koŋ^{55} kɑɪ̯^{55} t^{h}oŋ^{21} mɑ·ɪ̯^{21} wɑ·n^{35}
puppy chase rooster and play

t^{h}ɛ·k^{33} p^{h}eɪ̯^{21} k^{h}ɑʊ̯^{21}
kick leather ball

3. k^{h}øy^{23} fu·n^{53} heɪ̯^{35} wɑ·n^{35} sɪk^{33} jok^{55} kɛ^{33} jɛ·^{35}
it likes play moving part. thing

IV. 1. ŋɔ·^{23} sɑɪ^{33} loʊ^{35} hoʊ̯^{33} pɑ·k^{33} ji·m^{33}
my younger brother very naughty

2. (a) jɑʊ̯^{21} jət^{55} jət^{11} k^{h}øy^{11} kɛ·^{33} sɑʊ̯^{35} tɕi·^{35} meɪ̯^{55} jəp^{11}
is one day he poss. hand finger end put

tɕɔ·^{35} lɔ·k^{11} høy^{33} pɔ·^{55} leɪ̯^{55} tɕøn^{55} toʊ̯^{33}
in down go bottle

(b) jɑʊ̯^{21} jət^{55} jət^{11} k^{h}øy^{11} kɛ·^{33} sɑʊ̯^{35} tɕi·^{35} meɪ̯^{55} ham^{11}
is one day he poss. hand finger end whole,

pan^{11} lɑ·m^{23} jəp^{11} tɕɔ·^{35} lɔ·k^{11} høy^{33} pɔ·^{55} leɪ̯^{55} tɕøn^{55} toʊ̯^{33}
entire put in down go bottle

IV. 3. k^{h}øy^{11} moʊ̯^{23} ɕi·^{23} tsɘw^{21} wɑ·n^{35} ŋok^{55} k^{h}eɪ̯^{35} kɛ·^{33} jɛ·^{23}
he have not matter finish play home poss. thing

V. 1. sɛk^{11} ji·n^{55} tɕən^{53} hɑɪ̯^{11} hoʊ̯^{33} wu^{53} tɕoʊ̯^{53} kɛ^{33}
smoking certainly is very dirty

2. (a) m̩^{21} sɑɪ̯^{35} kɔ·ŋ^{35} $\boxed{\text{ji·n}^{55}}$ tɕɑɪ̯^{35} pin^{33} $\boxed{\text{ji·n}^{55}}$ fu·ɪ̯^{53} lɑ^{55}
not send say cigarette turn into ashes

(b) $\boxed{\text{ji·n}^{55}}$ tɕɑɪ̯^{35} t^{h}oŋ^{21} mɑ·ɪ̯^{21} ji·n^{55} tɕi·^{35} pi·n^{33}
cigarette and cigarette paper turn into

sɑɪ̯^{53} $\boxed{\text{ji·n}^{55}}$ fu·ɪ̯^{53} kɛ·^{33}
completely ashes

3. toʊ̯^{33} t^{h}cy·^{33} toʊ̯^{55} hɑɪ̯^{11} ji·n^{55} t^{h}ɑʊ̯^{35}
everywhere all be cigarette butt

VI. 1. sɪk^{11} ji·n^{55} tɑʊ̯^{35} hoʊ̯^{33} kwɔ·^{33} sɪk^{11} ji·n^{55} tɕɑɪ̯^{35}
smoking pipe better than smoking cigarette

2. (a) $\boxed{\text{ji·n}^{55}}$ tɑʊ̯^{35} hɑɪ̯^{11} moʊ̯^{23} $\boxed{\text{ji·n}^{55}}$ t^{h}ɑʊ̯^{35} kɛ·^{33}
smoking pipe be have not cigarette butt

(b) $\boxed{\text{ji·n}^{55}}$ tɑʊ̯^{35} m̩^{21} t^{h}ɕi·^{23} ji·n^{55} tɕɑɪ̯^{35} kɑm^{33} jɑʊ̯^{23} $\boxed{\text{ji·n}^{55}}$
smoking pipe not looks like cigarette so have cigarette

t^{h}ɑʊ̯^{35} kɛ·^{33}
butt

3. si·k^{33} ji·n^{55} tɑʊ̯^{35} kwɔŋ^{53} tʒɪn^{53} kwɔ·^{33} sɪk^{11} ji·n^{55} tɕɑɪ̯^{35}
smoking pipe clean more than smoking cigarette

OSU #36: Papers from the Linguistics Laboratory 1985-1987, Sept. 1987.

A Study of Toishan F0

Gina Lee
Linguistics, Ohio State University

Abstract: Like other Chinese languages, Toishan uses tone to signal differences in meaning among words. With the exception of certain morphologically conditioned tone changes, which must be memorized by speakers, Toishan exhibits no tonal modifications; in particular, there appear to be no tonal modifications which are strictly phonologically governed. Given the absence of such changes, Toishan provides an ideal situation for examining the hypothesis that declination--a phonologically unmotivated lowering of F0, independent of tone--is a relevant component in a model of intonation. Three native speakers read paragraphs containing five pragmatically connected sentences of different length. Two target tones within each sentence were measured. No evidence for declination was found. The F0 pattern is described in terms of an overall structure involving pitch range expansion for initial sentences and compression (or lowering) for final sentences. The results are discussed in light of evidence from other languages and implications for a model of local F0 implementation.

1. Introduction

1.1. Downtrend and declination

In examining the properties of intonation, many have noted a general tendency for F0 to gradually lower during the production of a syntactic unit. Such lowering of the intonation contour is believed by some to be universal. This phenomenon is known by a variety of terms; following Liberman and Pierrehumbert (1984), I will refer to the general lowering of F0 as downtrend.

In some cases, downtrend may be the result of a phonologically conditioned rule. For instance, phonologically identical high tones within an utterance are sometimes not phonetically identical. In many African tone languages, each non-initial high tone in the sequence HLHLHLH is produced with lower fundamental frequency relative to the preceding high tone, due to a phonetic rule which lowers high tones after a preceding low tone. Such lowering is a special case of downtrend known as downdrift.

Recent work has proposed that in some situations F0 lowering may be due to declination: a gradually declining phonetic frame of reference that is independent of tonal context (e.g. Pierrehumbert 1980). Various physiological explanations have been proposed. Among them is the argument that subglottal pressure decreases toward the end of a syntactic unit, and that--everything else being equal--this decrease corresponds to a decrease in F0. Although there is some doubt as to whether subglottal pressure actually decreases during the production of an utterance, and although other physiological factors may also contribute to F0 lowering, it is generally agreed that a decrease in subglottal pressure provides a clear physiological motivation for F0 lowering.

Unlike downdrift, declination is not dependent on specific phonological characteristics of the utterance but, rather, is dependent on time. Figures 1 (a) and (b) illustrate the relationship between utterance length and a declining F0 backdrop. Points x and y represent tones in an F0 contour. Distance d represents the difference between the heights of x and y. In 1 (b), the time interval separating x and y is longer relative to that in 1 (a). Because the pitch range falls to a greater extent in 1 (b), y is lower relative to x, i.e. d is larger in 1 (b). Thus, if declination is present in a set of utterances, and assuming no phonetic rules which might otherwise affect the F0, the difference between the F0 values of two points within a sentence is directly proportional to the length of the interval between the two points.

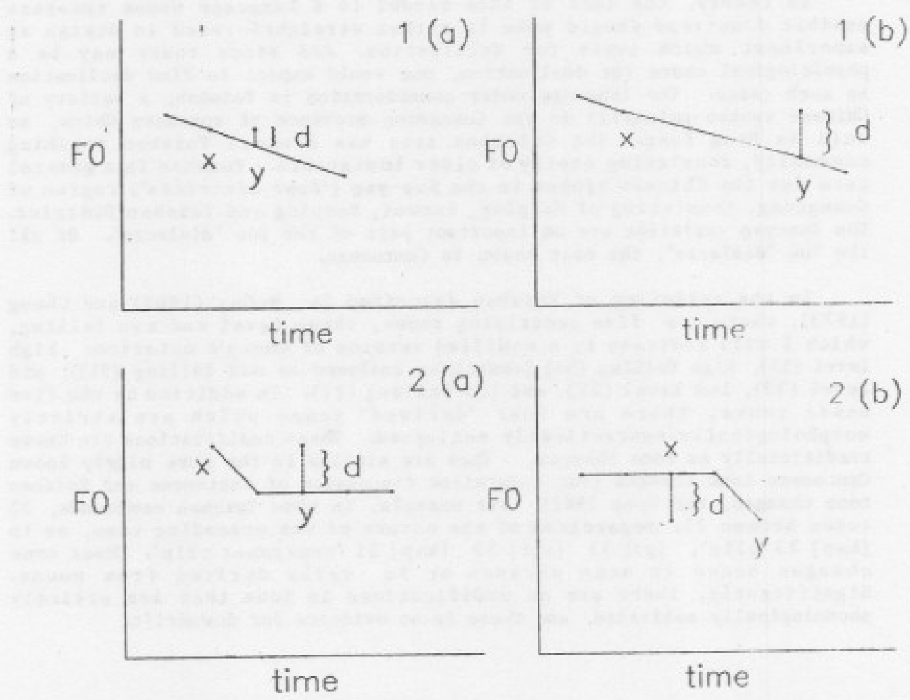


Figure 1: Declination as a function of time, and the consequential differences between the heights of points x and y, based on (a) short interval and (b) long interval lengths. Figure 2: Differences in height due to boundary effects of (a) initial raising and (b) final lowering, in each case without declination.

However, there is some question as to whether a declination component is necessary in a model of intonation. It may be the case that what appears to be a declination effect is actually the effects of initial raising and final lowering rules, particularly in short utterances. In 2 (a), the speaker raises the initial portion of the pitch range without applying declination. Since point x is higher, d is larger than it would be if x occurred later in the utterance. Final lowering, as described by Liberman and Pierrehumbert (1984), is a lowering and compression of the pitch range resulting from a speaker's anticipation of the end of the utterance. 2 (b) shows the effect of final lowering without declination. Point y is lower, and hence d is larger than it would be if y occurred earlier (relative to the end). If d can be accounted for exclusively in terms of such boundary effects--and if utterance length is not a significant factor--then it is unnecessary to posit a declination component.

In theory, the lack of tone sandhi in a language whose speakers exhibit downtrend should make it rather straightforward to design an experiment which tests for declination. And since there may be a physiological cause for declination, one would expect to find declination in such cases. The language under consideration is Toishan, a variety of Chinese spoken primarily in the Guangdong province of southern China, as well as Hong Kong.[1] The Columbus area has a small Toishan speaking community, consisting mostly of older immigrants. Toishan is a general term for the Chinese spoken in the Sze-yap ('four districts') region of Guangdong, consisting of Hoiping, Sunwui, Yanping and Toishan District. The Sze-yap varieties are an important part of the Yue "dialects". Of all the Yue "dialects", the best known is Cantonese.

In the varieties of Toishan described by McCoy (1969) and Cheng (1973), there are five underlying tones, three level and two falling, which I will indicate by a modified version of Cheng's notation: high level (55), high falling (52) (sometimes analyzed as mid falling (31)), mid level (33), low level (22), and low falling (21). In addition to the five basic tones, there are four "derived" tones which are strictly morphologically/syntactically motivated. These modifications are known traditionally as tone changes. They are similar to the more widely known Cantonese tone changes (for a detailed discussion of Cantonese and Toishan tone changes, see Wong 1982). For example, in some Toishan compounds, 33 tones become 21, regardless of the nature of the preceding tone, as in [kap] 33 'clip', [pɔ] 33 [tʃi] 55 [kap] 21 'newspaper clip'. Most tone changes occur in noun phrases or in verbs derived from nouns. Significantly, there are no modifications in tone that are strictly phonologically motivated, and there is no evidence for downdrift.

----

[1] This language is also known as Taishan(-ese). In the language itself it is called Hoisan-wa. Following Light (1981), who follows an old postal usage, and also to avoid confusion with phonetically similar names, I will use the name Toishan.

1.2. F0 implementation: global vs. local

In addition to the issue of declination, there is also the issue of the relationship between global trends and local events. In Thorsen's hierarchical representation of intonation, global patterns are direct, independent representations of a higher level component, such as the textual, sentence or phrase level. Using evidence from Standard Danish, Thorsen (1985) found that sentences in initial, medial and final positions within a text were differentiated by a sequential, evenly distributed declination. Lindau (1986) proposes a similar "layered" representation for Hausa, where the F0 contour is described in terms of an interaction between global and local factors.

In contrast, Liberman and Pierrehumbert (1984) argue that F0 is implemented only by means of locally triggered rules (a linear (or tonal) sequence view): " . . . we find no clear evidence for phrase level planning . . . The major factors shaping F0 contours appear to be local ones" (p. 166). In this model, the global trend is merely an artifact of the application of such local rules. A second point of interest (though not considered in this paper) involves the different claims these two views make concerning preplanning. In the "global" view, the implementation of F0 is by necessity preplanned. In the "local" view, preplanning is limited to phrase-level changes in pitch range specification (which has a direct effect only on the first phrasal pitch accent) and final lowering. The Toishan study may provide evidence for the preference of one view over the other.

2. Methods

The study involved three native speakers. Subject GFL was seventy-four years old at the time of the elicitation; he was born in Wailung Village, Sunwui District. He first came to the U. S. in 1927 before settling in Columbus after marriage. FKL, a sixty-seven year old female, was born in Taichong Village, Hoiping District, and came to the U. S. in approximately 1940. THY, born in the same village as FKL, was fifty-seven years old. He lived in Hong Kong before coming to the United States in 1982. GFL speaks fluent English; FKL understands some English, but does not speak it fluently. THY speaks little English. All three understand and can speak some Cantonese. None had any formal education beyond the grade school level.

Each read aloud a text consisting of four paragraphs. Each paragraph contained five pragmatically connected sentences. The story was about an elderly man and his two pets, a dog and a monkey. The sentences were neutral declarative sentences. For example, the first paragraph was: "Lɔ pak yu kaw toy. Kaw toy hay ɬay koy kaw toy. Kaw toy hay lɛŋ ɬuy. Lɔ pak tʃoŋyi həŋ kaw fan. Kaw toy han lɔ pak." (Which roughly translates as: 'Old Uncle has a puppy. The puppy is a small puppy. The puppy is two years old. Old Uncle likes to play with the dog. The puppy loves Old Uncle.') The corpus is given in its entirety in the appendix.

Each sentence had two target tones. The first target tone, which I will designate as t1, was always sentence initial; the second, t2, was always in penultimate position. Since tone changes do not affect underlying high level tones in Toishan, all target tones were high level.

The intervening amount of material between the target tones was either two syllables--considered a "short" interval, or four syllables--a "long" interval. To test for declination, one would need to see if the difference between the F0 values in a sentence such as "Lɔ pak yu kaw toy"--where lɔ and kaw are the test syllables--is smaller than in "Ma law yet ɬɛŋ lam lɔ pak"--where a longer interval separates the test syllables ma and lɔ.

Paragraphs 1 and 3 were similar in structure, in that the interval between each target tone alternated between two and four syllables, with the pattern 2-4-2-4-2. Paragraphs 2 and 4 contained the reverse pattern, 4-2-4-2-4.

Since a minimal amount of background noise on the recordings does not affect a program's capability to track F0, I was able to record GFL and FKL at their home, and THY in a relatively quiet area at work. A potential advantage of this was that the subjects are as comfortable as possible, which may result in more natural readings. I recorded GFL and FKL using their SONY cassette deck; each wore a SONY ECM-16 clip-on microphone. THY was recorded with a portable cassette recorder containing a built-in microphone.

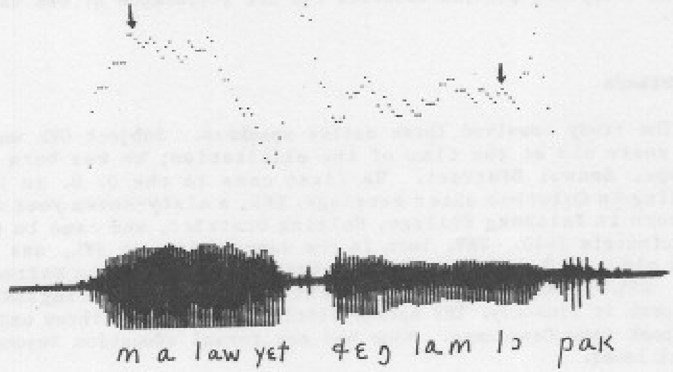


Figure 3: A sample F0 contour for speaker GFL. The arrows indicate measurement points.

Each speaker had a brief run-through session before the recording began. The subjects read each paragraph several times before going on to the next one. This appeared to be the best strategy (as opposed to reading the text in its entirety) because they were not fluent at reading aloud. They made numerous errors in reading (mainly by way of anticipatory mistakes), and reading by paragraph was the best way of obtaining good readings. The number of repetitions of each paragraph varied from paragraph to paragraph and from speaker to speaker; I simply asked them to repeat the paragraph until they had read it smoothly several times. FKL had a minimum of four usable repetitions. GFL and THY's reading went more smoothly than FKL's; they each had a minimum of seven usable repetitions. In all, a total of nearly 400 tokens were analyzed.

The recordings were digitized at a sampling rate of 10000 Hz on a DEC PDP 11/23, using the ILS signal processing program. Using LPC analysis, I obtained the fundamental frequency contour for each sentence. Figure 3 shows how the measurements of the target tones were made. The target tone value was the first peak F0 after the onset of the vowel. The initials consisted of both sonorants and obstruents. Since obstruents affect the F0 of adjacent vowels, the measurements were taken far enough into the vowel to eliminate such segmental effects.

3. Results

3.1. Effects of interval and position

For all three speakers, the ratio of t1/t2 is, in general, slightly greater than 1.00. Thus, the value of t2 was usually lower than t1. All three speakers showed a small amount of downtrend which was significant at the .001 level. However, the downtrend cannot be attributed to declination. The results of a two-way ANOVA indicate that the difference between the short interval sentences and the long interval sentences was not significant at the .05 level for any of the three speakers. This is shown in Figure 4, where the t1/t2 ratio (y-axis) is plotted against sentence position and interval length (x-axis). The lower the value of t2 is relative to t1, the larger the t1/t2 ratio. The open circles represent short interval sentences, the filled circles, long interval.

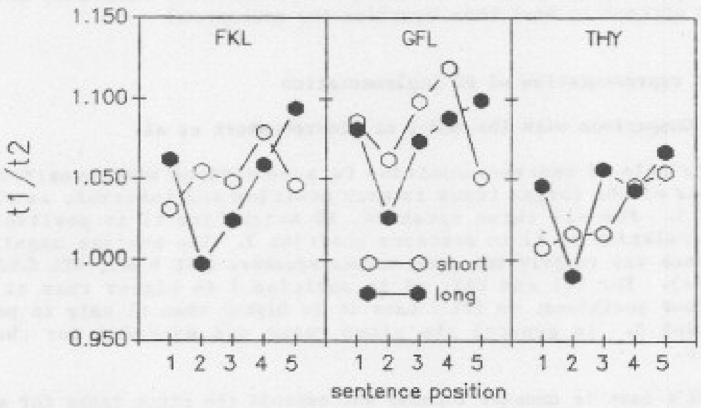


Figure 4: Average t1/t2 ratios (y-axis) according to both interval length and position within the paragraph (x-axis) for all three speakers. Open circles indicate short interval sentences; filled circles, long interval. If a declination effect is present, the filled circles should be consistently higher than the open circles. Hoever, none of the speakers show declination.

If declination is present, the filled circles should be consistently higher than the open circles (i.e. the ratios should have a relatively lower t2). This is clearly not the case for any of the three speakers. For GFL, the opposite occurs: the short interval sentences have a larger rate of downtrend in sentences 2 through 4, and roughly the same in sentence 1. FKL has a larger long interval downtrend only for positions 1 and 5, the peripheral sentences. THY likewise shows no declination effect.

The possibility of an interaction between position and interval was examined as well, where position refers to the order of each sentence within the paragraph. Perhaps early sentences with a short interval may show a smaller amount of downtrend than later sentences with a long interval. Across speakers, there was an unusual pattern involving a "dip" in the long interval sentences from sentence 1 to sentence 2, then a gradually increasing ratio from sentences 2 through 5, meaning that t2 became lower relative to t1. This is not true for the short interval sentences. This pattern, however, does not appear to be important. (A possible explanation is given below.) For all speakers, the interaction of position and interval was significant at the .05 level, but not at the .01 level.

Thus, there is no pattern of declination. But it would be interesting to see if the interval factor is significant in cases where the "long" interval is longer than four syllables--perhaps comparing two vs. eight syllable intervals. There is the possibility that declination may not have had an obvious effect in the current study because the difference between the syllable intervals was not large enough. The sentences were relatively short sentences because I did not want to make the reading difficult for the subjects. I had also hoped that they might with practice memorize the paragraphs and speak more naturally. (Because they made many errors, I did not attempt to have them memorize the sentences.)

3.2. A representation of F0 implementation

3.2.1. Comparison with the model of Pierrehumbert et al.

The role of sentence position is more obvious when considering the averages of the target tones in each position and interval, as shown in Figure 5. For all three speakers, F0 height for t1 in position 1 was higher relative to t1 in sentence position 2. The average magnitude of difference was roughly the same across speakers (FKL 6.903, GFL 8.6255, THY 8.929 Hz). For GFL and THY, t1 in position 1 is higher than t1 in the other four positions; in FKL's case it is higher than t1 only in positions 2, 3, and 5. In general the pitch range was expanded for the first sentence.

FKL's case is unusual because she expands the pitch range for sentence 4. The average value of t1 in sentence 4 is very close to that of t1 in sentence 1, showing a "resetting" of the pitch range. There was no semantic reason for her to do this. However, as noted earlier, FKL had difficulty reading the text, more so than the other speakers. The best explanation is that she split the text in half to make the reading easier.

In the production of sentence 5, GFL lowers and compresses the pitch range. t1 in position 5 is lower relative to positions 3 and 4, and the difference is significant at the .01 level. t2 in those positions remains the same. In the production of sentence 5, THY lowers the pitch range. t1 in position 5 remains the same, but t2 is lower relative to 3 and 4 (although the difference is not significant).

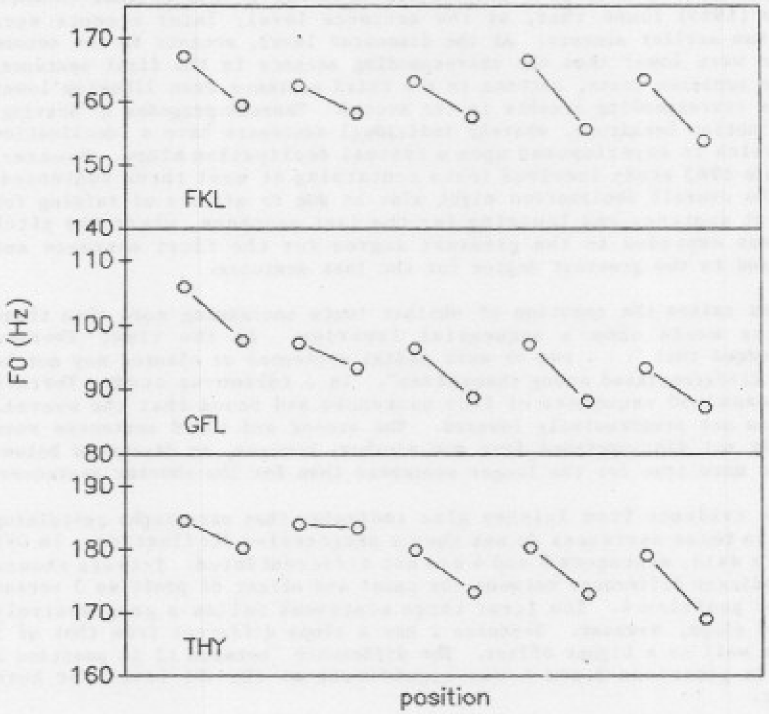


Figure 5: Averages of t1 and t2 by position for each speaker. The first connected pair represents t1 and t2 in sentence 1, and so on through t1 and t2 in sentence 5. The pitch range is expanded for the first sentence (which initiates the topic) and is compressed or lowered for the last sentence (which concludes the topic).

These results are consistent with Hirschberg and Pierrehumbert's descriptions of pitch range variations as pragmatic markers (1986). The position of the sentence within the discourse unit is signalled by its intonation. An initiation in topic is reflected by an increase in pitch range. The larger the change in topic, the larger the increase in the degree of expansion. End sentences are marked as well; if a sentence completes a topic, the pitch range is lowered and compressed to a relatively greater degree than for non-terminal sentences. The degree of final lowering depends on the position of the sentence relative to the discourse structure.

3.2.2. Comparison with Thorsen's model

A second point which is brought out when looking at tonal averages involves Thorsen's idea of a progressive lowering in the textual contour. Thorsen (1985) found that, at the sentence level, later accents were lower than earlier accents. At the discourse level, accents in the second sentence were lower than the corresponding accents in the first sentence; in three sentence texts, accents in the third sentence were likewise lower than the corresponding accents in the second. Thorsen proposes a "nesting" of declination backdrops, whereby individual sentences have a declination slope which is superimposed upon a textual declination slope. However, Thorsen's 1985 study involved texts containing at most three sentences. Thus, the overall declination might also be due to effects of raising for the first sentence and lowering for the last sentence, where the pitch range was expanded to the greatest degree for the first sentence and compressed to the greatest degree for the last sentence.

This raises the question of whether texts containing more than three sentences would show a sequential lowering. At the time, Thorsen acknowledged that ". . . two or more medial sentences or clauses may not be further differentiated among themselves". In a follow-up study, Thorsen (1986) examined sequences of four sentences and found that the overall slope was not progressively lowered. The second and third sentences were generally not distinguished from one another. However, as discussed below, this was more true for the longer sentences than for the shorter sentences.

The evidence from Toishan also indicates that paragraphs containing more than three sentences do not show a progressive declination. In GFL and THY's data, sentences 3 and 4 are not differentiated. T-tests showed no significant difference between the onset and offset of position 3 versus those of position 4. The first three sentences follow a progressively downward slope, however. Sentence 2 has a slope different from that of 3 and 4 as well as a higher offset. The difference between t2 in position 2 and t2 in positions 3 and 4 was significant at the .01 level for both speakers.

However, the difference between position 2 and positions 3 and 4 could be explained in terms of experimental design. The vowels in the target syllables were not always of the same height within each sentence. The syllables containing target tones had mid and low vowels (i.e. [ɔ], [a], and [aw]). The measurement in [aw] was made at a point in the [a] portion, and thus [aw] was considered equivalent to [a]. In cases where a mid vowel (usually [ɔ]) occurred in t1 and a low vowel (always [a]) in t2, there is the possibility that F0 might decrease for t2, resulting in a higher ratio. Conversely, a low-mid pattern might cause F0 to increase for t2, resulting in an intrinsically lower ratio. It essentially boils down to a difference between [ɔ] and [a].

In all positions except for position 2, there was a mixture of three patterns (mid-low, low-mid and same vowel), with half of the tokens containing one pattern and one-fourth containing each of the other two.

Position 2 differs from the others in that half had the same vowel, while the other half had a low-mid pattern. This suggests that the value of t2 may be intrinsically higher here than in the other positions, and that is precisely the pattern found in GFL and THY's cases. This might also explain the drop in the t1/t2 ratio for position 2 in the long interval sentences. Ideally the vowels should be of the same height. However, the extent to which vowel height affects F0 in the case of [ɔ] and [a] is not clear--especially in this case, where both vowels are close in height.

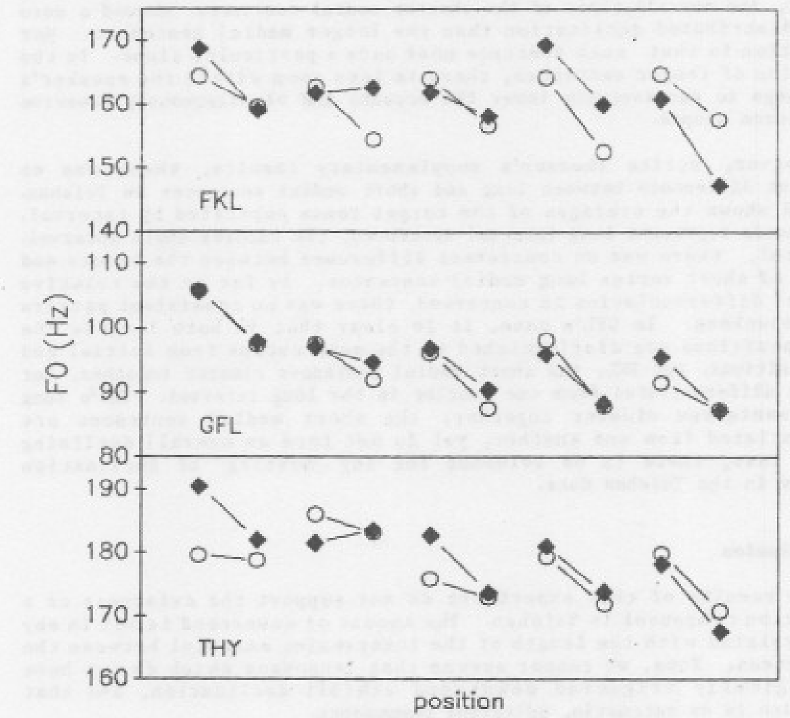


Figure 6: Averages of t1 and t2 by position for each speaker, separated by interval. The first connected pair represents t1 and t2 in sentence 1, and so on through t1 and t2 in sentence 5. Diamonds represent long interval sentences; circles, short interval. There is no difference between the onsets and offsets of long versus short sentences. Across speakers, there is no consistent pattern involving the relative degree of differentiation among medial sentences in both types of intervals.

Unlike the other speakers, FKL shows no difference between the second and third positions. This, together with the possibility that the relationship between the vowels in second position may have affected the quality of t2 for the other two speakers, suggests that under optimal conditions the three medial sentences might not have been distinguished from one another.

Thorsen (1986) also found that the shape of the textual contour differed depending on sentence length. The onsets and offsets of longer medial sentences (three stress groups) were higher and lower, respectively, than the onsets and offsets of shorter sentences (two stress groups). Moreover, the overall slope of the shorter medial sentences showed a more evenly distributed declination than the longer medial sentences. Her explanation is that each sentence must have a particular slope. In the production of longer sentences, there is less room within the speaker's pitch range to successively lower the accents and simultaneously preserve the sentence slopes.

However, unlike Thorsen's supplementary results, there was no consistent difference between long and short medial sentences in Toishan. Figure 6 shows the averages of the target tones separated by interval. The diamonds represent long interval sentences, the circles short interval. In general, there was no consistent difference between the onsets and offsets of short versus long medial sentences. As far as the relative degree of differentiation is concerned, there was no consistent pattern across speakers. In GFL's case, it is clear that in both lengths the medial positions are distinguished to the same extent from initial and final positions. For FKL, the short medial sentences cluster together, but are more differentiated from one another in the long interval. THY's long medial sentences cluster together; the short medial sentences are differentiated from one another, yet do not form an overall declining slope. Thus, there is no evidence for any "nesting" of declination backdrops in the Toishan data.

4. Conclusion

The results of this experiment do not support the existence of a declination component in Toishan. The amount of downtrend is not in any way correlated with the length of the intervening material between the target tones. Thus, we cannot assume that languages which do not have phonologically triggered downtrend exhibit declination, and that declination is an automatic, universal phenomenon.

The results are, on the other hand, consistent with a view of local F0 implementation. The positions of initial and final sentences are marked in the overall slope, and medial sentences as a whole are distinguished from the peripheral sentences (though probably not from each other). The hierarchical backdrop declination effects proposed in Thorsen's model did not appear in the five sentence paragraphs in the Toishan data. The results of the current study argue against a global implementation of F0.

Finally, the results are of particular interest when compared with the results of a parallel, simultaneous experiment involving Cantonese (see Johnson, this volume). Toishan and Cantonese are, to a large extent, structurally similar. In addition to finding evidence for initial raising, Johnson found that interval length was a significant factor--suggesting the effect of a declination component. Not only is this finding important in terms of language relatedness[2], but also in terms of language contact. Both Toishan and Cantonese are spoken in the same geographical areas, and many speakers can understand both varieties. Many speakers of Toishan are fluent in Cantonese as well (although the reverse is not always true; this may be due to the sociolinguistic standing of the two varieties, since Toishan has the position of lower prestige). Thus, some aspects of downtrend are cross-linguistic, but others are language specific.

---

[2]Chiu-yu Tseng (1984) found no evidence for a declination component in Mandarin ("Sentence intonation is context/discourse oriented: evidence from Mandarin Chinese", abstract, Journal of the Acoustical Society of America, Suppl. 1, Vol. 75, Spring).

Acknowledgements

A preliminary version of this paper, "Declination in Toishan", was presented at the 111th meeting of the Acoustical Society of America, Cleveland, May 12-16 1986. My thanks to Mary Beckman for her advice and comments throughout the investigation. I am also grateful to Timothy Light for sharing his unpublished work on Toishan. Keith Johnson and Ken de Jong gave me much of their time in discussing these issues and (especially) in relieving my occasional frustrations with the 11/23.

References

Beckman, Mary E. and Janet B. Pierrehumbert. 1986. "Intonational Structure in Japanese and English". Phonology Yearbook 3:255-309.

Cheng, Teresa M. 1973. "The Phonology of Taishan". Journal of Chinese Linguistics, Volume 1, Number 2 (May):256-322.

Hirschberg, Julia and Janet Pierrehumbert. 1986. "The Intonational Structuring of Discourse". Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics:136-144, June 10-13, Columbia University, New York.

Ladd, D. Robert. 1983. "Declination: A Review and Some Hypotheses". Phonology Yearbook 1:53-74.

______________. 1983. "Phonological Features of Intonational Peaks". Language 59:4 721-759.

Liberman, Mark and Janet Pierrehumbert. 1984. "Intonational Invariance under Changes in Pitch Range and Length". Language Sound Structure 157-233, Mark Aronoff and Richard T. Oehrle, ed. MIT Press.

Light, Timothy. 1981. "Toishan: An American Language (Problems for Learners of Standard Chinese)". Paper presented at the meeting of the National Association for Asian and Pacific American Education (April).

Lindau, Mona. 1986. "Testing a Model of Intonation in a Tone Language". *Journal of the Acoustical Society of America* 80:757-764.

McCoy, (William) John. 1969. "The Phonology of Toishan City". *Orbis* 18:108-122.

Pierrehumbert, Janet Breckinridge. 1980. *The Phonology and Phonetics of English Intonation*. PhD dissertation, MIT.

Thorsen, Nina Grønnum. 1985. "Intonation and Text in Standard Danish". *Journal of the Acoustical Society of America* 77:1205-1216.

________________. 1986. "Sentence Intonation in Textual Context--Supplementary Data". *Journal of the Acoustical Society of America* 80:1041-1047.

Wong, Maurice. 1982. *Tone Change in Cantonese*. PhD dissertation, University of Illinois. University Documents, Ann Arbor.

Appendix: Corpus

Target tones are indicated by an asterisk. Toishan also has a predictable alternation between long and short tones which is not under consideration here and, thus, is not indicated. In the noun phrase [lɔ pak], the tone on [pak] is underlyingly 33, which becomes 21 by a morphologically conditioned tone change. [pak] means 'paternal uncle', but can also be used, as in this situation, as a term of respect for an unrelated adult male. Because there is no exact equivalent in English, I have translated it as 'uncle'.

Paragraph 1:

*55	21	33	*55	55			
lɔ	pak	yu	kaw	toy			'Old Uncle has a puppy.'

*55	55	31	31	3	*55	55	
kaw	toy	hay	ɬay	koy	kaw	toy	'Puppy is a little puppy.'

*55	55	31	*55	33	
kaw	toy	hay	lɛŋ	ɬuy.	'Puppy is two years old.'

*55	21	33	33	33	*55	55	
lɔ	pak	tʃoŋ	yi	haŋ	kaw	fan.	'Old Uncle likes (to) play with (the) dog.'

*55	55	31	*55	21	
kaw	toy	han	lɔ	pak.	'Puppy loves Old Uncle.'

Paragraph 2:

*55	21	33	33	22	*55	55	
lɔ	pak	yɛt	yu	fi	ma	law	'Old Uncle also has (a) fat monkey.'

*55	55	21	*55	33	
ma	law	hay	lɛŋ	ɬuy	'Monkey is 2 years old.'

*55	55	33	33	33	*55	55	
kaw	toy	m̩	tʃoŋ	yi	ma	law	'Puppy doesn't like Monkey.'

*55	55	33	*55	55	
kaw	toy	tʃuy	ma	law	'Puppy chases Monkey.'

*55	55	55	55	22	*55	55	
kaw	toy	lɛŋ	ŋaw	fi	ma	law	'Puppy wants to bite (the) fat monkey.'

Paragraph 3:

```
*55  55   55  *55  55
kaw  toy  haw  ŋaw tʃi
```
'Puppy (with) mouth bites paper.'

```
*55  55   55  33  55   *55  21
kaw  toy  i   pɔ  tʃi   lɔ   pak
```
'Puppy gives (the) newspaper to Old Uncle'.

```
*55  21   33  *55  55
lɔ   pak  han  kaw  toy
```
'Old Uncle loves Puppy.'

```
*55  55   55   22  33   *55  21
kaw  toy  ɬɛŋ  na  ʃɛŋ   lɔ   pak
```
'Puppy wants to climb up on Old Uncle.'

```
*55  21   55   *55  55
lɔ   pak  lam  kaw  toy
```
'Old Uncle hugs Puppy.'

Paragraph 4:

```
*55  55   33   55   55   *55  21
ma   law  yɛt  ɬɛŋ  lam  lɔ   pak
```
'Monkey also wants to hug Old Uncle'.

```
*55  55   55   *55  55
ma   law  ŋaw  kaw  toy
```
'Monkey bites Puppy.'

```
*55  21   33     33  33  *55  55
lɔ   pak  fay[1] nɔ  fi  ma   law
```
'Old Uncle quickly scolds Monkey.'

```
*55  21   33   *55  55
lɔ   pak  han  ma   law
```
'Old Uncle loves Monkey.'

```
*55  55   33   22  33   *55  21
ma   law  yɛt  na  ʃɛŋ  lɔ   pak
```
'Monkey also climbs up on Old Uncle.'

[1] Read as [kʊt] by THY. THY also found the characters for 'monkey' unusual, but read them with the correct intonation.

OSU #36: Papers from the Linguistics Laboratory 1985-1987, Sept. 1987.

The Tonal Behavior of Osaka Japanese: An Interim Report

Shiro Kori
Osaka University of Foreign Studies

Abstract: This paper reports the results of a preliminary investigation of the F0 characteristics of Osaka Japanese. It describes the realization of word accents in sentences and the interaction of focus with word accents. It discusses the basic unit of phrasing in Osaka Japanese sentences, and suggests that this unit is more closely tied to the word than in Tokyo Japanese because a word does not easily lose its acccentual integrity in larger prosodic contexts. Several other important differences from Tokyo Japanese tonal patterns are also revealed. First, Osaka has two distinct kinds of sharp F0 falls, one at the accent nucleus (as in Tokyo) and the other at the boundary of a High-ending word and a following Low-beginning word. Second, while the F0 contour over a sequence of supposedly H-pitched moras has a slow decline, the rate of this downtrend is constant and independent of the size of sequence, suggesting that, unlike in Tokyo, there is H-tone spreading in Osaka Japanese. Uptrends in L-beginning words are also examined, as well as the interaction of L tones with focus.

0. Introduction

While there are a considerable number of experimental works devoted to the Japanese spoken in Tokyo, the Japanese spoken in Osaka and its neighbouring area has attracted little attention[1]. However, this variety of Japanese has several features which offer a very interesting and fruitful perspective for the study of prosody. For example, Osaka Japanese (henceforth OJ) has words whose constituting moras are all Low-pitched. How can one put focus on these words in sentences? Thus, how can one focus Nomura-san in a sentence such as:

(1) Are-wa Nomurasan-no kuruma-desu.
'That is Mr. Nomura'scar.'

where all the moras of Nomurasan-no 'Mr.Nomura's' are Low-pitched and all the other moras of the sentence are High-pitched? The answer to this question will contribute much to a general theory of focus and intonation. OJ also seems to have interesting characteristics of laryngeal control. In many languages, including Tokyo Japanese (TJ), pitch is believed to be controlled mainly by the onset and offset of cricothyroid activity, which correspond to the onset and offset of pitch rise. Sugito and Hirose (1978)

----

[1]See Appendix 1 for a discussion of the socio-linguistic status of Osaka Japanese with respect to Tokyo Japanese, and for a brief review of earlier work on Osaka and other related dialects.

found, however, that some OJ speakers have a systematic sternohyoid activity concomitant with a Low-pitched mora. The present acoustic study also yields some data which support the existence of an active lowering at word initial Low-pitched moras. Current theories of laryngeal control might need to be modified after a more detailed examination of OJ. Also, the generative F0 model proposed by Fujisaki and his colleagues, which is based on the onset and offset of F0 rise and claimed to be valid for OJ, would require modification within this perspective.

This paper gives a description of the tonal structure of Osaka Japanese, based on the acoustic analysis of more than 300 utterances. Section 2 provides a summary description of traditional analyses of word accent in OJ. Section 3 examines how word accent is realized acoustically in sentences, taking its pragmatic role (focus) into account. Section 4 examines the effect of focus per se. Section 6 presents a test to see whether OJ sentence contours can be explained by the interpolation between a few sparsely placed target tones, as has been proposed by Pierrehumbert and Beckman (in press) for Tokyo Japanese. I will refer to this sort of model as a 'target-tone model'. Since my examination of the realization of word accent in sentences has led me to an opinion that the basic unit of phrasing in OJ corresponds to the word, and because the validity of the target-tone model is dependent on the definition of such a prosodic unit, I will first discuss the problem of the basic accentual unit in section 5 before presenting the results of the test for the target-tone hypothesis.

1. Methods

The methods used in obtaining the acoustic analysis are as follows: Utterances with various syntactic structures and with variable focus location[2] were analyzed using pitch-tracking routines in ILS to examine the realization of word accents in sentences. Three sets of utterances were used to test the effect of focus and the validity of the target tone hypothesis. The phonological structure of these sets of utterances will be described later. In order to obtain continuous F0 contours, the sentences used in the experiments were chosen in such a way that most of the constituting sounds are voiced. All utterances were uttered by the author himself, who was born and brought up in the city of Osaka. An attempt was made to utter the sentences as if they appeared in naturally occurring conversation, without emphasis or emotions. In this paper only F0 characteristics will be discussed and no other possible correlates of accent will be taken into account.

---

[2] My notion of focus here corresponds to Dik et al.'s (1981) 'completive focus'. A word in completive focus carries the most important information, in the sense that it is meant to fill in blanks in the pragmatic information for the addressee. For example, in John bought coffee. as an answer to What did John buy?, the word coffee has completive focus. Completive focus is distinguishable from 'selective focus', 'replacing focus', 'expanding focus', 'restrictive focus' and from 'parallel focus'. See Dik et al. (1981) for further discussion.

2. The Osaka Japanese accent system

Tokyo Japanese can be labeled an 'accent language' in the sense that the tonal pattern of a word is automatically determined by specifying only the location of an 'accent nucleus'[3], which is realized as a local fall in pitch. OJ, on the other hand, has some characteristics of a tone language. One of these characteristics is that OJ distinguishes two types of word pitch pattern for the same location of the nucleus, as in:

(2) a. kaminari 'thunder' b. nokogiri 'saw'

In the first type, all the moras[4] that precede the nucleus and the nucleus-bearing mora itself are High-pitched, while in the second one all the moras before the nucleus are Low-pitched. (I will distinguish these two types by calling the first a High-beginning pattern and the second a Low-beginning pattern.) According to the phonological interpretation of scholars such as Kindaichi (1950), Hirayama (1960) and McCawley (1968), the second type of word has a pitch fall (an accent nucleus) preceding the initial syllable. (However, as we will see later in section 3, this initial 'fall' in Low-beginning words is acoustically distinguishable from the fall at the usual medial accent nucleus.) Phonetic characterizations include Ikeda's (1942) claim that the stretch of low moras preceding the nucleus in Low-beginning words has a continuous pitch rise. Hattori and Kawakami are of the same opinion, but certain other scholars distinguish two low-beginning patterns, "Low flat" and "ascending from Low", a difference which they ascribe to individual option. However, all these researchers agree as to the phonological irrelevance of this pitch rise in Low-beginning words.

In OJ, the nucleus can occur on the word-final mora only in Low-beginning words, and on the word-initial mora only in High-begining words. As in TJ, a word may also have no nucleus. Thus two-mora words have four phonologically possible tonal patterns (instead of the three possible in TJ), as shown in Table I.

Table I. Accent patterns of two-mora words

nucleus location	high-beginning		low-beginning	
1	yama	'mountain'	–	
2	–		ame	'rain'
none	mizu	'water'	sora	'sky'

[3]The term akusento 'accent' in Japanese refers to the overall tonal pattern of a word rather than to any local tonal shape around the accent nucleus, or to the tone-bearing unit associated to that shape. Since this difference in usage causes some confusion for translating between the scholarship in Japanese and that in English, I will avoid the term 'accent', and use instead the term 'nucleus' to refer to the lexically specified location of a pitch fall.

[4]The mora in OJ corresponds to the syllable except for the following cases: moraic nasal, long vowels, diphthongs, double consonants; thus ho-o-ge-n, ha-i-t-ta, etc.

The nucleus usually triggers a perceptible pitch fall at the moraic boundary between the nuclear and post-nuclear moras. However, Low-beginning words with the nucleus on the final mora are pronounced in two ways depending on the following context. When these words are uttered in isolation, a fall occurs in the final nucleus-bearing mora and the final mora is lengthened, as shown in (3a). (Some scholars interpret this lengthening as the word having one more mora and the nucleus being on the penultimate mora). When such words are followed by a Low-pitched enclitic particle, on the other hand, the final mora is not long and remains High throughout, as shown in (3b). This is the traditional description of this pattern.

(3) a. ame 'rain' b. ame-ga ... 'rain-nom...'

Peculiarly, however, this type of pattern is found only among two-mora words and in a few three-mora words. Words of more than three moras with final nucleus are not known. Moreover, there seems to be a great deal of variation among speakers as to the treatment of these words. For example, there are a lot of speakers, especially among young people, for whom the final mora is always High throughout regardless of what is coming next, as shown in (4):

(4) alternative pattern

a. ame 'rain' b. ame-ga ... 'rain-nom. ...'

This alternative seems to be characteristic mainly of speakers in the central and northern part of the province (see Sugito et al. 1981). There seems also to be a third type with a rise-fall pattern in the final mora in all contexts. Yet another source of variation concerns the treatment of this pattern in the context of declarative versus interrogative intonations. The expected pattern in an interrogative contour is LHL¶, as shown in (5), but I found some speakers who say 'ame' with LHL as declarative and with LH¶ as interrogative, and other speakers who have LH as the declarative pattern and LHL¶ as interrogative (see Kori 1984a).

(5) ame 'rain?'

As for the Low-beginning words with no nucleus, their final mora is High-pitched when uttered in isolation or followed by a Low-pitched particle, as shown in (6a-c). When they are followed by a High-beginning word or by a tonally neutral particle (such as the nominative -ga), the final mora assumes Low value, as shown in (6d-f). This rule applies also across a syntactic break or a theme/rheme boundary or even a pause. There is no idiolectal variation here, but note that for some younger speakers the tonal opposition between Low-beginning with final nucleus and Low-beginning words with no nucleus becomes neutralized in some conditions.

(6) a. sora 'sky'

b. sora aoiro-ya-wa 'The sky is blue!'

c. sora-ya-wa 'It's the sky!'

d. sora-ga ... 'sky-nom. ...'

e. sora aoi-wa 'The sky is blue!'

f. sora-ga aoi-wa 'The SKY is blue!'

High-beginning words with no nucleus have only high-pitched moras in all contexts, as shown in (7).

(7) a. mizu 'water' c. mizu abiru-nen '(One) bathes in the water.'

b. mizu-ga ... 'water-nom' d. mizu-ga umai-wa 'The water is good!'

Besides the distinction between High-beginning and Low-beginning patterns and the individual differences in realizing the nucleus on final moras, there are some other factors that render the accent system of OJ rather complicated. One such factor is the behavior of some monosyllabic words composed of a consonant plus a monophthong or only of a monophthong, such as ka 'mosquito', ki 'tree', and ke 'hair'. These words undergo a lengthening of the vowel when uttered in isolation or when not followed by a particle. The lengthening is traditionally considered to make them two moras long. But acoustic examination shows that they are shorter than real two-mora words (Sugito 1981). Some speakers have the longer vowel also in other conditions. There are still other speakers, mostly among young people, who pronounce them short in all conditions. Three tonal patterns -- flat, rising, or falling -- are distinguished for these words in isolation for conditions and speakers that make them long. Conditions or speakers that do not lengthen the vowel, on the other hand, have only two patterns -- high and low -- and the rising pattern is replaced by a low flat pattern in positions before a Low-beginning word. This is parallel to the behavior of nucleusless Low-beginning words before another Low-beginning word.

One final peculiarity of OJ is that the nucleus may occur on a non-syllabic mora -- on a moraic nasal, as in (8a), or on the first half of a geminate consonant, as shown in (8b). Thus, unlike in TJ, the nucleus-bearing unit in OJ is without question the mora. However, for the words in which the nucleus is on a moraic stop, there seems to be always an alternative pronunciation in which the nucleus in such moras is avoided[5].

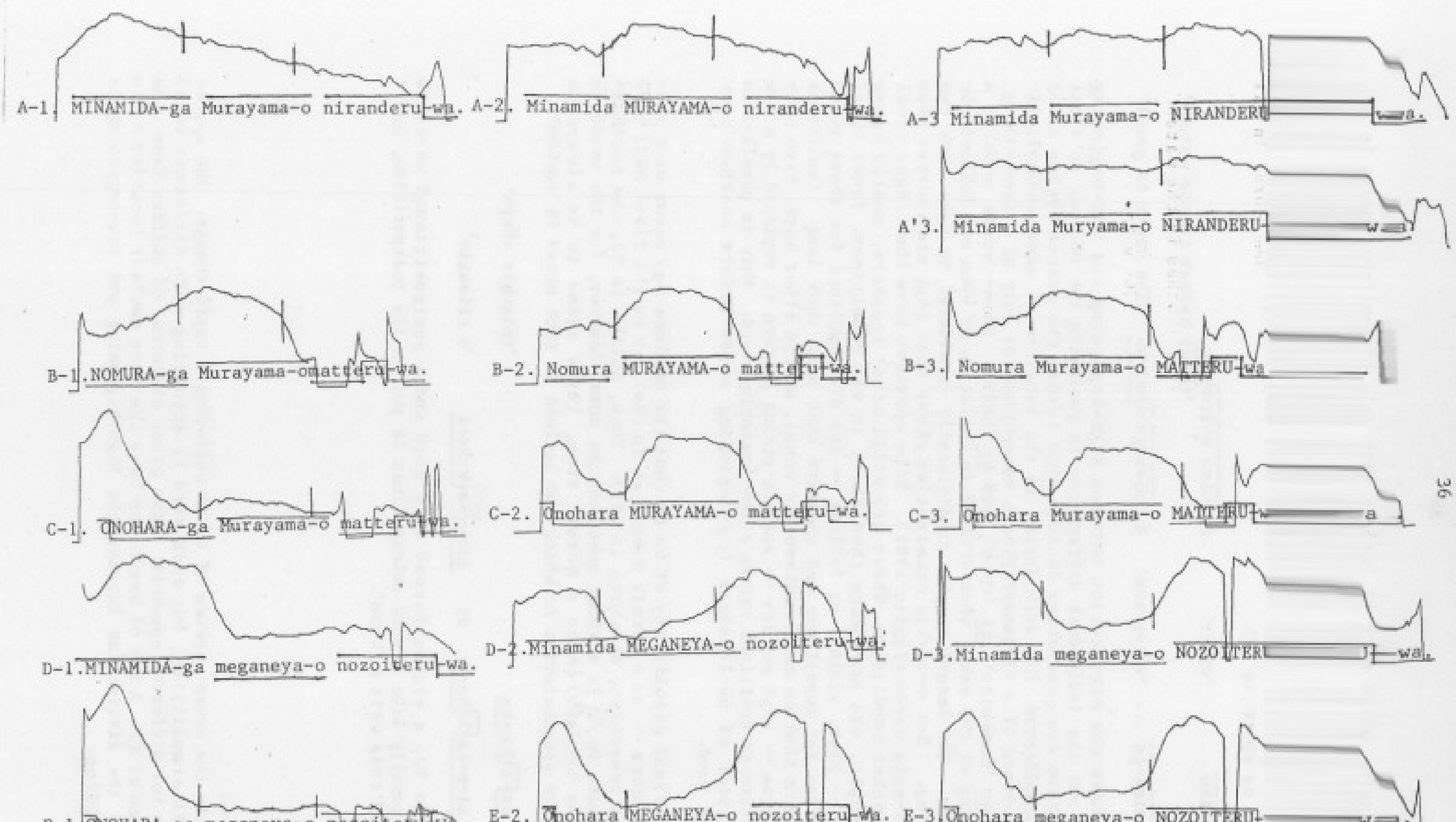
(8) a. sa-n-ka-ku 'triangle shape'

b. sa-n-ka-k-ke-i or sa-n-ka-k-ke-i 'a triangle'

As in TJ, a clause formed by a word and a postpositional particle behaves tonally like a word unless focus is put on the postposition, and I will call this unit a 'word'.

----

[5]Currently the accent system of OJ is suffering a rapid change. The system itself is simplifying, but each word is acquiring more different tonal patterns for different speakers. Like other dialects, OJ suffers from the influence of TJ. A lot of words now have the same nuclear location as in TJ. Yet the distinction between the High-beginning and Low-beginning is still holding.


A-1. MINAMIDA-ga Murayama-o niranderu-wa.
A-2. Minamida MURAYAMA-o niranderu-wa.
A-3 Minamida Murayama-o NIRANDERU-wa.
A'3. Minamida Muryama-o NIRANDERU-wa
B-1. NOMURA-ga Murayama-o matteru-wa.
B-2. Nomura MURAYAMA-o matteru-wa.
B-3. Nomura Murayama-o MATTERU-wa
C-1. ONOHARA-ga Murayama-o matteru-wa.
C-2. Onohara MURAYAMA-o matteru-wa.
C-3. Onohara Murayama-o MATTERU-w a
D-1. MINAMIDA-ga meganeya-o nozoiteru-wa.
D-2. Minamida MEGANEYA-o nozoiteru-wa.
D-3. Minamida meganeya-o NOZOITERU-wa.
E-1. ONOHARA-ga meganeya-o nozoiteru-wa.
E-2. Onohara MEGANEYA-o nozoiteru-wa.
E-3. Onohara meganeya-o NOZOITERU-wa.

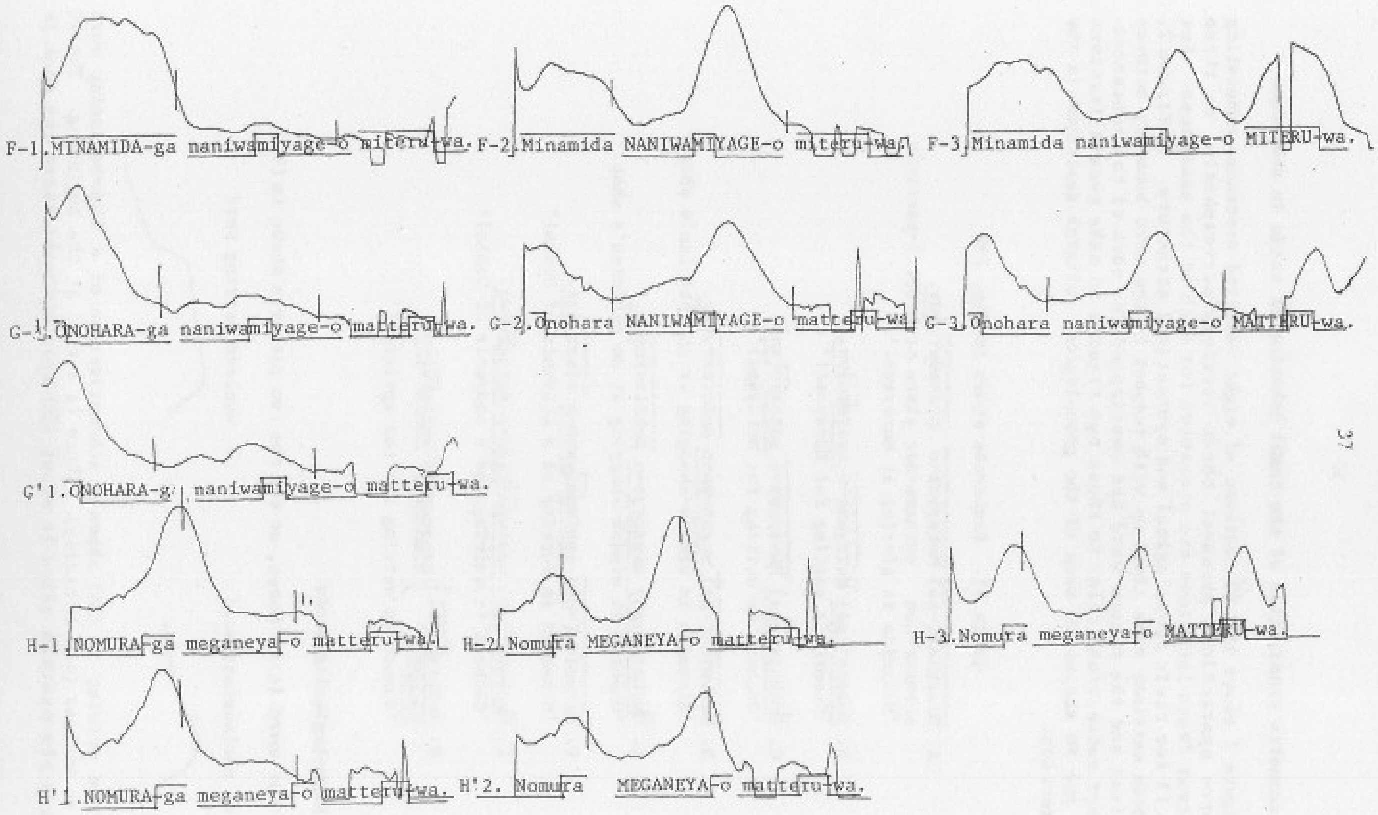


Figure 1. Representative utterances of each type listed in Table II, with each of three focus positions.

3. An acoustic examination of the tonal behavior of words in sentences

Figure 1 shows the F0 contours of eight selected sentences consisting of three syntactical phrases: three versions corresponding to three different focus locations are presented for each of the sentences. See Table II for their phonological and syntactical structure. In Figure 2, the three versions were lined up with respect to the word boundary between the first and the second word and overlapped for each of the sentences. The next subsections refer to these two figures to make generalizations about the F0 shapes of many of the phonological patterns described in the last section.

Table II. Sentences shown in Fig. 1-2

A. Minamida(-ga) Murayama-o niranderu-wa.
surname-nom surname-acc glare at (cont.)-particle
'Minamida is glaring at Murayama!'

B. Nomura(-ga) Murayama-o matteru-wa.
'Nomura is waiting for Murayama!'

C. Onohara(-ga) Murayama-o matteru-wa.
'Onohara is waiting for Murayama!'

D. Minamida(-ga) meganeya-o nozoiteru-wa.
'Minamida is window-shopping at an optician's shop!'

E. Onohara(-ga) meganeya-o nozoiteru-wa.
'Onohara is window-shopping at an optician's shop!'

F. Minamida(-ga) naniwamiyage-o miteru-wa.
'Minamida is looking at a souvenir of Osaka!'

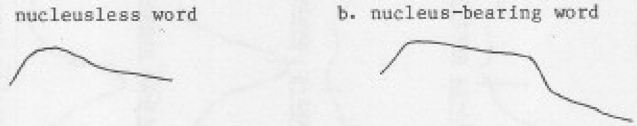
G. Onohara(-ga) naniwamiyage-o matteru-wa.
'Onohara is waiting for a souvenir of Osaka!'

H. Nomura(-ga) } meganeya-o matteru-wa.
Nomura }
'Nomura is waiting for the optician!'

3.1 High-beginning words

When the word is in focus, we see the two patterns shown in (9):

(9) a. nucleusless word b. nucleus-bearing word



The line drawing in (9a) shows a schematization of a High-beginning word with no nucleus (e.g. Minamida). There is a rise at the beginning. The F0 reaches its maximum value in about 150 msec (in most cases the peak is located in the second mora or sometimes in the third mora of the word), and then the F0 declines in a quasi-linear way until the end of the word. If the following word is Low-beginning, there is another steeper fall prior to the word boundary. Otherwise, the gradual declination continues onto the following High-beginning word (see the first words of A-1, D-1 and the second words of A-2, B-2, and C-2 in Fig. 1). The rate of decline in my data is about 60 Hz/sec (8 samples) and the word size does not significantly affect the rate (see section 4.1).

When a High-beginning word has a nucleus (e.g. miteru-wa) the pattern is the same as when there is no nucleus, except that the F0 falls just after the nucleus-bearing mora. This fall, however, is distinct from the

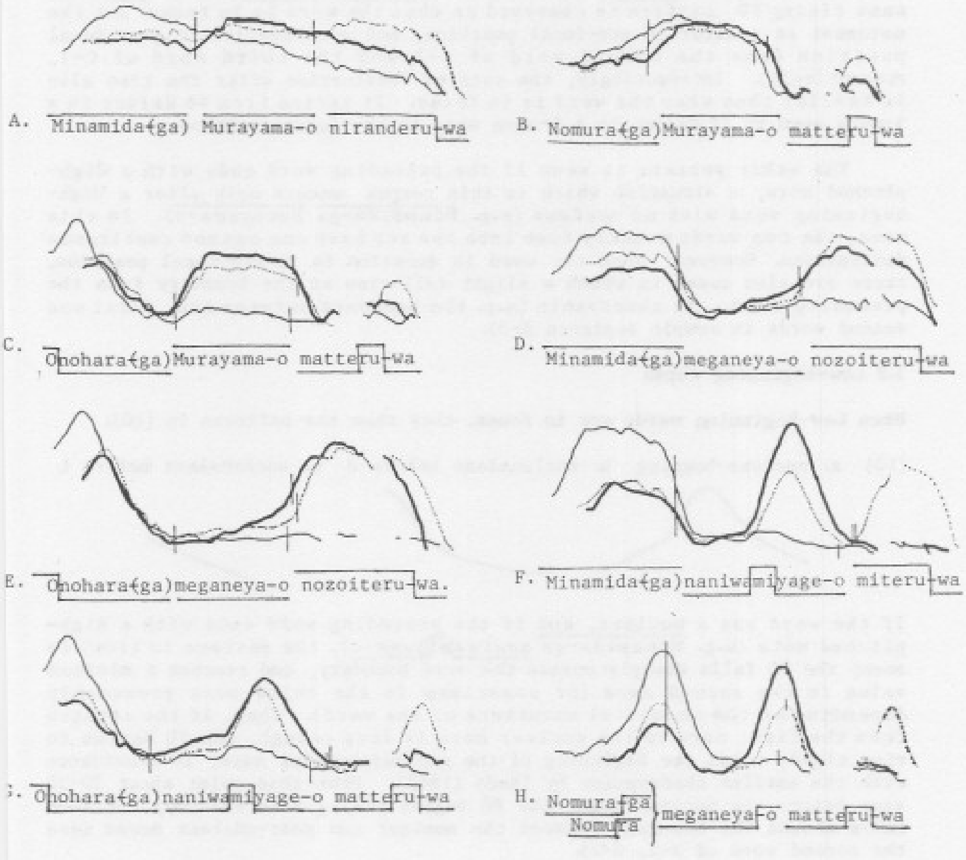


Figure 2. Representative utterances of each type shown in Figure 1, with contours for the three focus patterns overlaid using the word boundary after the first word as the line-up point.

fall at the word boundary between a nucleusless High-beginning word and a following Low-beginning word; there is a difference both in timing and steepness. In the first case, the F0 begins to fall in most cases just after the moraic boundary between the High-pitched nucleus-bearing mora and the following Low-pitched mora. In the second case the fall starts just before the word boundary, usually in the middle of the word-final mora, and the fall has more steepness than the fall at the nucleus. It seems as if the following Low-beginning word forcibly pulls down the pitch. The fall at the nucleus, in contrast, seems a gentle gliding down of pitch.

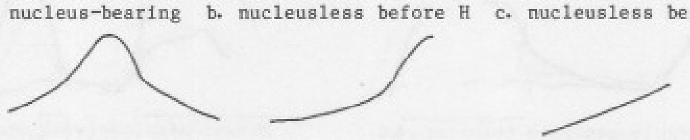
When a High-beginning word is out of focus, there are two possible patterns. If the preceding word is nucleus-bearing and ends with a Low-pitched mora (e.g. Oohara Murayama-o or naniwamiyage-o miteru-wa), the same rising F0 pattern is observed as when the word is in focus. But the movement is smaller in pre-focal position, and even smaller in post-focal position (see the second word of C-3 and the third word of C-1, respectively). Interestingly, the rate of declination after the rise also is smaller than when the word is in focus. It varies from 46 Hz/sec in a 5-mora word to 22 Hz/sec in a 9-mora word in pre-focus position.

The other pattern is seen if the preceding word ends with a High-pitched mora, a situation which in this corpus occurs only after a High-beginning word with no nucleus (e.g. Minamida-ga Murayama-o). In this case, the two words usually fuse into one and have one common continuous declination. However, when the word in question is in pre-focal position, there are also cases in which a slight fall-rise at the boundary from the preceding word is observable (e.g. the boundary between the first and second words in sample sentence A-3).

3.2 Low-beginning words

When Low-beginning words are in focus, they show the patterns in (10):

(10) a. nucleus-bearing b. nucleusless before H c. nucleusless before L



If the word has a nucleus, and if the preceding word ends with a High-pitched mora (e.g. Minamida-ga naniwamiyage-o), the pattern in (10a) is seen; the F0 falls steeply across the word boundary, and reaches a minimum value in the second mora (or sometimes in the third mora presumably depending on the segmental structure of the word). Then, if the stretch from the first mora to the nuclear mora is long enough, the F0 begins to rise slowly until the beginning of the nucleus-bearing mora, in accordance with the earlier observation by Ikeda (1942). From this point about 20-30 msec before the nuclear mora, the F0 begins rising more steeply, and then falls around the boundary between the nuclear and post-nuclear moras (see the second word of F-2, H-2).

When the preceding word ends with one or two Low-pitched moras, the Low-beginning word shows the same pattern, except that the F0 reaches a minimum and the slow rise begins earlier, somewhere between the first and the second mora. When the word is preceded by a word ending with a sequence of more than two Low-pitched moras, the turning point from falling to slow rising is even earlier, right at the word boundary (see the second word of G-2).

When a Low-beginning word has no nucleus, and is followed by another Low-beginning word, the pattern in (10b) is seen; the final mora assumes High value (e.g. Nomura-ga naniwamiyage-o). (This was discussed above around example (5).) As in the initial part of a Low-beginning word with a nucleus, the F0 rises slowly until it makes a sharp upturn to rise steeply for the final High-pitched mora. The rate of this slow upward inclination varies from 71 Hz/sec for a 9-mora word to 149 Hz/sec for a 3-mora word, with a mean of 107 Hz/sec (8 samples). This rate systematically decreases as the word size increases. After reaching a peak, the F0 then falls rapidly. The starting point of this rapid fall is on the word boundary or just before the boundary and not much earlier. Thus, it is different from the fall at the end of a High-beginning nucleusless word followed by a Low-beginning word, probably because here the F0 has to rise in the last mora of the word and there is no time for the F0 to fall in the same mora.

When the nucleusless Low-beginning word is followed by a High-beginning word, the pattern in (10c) is seen; all the moras of the word assume Low value (e.g. meganeya-o nozoiteru-wa). See, for example, the second word of D-2, E-2. In this case the slow rise continues until the end of the word. Interestingly, however, its rate is smaller than when followed by a Low-beginning word (mean= 63 Hz/sec (12 samples): no significant effect of word size). See section 4.2.1 for further discussion about this difference. For the location of the initial F0 minimum, the same rule as for the nucleus-bearing case applies.

The form of the slow rise in Low-beginning words is in most cases a quasi-linear or slightly concave line, as can be seen in Figures 1 and 2, but sometimes it is slightly convex, especially when the word has a nucleus.

When the Low-beginning word is out of focus, the F0 shapes are in general the same as those found when the word is focused, but with a smaller amount of movement. Again, as in a High-beginning word, any movement is much smaller when the word is in post-focal position than in pre-focal position. However, there seem to be two exceptional cases. The first is when a focused word with a High final mora precedes a short Low-beginning word (e.g. Minamida-ga meganeya-o...); here, the slow rise of the short post-focal Low-beginning word is masked by the more prominent fall around the word boundary. The second exception is when a Low-beginning word with nucleus is preceded by a focused word with a final Low mora (e.g. Onohara-ga naniwamiyage-o); here, the Low-beginning word with nucleus has two small F0 peaks, the first corresponding to the slow rise in the other conditions and the second to the rise-fall due to the nucleus (see the second word in samples sentences G-1 and G'-1). We have already seen that the slow rise in a focused Low-beginning nuclear word often assumes a convex shape, which might explain the first peak. On the other hand, these two small F0 peaks might be due to the fact that all the Low-beginning-nucleus-bearing words examined in our corpus are compound words (see section 5.2 for a discussion of the tonal idiosyncrasy of the compound words).

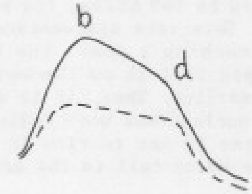
4. The realization of focus

Focus is realized in somewhat different ways depending on the tonal structure of the word in focus. The behavior of the Low-beginning words is particularly interesting. In this section, I will propose "rules" for the realization of focus for the different structures.

4.1 High-beginning words

The effect of focus on High-beginning rules can be schematized as:

(11)



As a rule, focus raises the over-all F0 contour. The amount of raising is greater at the beginning of the slow declination than at its end (points B and D, respectively, in the schematic representation). So the effect of focus could be represented as a curve like this:

(12)

An exception to this rule occurs when the High-beginning word is preceeded by a Low-beginning nucleusless word (e.g. meganeya-o # nozoiteru-wa) and I will treat this case separately later in section 4.2.1.

Figure 3 shows how focus is realized in a set of sentences with the structure:

(13) Uchi-wa # Q # mukoo-ya-nen.
'My house is on the other side of Q'

where Q is a 3-, 5-, 7- or 9-mora High-beginning nucleusless word. The F0 contours in the lefthand column in Figure 3 are versions with focus on the noun in Q and those in the righthand column are versions with focus on mukoo 'over there, the other side' (a noun). These contours are summarized in Figure 4, where each curve consists of linear interpolations between the mean values of some measurement points. The filled circles represent the versions with focus on Q, and the empty circles represent the versions with focus on mukoo. Below each measurement point, ANOVA outputs are summarized. Two repetitions for each of the eight versions were obtained and pooled for the lines and for the statistical treatment.

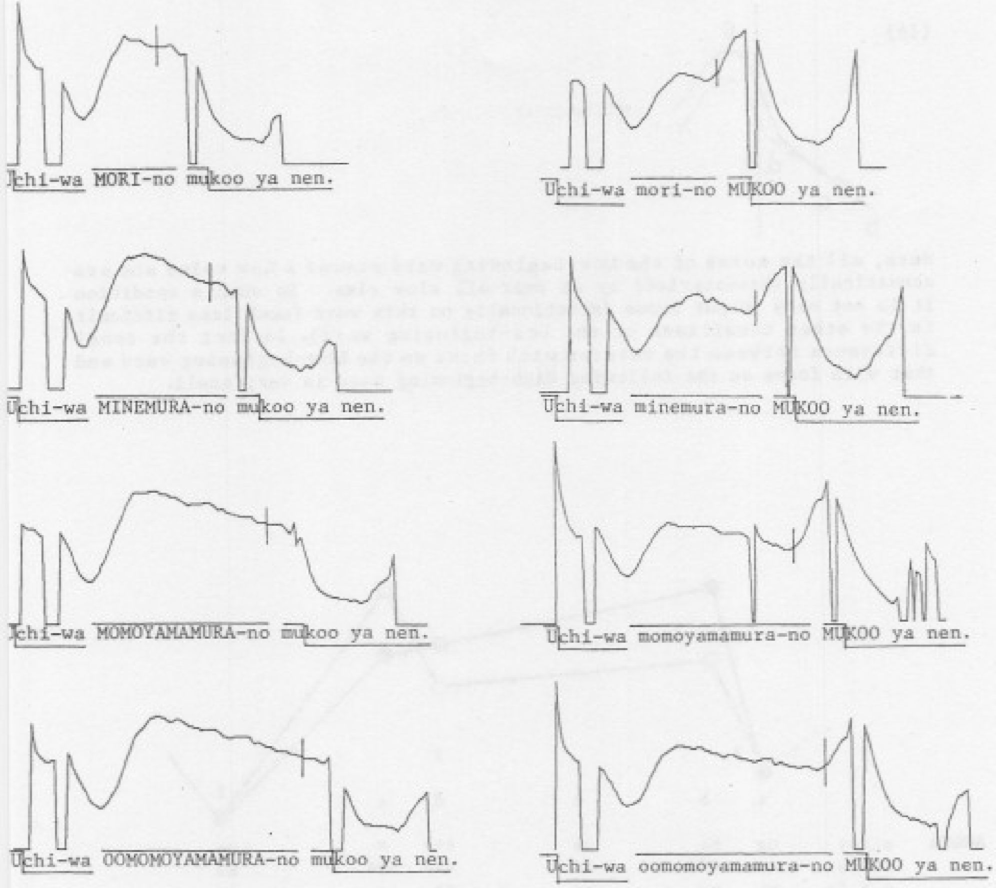


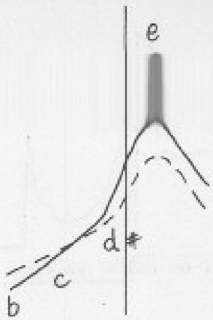
Figure 3. Representative F0 contours illustrating the interaction of focus with declination in High-beginning unaccented words of different lengths.

4.2 Low-beginning words

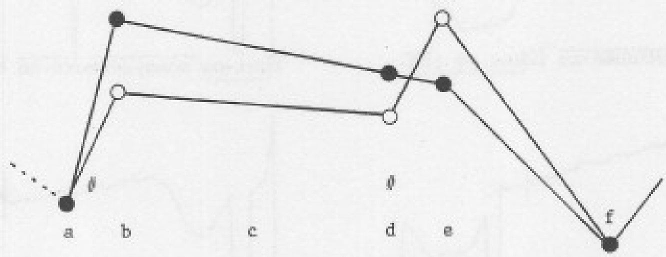
Focus in Low-beginning words is realized by increasing the steepness of the slow rise and that of the rise to the High-pitched mora if there is one. The amount of increase for the steepness of the slow rise is not always the same: it depends on the presence or absence of the nucleus. Focus may lower the beginning of the word, but again there is an interaction with whether the word has a nucleus.

When a nucleusless Low-beginning word is followed by a High-beginning word (e.g. meganeya-o # nozoiteru-wa), focus has little effect on the F0 shape of either of the two words. This can be schematized as in (14):

(14)



Here, all the moras of the Low-beginning word assume a Low value and are acoustically characterized by an over-all slow rise. In such a condition it is not easy to put focus intentionally on this word (much less difficult in the other conditions of the Low-beginning word). In fact the tonal difference between the version with focus on the Low-beginning word and that with focus on the following High-beginning word is very small.



		a	b	c	d	e	f
ANOVA	size:	ns	ns	ns	***	*	ns
	focus:	ns	***	**	***	***	ns
	s X f:	ns	ns	ns	**	ns	ns

(*** p 0.01, ** p 0.01, * p 0.05)

		a	b	c	d	e	f
ANOVA for the versions with focus on target word							
	size:	ns	ns	ns	**	ns	
LINEAR TREND:				**			
ANOVA for the versions with focus on <u>mukoo</u>							
	size:	ns	*	*	ns	ns	
LINEAR TREND				ns			

<u>Figure</u> <u>4</u>. Mean F0 values for selected measurement points in phrases containing High-beginning words of various lengths preceding another High-beginning word, with focus on the target word (filled circles) or on the following word (empty circles). Statistical tests of differences among the different lengths are summarized below the graph.

Figure 6 schematically shows the F0 shape of utterances in this condition. The sentence frame is:

(15) Are-wa # Q # mukoo-ya-nen.

where Q is a 4, 6, 8 or 10-mora Low-beginning word with no nucleus. Filled circles again represent the versions with focus on the Q word, while empty circles are for the versions with focus on mukoo. Figure 6 contains also a summary of ANOVA outputs. Three or four repetitions for each of the eight versions were pooled to obtain the mean values represented by the circles and for the statistical procedure. Some of the original F0 contours are shown in Figure 5.

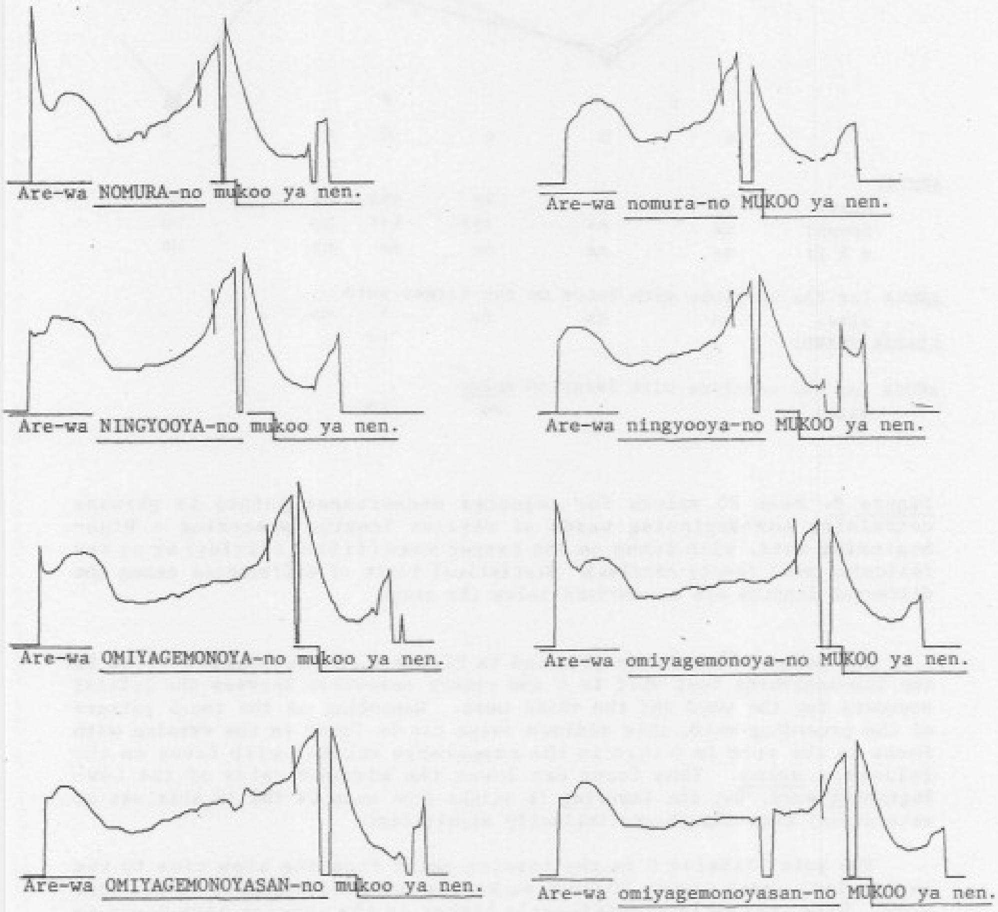
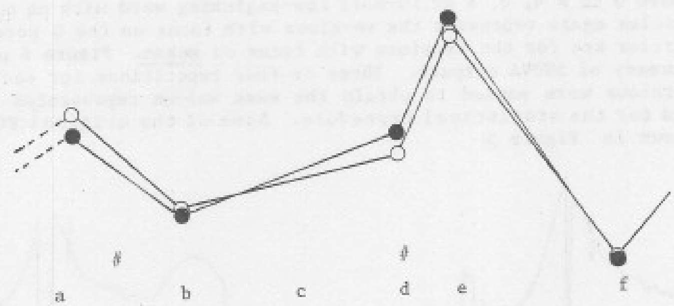


Figure 5. F0 contours illustrating the interaction of focus with the inclination of Low-beginning unaccented words of different lengths.



	a	b	c	d	e	f
ANOVA:						
size:	ns	*	**	***	ns	ns
focus:	ns	**	***	***	ns	ns
s X f:	ns	ns	ns	ns	ns	ns
ANOVA for the versions with focus on the target word						
size:	ns	ns	ns	*	ns	
LINEAR TREND:				ns		
ANOVA for the versions with focus on <u>mukoo</u>						
size:			ns	ns		

<u>Figure 6</u>. Mean F0 values for selected measurement points in phrases containing Low-beginning words of various lengths preceding a High-beginning word, with focus on the target word (filled circles) or on the following word (empty circles). Statistical tests of differences among the different lengths are summarized below the graph.

The point labeled B in (14) and in Figure 6 is the minimum value in the Low-beginning test word in Q and occurs somewhere between the initial boundary for the word and the third mora. Depending on the tonal pattern of the preceding word, this minimum value can be lower in the version with focus on the word in Q than in the comparable version with focus on the following <u>mukoo</u>. Thus focus can lower the minimum value of the Low-beginning word, but the lowering is slight (the mean is 4Hz in this set of utterances) even though statistically significant.

The point labeled D is the turning point from the slow rise to the rapid rise, which occurs 20-30 msec before the word boundary to the next word. Here, the F0 is significantly higher in the version with focus on the Low-beginning- nucleusless word (Q) than in the version with focus on the following High-beginning word, but again the difference between the two versions is very small (mean = 7 Hz).

The value labeled C is the inclination ratio from B to D. It is significantly steeper in the version with focus on the Low-beginning word, though the difference is still small, with a mean of 63 versus 45 Hz/sec when the mean is averaged over all words (i.e., over the 4-, 6-, 8-, and 10-mora words).

Summing up, the effect of focus on the F0 shape within a Low-beginning word itself in this context is to slightly emphasize the slow rise by making it steeper. It is possible that the small effect of focus on the F0 in this condition might be compensated by other acoustical cues such as intensity. These other cues, however, were not measured in this experiment.

The F0 value at point E, which is the peak value in the following High-beginning word, seems almost constant whether the focus is on the High-beginning word or on the preceding Low-beginning word. As was discussed above in section 4.1, High-beginning words as a rule realize focus as a over-all raising of the F0. But here, the presence of a preceding Low-beginning nucleusless word seems to preclude this effect. Not only the peak value at point E but also the whole F0 configuration of the word is left almost unchanged.

Thus it seems unlikely that the presence of focus in the High-beginning word in this condition (e.g. nomura-no # MUKOO) is signalled by the F0 characteristics of the High-beginning word itself. The F0 contour of the preceding word might help identify the focus location. The possible contribution of other acoustic parameters such as intensity was not examined.

My interpretation of the evident lack of an effect of focus on the F0 in either of the two words in this sequence (nomura-no # mukoo) is as follows. When focus is on the first word of this sequence, the F0 value at the end of the slow-rise (i.e., just prior to the word boundary) is higher than when the same word is in pre-focal position, because focus increases the steepness of the slow rise. After this point, the F0 will begin to rise steeply in order to realize the initial High-pitched mora of the second word. Since this steeper rise begins from a relatively high value, it could result in a very high value for the F0 peak at the first High-pitched mora of the second word. However, if this peak were very high, it would give a contextual prominence to the second word, and focus would be perceived incorrectly as occurring on the following High-beginning word rather than on the Low-beginning word. In order not to give prominence to the second word, the starting point of this steep rise (i.e. the value at the turning point D near the word boundary), should be as low as possible. For the F0 to be low at the turning point, however, the steepness of the preceding slow rise must not be much increased by the focus on the Low-beginnning word. As a result, the effect of focus on the F0 shape of the first word will be reduced. On the other hand, focus on the second High-beginning word of the sequence would be realized as a more rapid rise at the word boundary, as schematized in (11). However, the starting value of this rapid rise is relatively low because of the lack of focus in the first word, so that the peak value reached may not be any higher than when the preceding Low-beginning word is focused instead. Thus the F0 peak value at the first mora of the second word may assume a similar value whether the focus is on the first or on the second word. In most of my data in the present experiment the peak F0 values of these two versions are similar, but in emphatic speech, the peak of the second word would be much higher under focus than in the version with first-word focus.

Low-beginning words followed by a High-beginning word have another singularity. In realizing focus, they interact with the tonal structure of the preceding word. Let us compare the two sentences below:

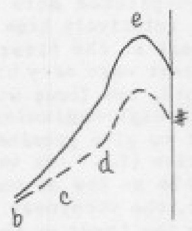
(16) a. Omohara # meganeya-o#nozoiteru-wa.

b. Minamida # meganeya-o # nozoiteru-wa.

The only difference here is the tonal pattern of the first word. Yet in the (a) sentence, it is more difficult to put focus on the second word than in the (b) sentence. The F0 contours for these two sentences are shown in E-2 and D-2 of Figure 1. The difficulty of putting focus on the second word in (a) probably comes from the difficulty of lowering the F0 any further than it already is at the minimum value just after the boundary between the first and the second word. The speaker attempts to lower the pitch here to increase the steepness of the slow rise and to give full focus to the second word. In the (a) sentence, however, the F0 has been declining smoothly from just after the nuclear mora of the first word, so that, at the word boundary the F0 arrives at a very low value. This value is about 70 Hz when both words are out of focus (i.e., when focus is on the third word of the sentence), which is already near the lowest value the speaker can produce (about 65 Hz). If the speaker cannot further lower the F0 at this point (because he has already reached the floor of his pitch range), it would become difficult to realize focus only by manipulating the rate of the following slow rise, because the end point of this slow rise portion in this condition has little flexibility for the reason we have seen in the previous paragraph. Thus the F0 range that the speaker can use to give the maximum focal effect to the second word is very small.

When the nucleusless Low-beginning word is followed by another Low-beginning word, we see the patterns in (17):

(17)



In this condition all moras but the last one assume Low value (e.g. meganeya-o # matteru-wa). Figure 8 summarizes mean values for various aspects of the F0 contours of sentences of the form:

(18) Are-wa # Q # mae-ya-nen.
'That is in front of Q.'

where mae 'front' is a Low-beginning nucleus-bearing word, and the preceding word in Q is a 4-, 6-, 8- or 10-mora Low-beginning nucleusless

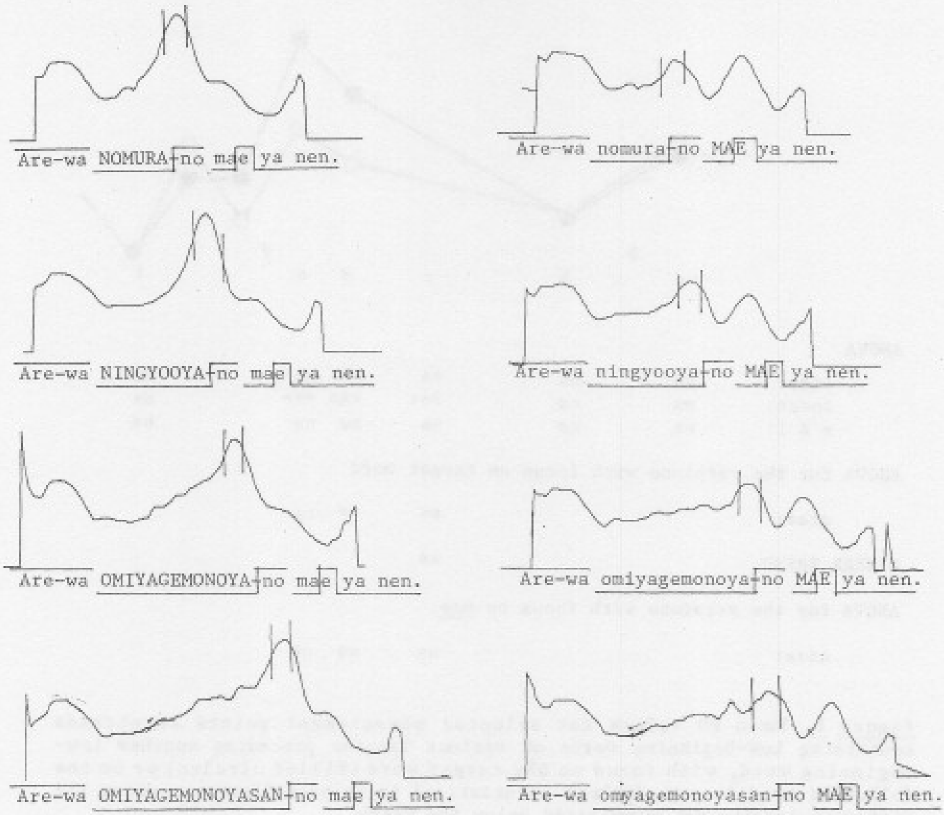
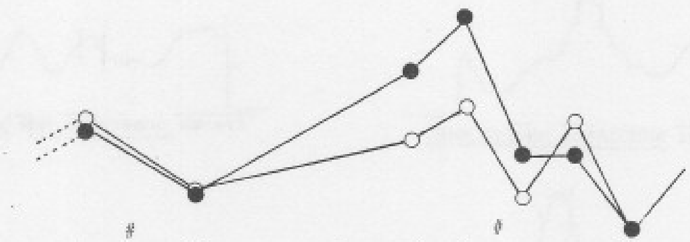


Figure 7. F0 contours showing the inclination in Low-beginning unaccented words of different lengths before a Low-beginning word.

word. Focus was put either on the word in Q or on mae. Two repetitions for each version were analyzed. See Figure 7 for some of the F0 contours being summarized.

Focus emphasizes the slow rise more prominently than in the previous sentence set in (15), where all the moras of the word were Low-pitched. The values labelled C and D in Figure 8, which are the slope and end-point of the slow rise, are clearly higher in the versions with focus on the Q word. The minimum F0 value in the Low-beginning word (point B) does not differ significantly under different placments of focus. However, in some other example sentences not shown here, focus clearly lowered the F0 at this point.



ANOVA

	a	b	c	d	e	f
size:	ns	ns	**	**	ns	ns
focus:	ns	ns	***	***	***	ns
s X f:	ns	ns	ns	ns	ns	ns

ANOVA for the versions with focus on target word

	c	d	e
size:	**	**	ns

LINEAR TREND: **

ANOVA for the versions with focus on mae

	c	d	e
size:	ns	ns	ns

Figure 8. Mean F0 values for selected measurement points in phrases containing Low-beginning words of various lengths preceding another Low-beginning word, with focus on the target word (filled circles) or on the following word (empty circles). Statistical tests of differences among the different lengths are summarized below the graph.

The point labelled D, which is the F0 value at the turning point from the slow rise to the rapid rise (about 20 msec before the beginning of the last High-pitched mora), is very clearly raised by focus. The peak value in the final High-pitched mora (point E) is also clearly raised in versions in which the Low-beginnning word is under focus.

In contrast with the sequences in (15), however, there is no significant difference for the value at the beginning of the low-rise (point B). The reason why focus lowers this value more consistently at the beginning of the rise before a High-beginning word would be as follows. When the Low-beginning word is followed by another Low-beginning word, focus can be easily realized by the rise to the word-final High-pitched mora. When the Low-beginning word is followed by a High-beginning word, on the other hand, there is no High-pitched mora in the first word that might be raised by the focus, and the slow-rise portion in this sequence has little flexibility to realize focus, because emphasizing the slow-rise may give contextual prominence to the next word. Thus the focus would be obliged to resort to manipulating the initial portion of the first focused word in order to best realize the focus effect.

When the Low-beginning word has a nucleus, focus seems to be realized in a similar way as when the nucleus is not present and followed by a Low-beginning word, as just discussed. However, this generalization was based on a small number of samples, all of which had compound nouns for the Low-beginning nucleus-bearing word, and hence may reflect peculiarities of compound nouns instead of the general behavior of this type of word. See section 5 for a more detailed discussion of the behavior of compound nouns.

5. The basic unit of phrasing in OJ

5.1 The phrase in TJ

It is generally recognized that in Tokyo Japanese, words may lose their accentual independence and fuse into one tonal unit in a larger prosodic context. This unit is called 'accentual phrase' (c.f. Hirayama 1960) or simply 'phrase'. For example, McCawley (1968) writes:

> I described the accentuation of a phrase by saying that a phrase contains at most one accented syllable and that everything up to the first mora of that syllable is high pitched and everything after the first mora of that syllable is on a low pitch, except that the first mora of a phrase will be low pitched unless it itself is accented. It will be necessary for a grammar of Japanese to give rules which mark the places where phrases begin and end, since those boundary points must be known in order to apply the above rule of pitch assignment. Furthermore, it will be necessary to revise the pitch assignment rule somewhat since there are actually two kinds of phrases, "major phrases" and "minor phrases"; each minor phrase follows the rule given for pitch shape except that only the first minor phrase of a major phrase will contain a high pitch. (p.137)

According to such a model there would be two types of phrases with respect to the tonal configuration:

(19) a. a phrase consisting of i nucleusless words, j nucleus-bearing words, and k nucleusless words ($i >= 1$, $j = 0,1$, $k>=1$ if $j=1$, else $k=0$). All the moras before the nucleus except for the first one assume High pitch. The first mora may be high if it belongs to a long syllable.

ooo # ooo # .... (# ooo # ooo # ....)

i j k

b. a phrase consisting of one nucleus-bearing word + i following nucleusless words ($i>=0$). All the moras following the nuclear mora are Low-pitched.

ooo (# ooo # ....) or ooo (# ooo # ....)

These two types of phrase have some tonal characteristics in common. Aside from the fact that they have at most one nucleus, a phrase contains only one High-pitched mora sequence. Furthermore, the heights of the first and the second mora are different unless they form a long syllable. Acoustically, both types of phrase begin with a F0 rise (Poser 1984, Pierrehumbert and Beckman in press).

McCawley seems to have taken into account only the cases where the sentences are uttered neutrally, without putting focus on any particular syntactical components. But, as pointed out by Pierrehumbert and Beckman, focus restructures phrases. A sequence of two nucleusless words becomes one phrase when uttered as a neutral utterance or when focus is on the first word (e.g. ooo # ooo), but when focus is put on the second word (both marked and unmarked cases) it splits into two independent phrases (e.g. ooo # ooo). I think this kind of phrase restructuring occurs also at the theme/rheme boundary.

This concept of phrase, however, is not directly applicable to the other Japanese varieties. First of all, the initial tone differentiation rule does not apply to all Japanese varieties. Second, a sequence of words that contains at most one nucleus may have more than one sequence of High-pitched moras even in non-emphatic speech, as it is the case in OJ[6]. In order for the "phrase" to be a generally applicable concept, I will redefine it as a sequence of words (including enclitics) that contains at most one High-pitched portion not blocked by focus.

5.2 Applicability to OJ

Let us now examine if this concept of phrase is applicable to OJ. If such a phrase is applied to OJ, there would be the following two types of phrase, with every phrase having at most one High-pitched mora sequence:

(20) a. The <u>Low-beginning</u> <u>phrase</u>, which consists of one Low-beginning nucleusless word + i High-beginning nucleusless words (i>=0) + j High-beginning-nucleus-bearing words (j=0,1):

ooo (# ooo # ...) (# ooo)

or of one Low-beginning-nucleus-bearing word:

ooo

b. The <u>High-beginning</u> <u>phrase</u>, which consists of i High-beginning-nucleusless words + j High-beginning nucleus-bearing words (i>=0, j=0,1; i or j >0):

(ooo # ...) (# ooo)

Focus restructures phrases only when it is on the second word or later of a word sequence which can form a High-beginning phrase.

----

[6] For example, the concatenation of a High-beginning nucleusless word followed by a Low-beginning nucleusless word and then a High-beginning nucleusless word has no nucleus, but it has two High-pitched sequences.

Do these two types of phrase have some acoustic characteristics in common like a initial rise in TJ? It seems so: both types begin with a rise. But there is evidence that makes us hesitant to analyze the OJ tonal system in terms of this notion of phrase. Let us examine two cases in (21), where % represents a phrase boundary:

(21) a. % oo..oo #% oooo .. %

b. % oo...o # oooo .. %

In both cases, the first word belongs to a Low-beginning phrase, and the phonological tone sequence over the first few moras is the same. So we can expect the acoustic realization of the Low tones in this sequence (i.e. the slow rising portion of the word) to be the same in both cases. But as we have already seen in Section 4, the slope of the slow F0 rise up through the last Low-pitched is much steeper in (21a) than in (21b) when focus is on the first word. (In sentences that fit the specification of (21a), the rise averages 77, 62, 59 and 54 Hz/sec for 4-, 6-, 8- and 10-mora sequences; in sentences that fit (18b), the rise is 159, 109, 99 and 71 Hz/sec for 3-, 5-, 7- and 0-mora sequences.) The reason why these two types of sequence show different F0 behavior was discussed above. It can be ascribed to the presence versus absence of a word boundary at the end of the slow rise: focus cannot drastically raise the F0 at the word boundary in this condition. It is clear that the word boundary influences the F0 shape of the phrase. Thus the words constituting a phrase hold their integrity and do not fuse into one. This observation has led me to the conclusion that the basic unit of accentuation in OJ is not a previously defined phrase but a word.

Yet words can really fuse into one in a sequence of one or more High-beginning nucleusless words ending in at most one High-beginning nucleus-bearing words (i.e. ooo # ooo ... (# ooo)) with focus on the first word in the sequence.

Another exceptional case is found in some compound words. For example, sanzyuu+hati 'thirty-eight' has two independent High-pitched moras and it is best treated as a word containing two accentual units. In compound words such as naniwa+miyage 'souvenir of Osaka' (the second word in panels G-1 and H-1 in Figure 1) and omiai+bangumi '(TV) dating-game program', I encountered cases in which these words have two F0 rise-falls when they are in post-focal position. This phenomenon could be easily interpreted if these words consist of two accentual units. Furthermore, the existence of such compound words as niwaka+bugensya 'a nouveau riche' encourages this interpretation: a sequence of more than one High-pitched mora in Low-beginning words occurs only in compound words (c.f. Wada 1942).

Thus the unit of accentuation in OJ is different from the phrase previously defined, necessitating a modification of the traditional concept of phrase originally defined for TJ, such as McCawley's minor phrase. However one can still call the OJ unit of accentuation a phrase in the sense that it is not exactly the same as the word: it may be larger or smaller than the word. In the following discussion, I will call this accentual unit the 'OJ phrase' in order to avoid confusion with the traditional concept of phrase in TJ.

6. A target-tone model for OJ?

6.1 The target-tone model for TJ

In their analysis of the tone structure of TJ, Pierrehumbert and Beckman (in press) propose a new model to account for the F0 contours of utterances. They claim that the traditional account in terms of moraic pitch height does not explain the phonetic event. Instead they posit four target tones that characterize the phrase: a boundary tone (L%) which occurs at the phrase boundary, a phrasal High tone which occurs at its beginning, an accentual High tone which occurs at the accent nucleus, and an accentual Low tone which occurs shortly after the accentual High tone. The last two tones occur only in an accented phrase (i.e., a nucleus-bearing phrase in my terminology). Actual values that these tones assume in an utterance are determined mainly by focus and 'catathesis'. Two kinds of L% are distinguished for long and short phrase-initial syllables. Linear interpolation between these tones gives a good approximation of the F0 contour of a phrase in utterance.

The main reason for Pierrehumbert and Beckman's rejection of the moraic pitch-height model is the absence of High-spreading and Low-spreading. In a nucleusless phrase (oooo...), the F0 values of the High-pitched moras are expected to be same. But the actual F0 contour shows a slow falling. This slow fall has been generally interpreted as a non-phonological or physiological downdrift (Kawakami 1962, Kobayashi 1963). Pierrehumbert and Beckman, however, found that the rate of this downtrend decreases as the number of the High-pitched moras increases. It seems as if the boundary to the next phrase would have a target value to which the F0 is interpolating. They also claim that the declination rate of a Low-pitched mora sequence decreases as a function of the number of Low-pitched moras.

6.2 Applicability to OJ

I examined the F0 contours of sequences of High-pitched moras and sequences of Low-pitched moras with three sets of sentences, with a view to see whether similar evidence against tone spreading exists in OJ.

The sentence set that was used to examine the F0 in the sequence of High-pitched moras consists of four sentences differing in the length of the second word, which had 4, 6, 8 or 10 moras. All the moras of the second word, which is the test word, are High-pitched. The sentences were uttered in two ways, one with focus on the test word and another with focus on the following word. Figure 3 above gave the F0 contours of selected utterances from this set of sentences. Figure 4 gave mean values for various measurements, as well as the results of statistical analyses involving these measures. The value C in Figure 4 is the inclination rate (Hz/sec) from the point B to D, as defined by equation (22)[7]. Figure 9 shows the detailed relationship between B and D in all utterances.

(22) C = (F0 value at B - F0 value at D) / interval from B to D

----

[7]Pierrehumbert and Beckman (in press) define the rate of declination as the coefficient of a linear regression between the endpoint tones.

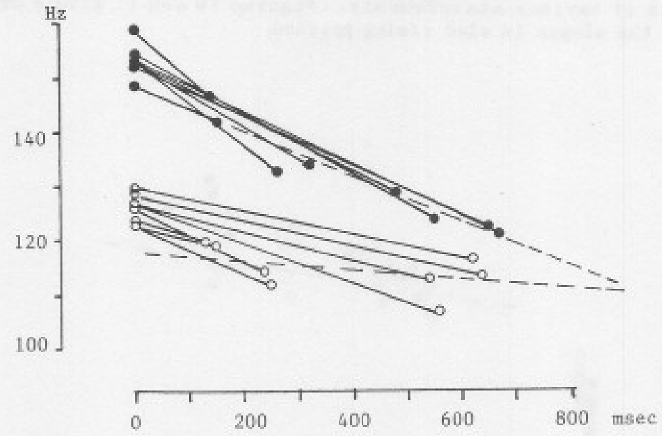


Figure 9. Beginning and end point values (in Hz) of declining F0 portion in High-beginning unaccented words plotted against the distance (in ms) between the plotted points for all utterances in the corpus in Figure 4.

Let us examine only the case in which the test word is in pre-focal position (empty circles in Figures 4 and 9). In this case, the point D in Figure 4 roughly corresponds to the phrase boundary and the measure C is the declination over the High-pitched mora sequence.

A trend analysis (Winer 1971) shows that the value C does not vary systematically as the size of the test word increases ($p>.25$). This seems to be counter-evidence to the target-tone hypothesis. But an ANOVA indicates that the F0 value at D does not vary significantly with the variation in word length ($p>.05$). This seems in favor of the tone-target hypothesis. This problematic set of results is probably caused by the behavior around point B, which increases systematically as the word size increases ($p<.05$, trend analysis; see also Figure 9). Points B and D are both places where a tone must be posited in the framework of the tone-target theory. (I have labeled these a High phrasal tone and High boundary tone, respectively: see Appendix 2).

The F0 shape of a Low-pitched mora sequence was examined with the sentence sets described by examples (15) and (18) above. The first set contained four sentences whose second word, the test word, is a Low-beginning nucleusless word containing 4, 6, 8 or 10 moras, which in this context would all be Low toned in a traditional moraic tone account. Two versions, one with focus on the test word and another on the following word, were uttered. Three or four utterances for each version were obtained and used in the analysis. The same test words were used in the sentence set exemplified in (18). This set differs from (15) only in that its third word is Low-beginning, which makes the final mora of the preceding test word High-pitched. See Figures 5 and 7 for sample F0 contours for some of the utterances, and Figures 6 and 8 for a summary of mean values of various measurements. Figures 10 and 11 give a detailed summary of the slopes in slow rising portion.

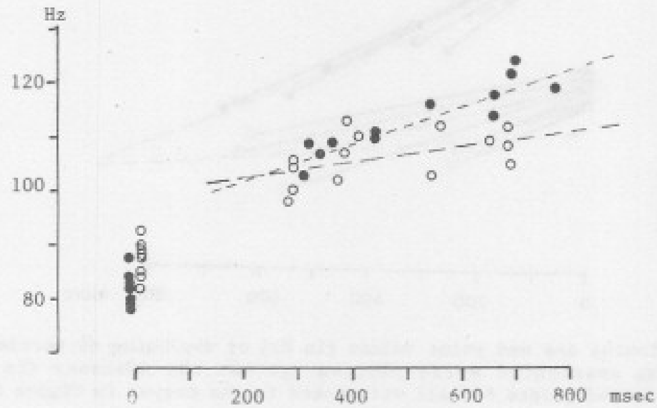


Figure 10. Beginning and end point values (in Hz) of slow rise portion in Low-beginning unaccented words plotted against the distance (in ms) between the plotted points for all utterances in the corpus in Figure 6.

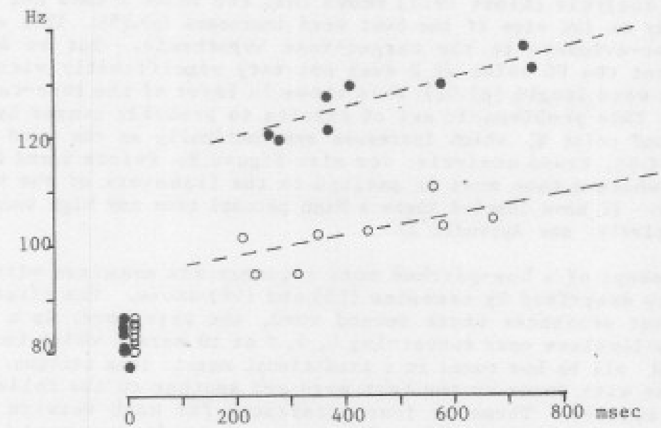


Figure 11. Beginning and end point values (in Hz) of slow rise portion in Low-beginning unaccented words plotted against the distance (in ms) between the plotted points for all utterances in the corpus in Figure 8.

In these sets of utterances the point D roughly corresponds to the turning point from the slow rise to the rapid rise, where in terms of target hypothesis a target tone should be posited (I call this a Low phrasal tone). The value C is the rate of the slow rise from B to D.

In sentences like (18), when the test words are in focus, the rate of the slow rise systematically decreases as the word length increases, a trend that is significant according to the trend analysis. This would support a target-tone model. But at the same time, the F0 value at D increases significantly, as analyzed using trend analysis. Moreover, when the test words are out of focus, there is no statistically significant effect of word size, as shown by an ANOVA. These last two results seem to be against the target hypothesis.

In sentences like (15), in the versions with focus on the test word, the rate of the slow rise seems to decrease as the word length increases, and the F0 value at point D also seems to increase, though the linear trend is significant in neither case. When the test words of the same sentence set are out of focus, the effect of word size is not significant, as shown by an ANOVA.

In sum, these two data sets do not provide a definite answer to the target hypothesis. With one other data set which I have not reported here, the posited targets seemed to have fixed values, though no statistical analysis can be applied because the number of observations was too small. Thus, whether a target-tone type of account can apply to OJ is still an open question.

Acknowledgements

I am deeply grateful to Prof. I. Lehiste for having made my stay at OSU possible and for her constant support. I would like to thank Dr. M. Beckman for having enlightened me with a new perspective of study of intonation, and for discussion.

References

Dik, S., et al. [sic]. "On the Typology of Focus Phenomena," In T. Hoekstra, H. van der Hulst, and M. Moortgat (eds.) Perspectives on functional grammar, Foris, 1981.

Hirayama, T. Zenkoku Akusento Ziten, Tookyoodoo. 1960.

Homma, Y. "Pitch accent in Kyoto Speech through a Visicorder", Bull. Phon. Soc. Japan, 137, 11-15, 1971. (in Japanese)

Homma, Y. "Tonal Assimilation in the Kyoto Dialect," Bull. Phon. Soc. Japan, 158, 12-15, 1978.

Ikeda, K. "Kinki akusento keisikikan no mondai," in Nihongo no Akusento, Nihon Hogen Gakkai, 1942.

Kawakami, S. "An Analytic Notation System for the Pitch Accent in the Kyoto-Osaka Dialect," Bull. Phon. Soc. Japan, 109, 10-12, 1962. (in Japanese)

Kawakami, S. "The System of the Accent Patterns in the Dialect of the Kyoto-Osaka District," Study of Sounds, 10, 249-258, 1962. (in Japanese)

Kawakami,S. "Japanese Accent Observed by the Aid of Pitchgraph," Study of Sounds, 10, 115-129, 1962. (in Japanese)

Kindaichi, H. "Keihan akusento no atarasii mikata," Kinki Hoogen, 3, 1950.

Kobayashi, R. "On the Relation between Accent Forms and Speech Melodic Curves," Bull. Phon. Soc. Japan, 112, 5-8, 1963. (in Japanese)

Kori, S. "F0 contour of Interrogative Utterances in Osaka Japanese," Proc. First Meeting, Kinki Soc. Phon., 1984a. (in Japanese)

Kori, S. "Individual Differences in the Perception of Prosodic Features," Proc. Second Meeting, Kinki Soc. Phon., 1984b. (in Japanese)

McCawley, J. D. The Phonological Component of a Grammar of Japanese, Mouton, 1968.

Pierrehumbert, J. and Beckman, M. Japanese Tone Structure, to appear in Linguistic Inquiry Monograph Series. In press.

Sakuma, K. "Kyootogo ni okeru akusento," Study of Sounds, 4, 61-77, 1931.

Sugito, M. "Oosaka no akusento," in Daitosi no Gengoseikatu, Sanseido, 1981.

Sugito, M. Nihongo Akusento no Kenkyuu, Sanseido, 1982.

Sugito, M. and Hirose, H. "An Elecctromyographic Study of the Kinki Accent," Ann. Bull., Research Institute of Logopedics and Phoniatrics, University of Tokyo, vol.12, 1978.

Wada, M. "Kinki-akusento ni okeru meesi no hukugoo-keetai," Bull. Phon. Soc. Japan, 71, 10-13, 1942.

Winer, B. J. (1971) Statistical Principles in Experimental Design, McGraw-Hill, 1971 (2nd ed).

Yamada, T., Okajima, H., & Miura, I. (1982). "On combinations of word accents in Tokyo, Nagoya and Osaka dialects", Bull. Phon. Soc. Japan 169. 10-14. (In Japanese)

Appendix 1: The sociolinguistics status of Osaka Japanese

It should be pointed out that for most Japanese people the term 'Japanese' does not mean Standard Japanese (usually defined as the Japanese spoken in Tokyo). For them Japanese consists of all the various but similar linguistic systems used in Japan. Moreover, as far as accentuation is concerned, Tokyo Japanese is not an exact synonym for the Standard Japanese, even though it is considered to be the model language. In fact, many people believe themselves to be speaking Standard Japanese even though their accent system does not coincide exactly with the Tokyo accentuation. Japanese pronunciation dictionaries are usually nothing more than accent dictionaries for Tokyo Japanese. But it is not sufficient to know only the accentuation of lexical words. These pronunciation dictionaries usually list many pages of rules to be applied when the words are used in combination with enclitics or in compound form. For the people from the other areas and for learners of Japanese as a second language, it would be very difficult to learn all those ad-hoc rules in addition to the accentuation of each word. The standard language must be simple or at least more flexible in regards to the accentuation.

The study of word accent is the most advanced area in Japanese linguistics. The main concerns of researchers in the field have been the phonological description of the accent systems of various dialects based on auditory observation and the discovery of the historical relationships among these systems. Since information about word accent is believed to be sufficient to reconstruct the history of an accent system, little attention has been paid to sentence prosody. Exceptions are the auditory-based works of Y. Fujiwara and his students and of Y. Yamaguchi, who call the utterance pitch patterns the 'sentence accent'.

Osaka Japanese and the similar varieties of Japanese spoken in the vicinity of Osaka (Kyoto, Kobe, Nara, etc.) are grouped together as the 'kansai' dialect by the Japanese from the other areas. There are somewhere between 10 and 20 million native speakers of these varieties, and the sociolinguistic status of these varieties is unique. The native speakers believe theirs to be the only non-standard varieties of Japanese which have social prestige. Most teachers in local schools (from elementary to the highest education) give their lessons in OJ or in Standard Japanese with an Osaka tonal pattern. This is a surprising situation to the Japanese from the other areas, where the local variety has little prestige relative to Standard Japanese. Most OJ speakers use OJ in every circumstance, even in Tokyo or on TV, while other non-Standard speakers are willing to assimilate their speech to Tokyo Japanese. Some OJ speakers seem not even to notice the difference between the accent system of OJ and that of TJ, while they are sometimes very sensitive to the difference among other dialects.

Meyer (1906), Polivanov(1925), and Sakuma (1931) were the first to use acoustic data to study Kyoto Japanese, which is very similar to that of Osaka. Sakuma discussed the F0 characteristics of short words in sentences. But as the focus of interest of researchers in accent shifted to the phonological treatment of dialectal accent systems and the establishment of their historical relationship, experimental study came to be neglected.

In the 1960's, after a lapse of some decades, Sugito began to publish a series of important articles on Osaka accent using various experimental techniques (cf. Sugito 1982). Although the majority of her work in this area deals with word accents in isolation, there are also several remarks on their realization in larger prosodic contexts. For word accents in Kyoto Japanese there are studies performed by Homma (1971, 1978). In 1982, Yamada et al. discussed in a brief report how word accents interact in two-word utterances in TJ, OJ, and Nagoya Japanese and provided evidence for the generally accepted view that the sentence tonal pattern in OJ is a mere concatenation of word accents. Kori (1984a, 1984b) discussed the interaction of sentence intonation with word accents in two-mora word utterances. Yoshizawa and his co-workers are now doing research on the tonal characteristics of the sentence final particles of OJ as a part of their comparative study of several varieties of Japanese. As for phonological treatments, Wada, Hattori, Kindaichi and Kawakami are major contributors.

Appendix 2: A target-tone model for Osaka Japanese

In this appendix, I will propose a target-tone model of OJ, though the applicability of such a model is not yet proved.

In Section 5, I discussed the basic unit of phrasing in OJ, and argued for a particular definition of the basic unit, which I called the OJ-phrase. I will use the OJ-phrase as the unit of tone assignment in the tone-target model. Here I will show only a simple sketch of what tones this model might need to include, without going much into detail.

Six tones are necessary to account for lexical contrasts between Low-beginning and High-beginning and between nuclear and nucleus-less words and to model the various inflexion points discussed above. I posit the following inventory of tone types for these six tones:

Boundary tones:	High boundary tone	(H%)
	Low boundary tone	(L%)
Phrasal tones:	High phrasal tone	(pH)
	Low phrasal tone	(pL)
Nuclear tones:	Nuclear high tone	(nH)
	Nuclear low tone	(nL)

The two types of phrasal tones, pH and pL, cannot occur together in the same phrase, because they characterize the High-beginning phrase and the Low-beginning phrase respectively. These six tones have the following intrinsic tonal relationships:

$$H\% < pH > nH > nL > L\% \quad \text{and} \quad L\% < pL < nH$$

(Here A > B means that A is intrinsically higher than B)

There would be four types of OJ-phrase, and they would have the following tone target configurations.

High-beginning phrase:

nucleusless HZ pH HZ

nucleus-bearing HZ pH nH nL LZ

Low-beginning phrase:

nucleusless LZ pL HZ

nucleus-bearing LZ pL nH nL LZ

Examples:

ooo # ooo LZ pL HZ LZ pL nH nL LZ

ooo # ooo LZ pL HZ pH nH nL LZ

OSU #36: Papers from the Linguistics Laboratory 1985-1987, Sept. 1987.

Perceiving by syllables or by segments: Evidence from the perception of subcategorical mismatches

Keith Johnson
Linguistics, Ohio State University

Abstract: This paper describes an experiment in which two general hypotheses concerning speech perception are tested. According to the segment perception hypothesis the acoustic signal is interpreted in terms of segments analogous to those used by phoneticians in transcribing speech. The syllable perception hypothesis on the other hand holds that the speech signal is perceived in terms of syllable sized units. The experiment tests these two hypotheses by presenting subjects with a perceptual task for which the two make opposite predictions. Tokens with subcategorical mismatches were produced by cutting the fricatives [s] and [ʃ] from VC syllables (vowels were [i,a,o,u]) and recombining them with vowels which differed from the original context in terms of transitions and rounding. The segment perception hypothesis predicts that in syllables with transition mismatches (ie. transitions for [s] and with [ʃ] actually occurring) coarticulatory rounding on the actually occurring fricative will aid in the perception of [ʃ] and slow the perception of [s], while the lack of rounding on the actually occurring fricative will have the opposite effect. This is because the rounding makes [ʃ] a more extreme example of [ʃ] (and thus easier to categorize as such) while rounding makes an [s] less distinctly an [s]. The syllable perception hypothesis predicts that in syllables with transition mismatches coarticulatory rounding on the actually occurring fricative will aid the perception of [s] and hinder [ʃ] perception. This is because the [s] with rounding is acoustically closer to the prediction made on the basis of the transition on the vowel. Similarly, the [ʃ] with rounding is acoustically further removed from the [s] which is expected as a result of the transitions on the vowel in a mismatched syllable and thus should require more time to be perceived as [ʃ]. The results of the experiment reported here support the segment perception hypothesis. Subjects' perception of [s] in syllables with transition mismatches was inhibited by coarticulatory rounding while their perception of [ʃ] in syllables with transition mismatches was facilitated by coarticulatory rounding.

1. Introduction

The experiment described in this paper was designed to test two hypotheses about speech perception. These two hypotheses will be called syllable perception and segment perception. As their names indicate they differ in so far as they entail that the basic units of speech perception are respectively syllables and segments. Advocates of the first approach include Klatt (1980)[1], Massaro (1972), and Morton and Broadbent (1967). The theorists who suppose that segments are perceived include Bondarko et al. (1970), Fant (1967), Stevens and Halle (1967), Liberman et al. (1967), and Pisoni and Sawusch (1975). As these lists indicate there are a wide variety of ways that syllable or segment perception might be conceived.

The unifying feature of the different perception by segments approaches is that they all hold that the objects of speech perception are phonemes and that subsequent percepts (syllables or words) are sequences of phonemes. Figure 1 shows the sequence of perceptual events as envisioned in a perception by segments approach.

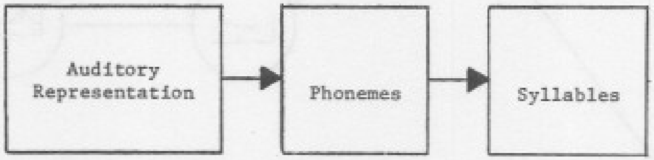


Figure 1: Organization of the speech perception process from a perception by segments approach.

In this kind of model the recognition device takes as input a preliminary auditory analysis and computes as output phones. Pisoni and Sawusch (1975) proposed such a model of perception. In their view speech perception is accomplished via (1) acoustic feature analysis, (2) phonetic feature analysis, (3) a feature buffer, and (4) phonetic feature combination. The key element in the segment perception approach is that perception is accomplished in terms of units which correspond to the symbols a phonetician might use to transcribe the utterance. Thus, in the segmental model the things being perceived are segments which are then combined into syllables.

An example of the syllable perception approach is found in Klatt (1980) and is illustrated in figure 2. This figure shows a phonetic decoding network. The network defines possible sequences of spectra. When the perceptual system matches a particular sequence (i.e. a particular path through the network is followed) the syllable defined by the sequence is perceived. There is no intermediate perceptual stage between auditory analysis and identification of a syllable. In this model the identification of component phonemes can only be accomplished after the entire syllable is identified.

[1] Klatt's (1980) model actually includes both a version of syllable perception and segment perception. Since he argues in the body of his paper for syllable perception I am including him in that camp.

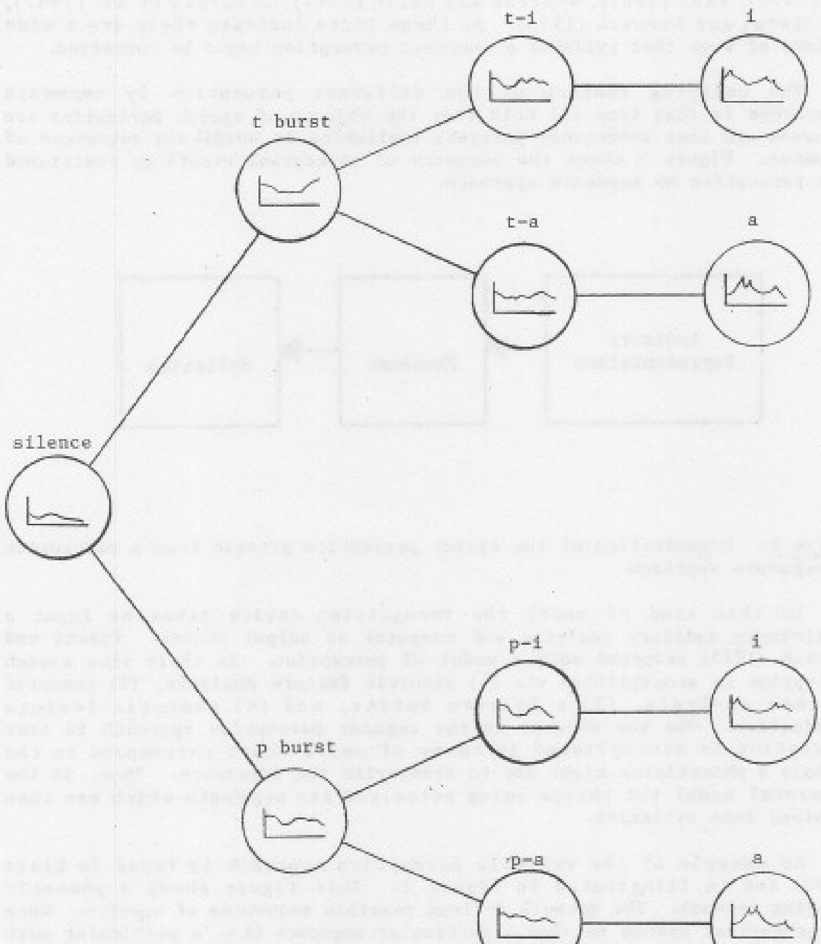


Figure 2: A spectral decoding network for the syllables [ti], [ta], [pi], and [pa].

The experiment reported here used naturally produced speech tokens which were edited to produce subcategorical mismatches, following the technique used by Whalen (1984). The subcategorical mismatches used in this experiment were produced by cutting VC syllables and recombining the resulting segments with V's or C's from other syllables. More specifically, the fricatives [s] and [ʃ] were produced after four different vowels [i,a,o,u]. The mismatches which resulted from recombining these segments were of two types. Transition mismatches resulted when vowels from the context [__s] were recombined with [ʃ] or when vowels from the context [__] were paired with [s]. Rounding mismatches resulted when fricatives from the context [V_{+rnd}__] were recombined with [-rnd] vowels, or when fricatives from the context [V_{-rnd}__] were recombined with [+rnd] vowels.

Mismatches pose an interesting problem for speech perception theories because coarticulatory information remains in the segments which are separated from each other. In the case of transition mismatches the discrepancy between the place of articulation information in the vowel and that in the fricative itself produces a fairly large and stable reaction time lag in perception (Whalen 1984). It is also the case, though, that the discrepancy between the rounding of the vowel and the effects of rounding coarticulation in the fricative could affect perception.

When subjects are asked to identify the fricative noise in mismatched tokens such as these it is often the case that they respond before the completion of the acoustic signal. Due to the fact that the subject's reaction time involves both perception time and response time it is very likely that the subject has established some predictions concerning the identity of the fricative during the vowel - based on the transitions and the roundness of the vowel. Predictions such as these seem also to be the most plausible explanation of the effect of subcategorical mismatches on reaction time in identifying the fricative in VC syllables. The vowel portion of the syllable allows the listener to set up some expectations concerning the following fricative. When the expectations are not met identification is slowed.

The two hypotheses characterize the listener's predictions in quite different ways. In the syllable perception model it must be assumed that the hearer makes predictions which are below the level of segmental identities. The predictions are acoustical in nature because within a syllable the perceptual system is seen as progressing from one spectral template to the next. In the segmental model predictions are made in terms of categories instead of acoustic values. For instance, the occurrence of [u] leads to a prediction that the following consonant will have rounding.

Figure 3 illustrates the syllable model for the experimental tokens. This figure is analogous to figure 2. If we suppose that the vowel [u] from [uʃ] is presented, then the prediction for the following state in the network is the spectral template for [ʃ]. If the speech token being presented is a mismatched token which has [s] instead of [ʃ] then the fact that the perceptual prediction was a spectrum suggests a strategy for recovering from the mismatch. The general requirement is that another spectral template be found which will match the actual fricative spectrum.

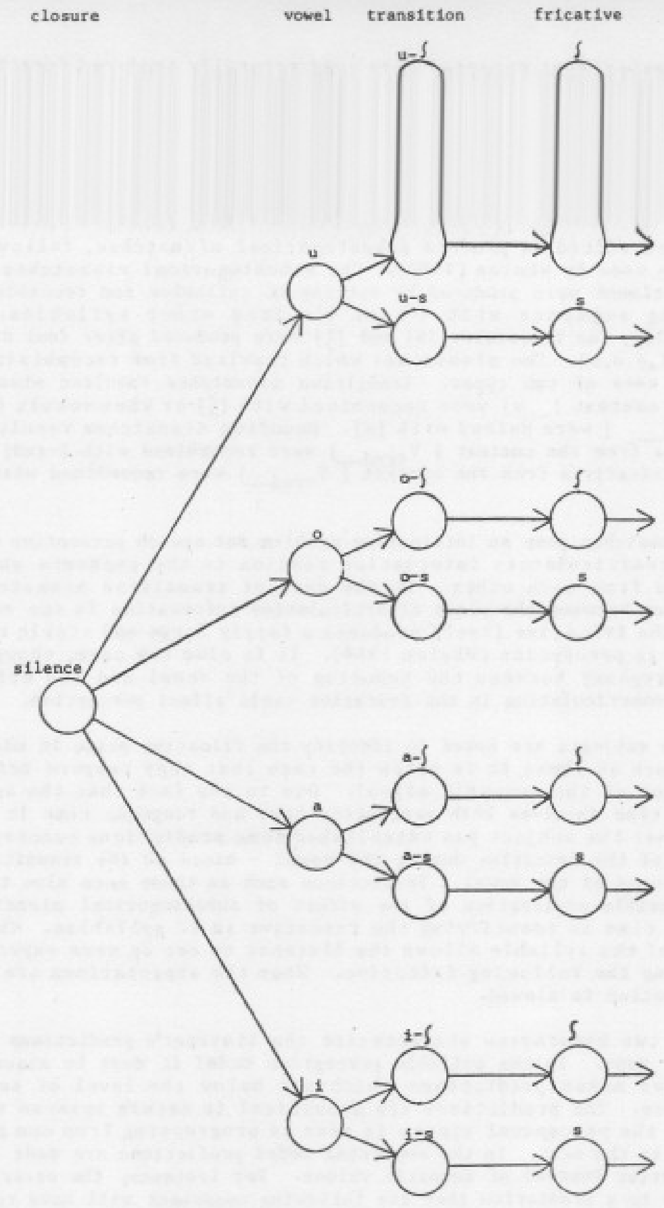


Figure 3: Perception of the VC test tokens in a transition network.

One way for the syllable perception model to recover from a mismatch is for the perceptual mechanism to attempt to interpret the auditory presentation as a well-formed syllable by revising the template or the auditory representation of the sound or both. Thus, if the expected fricative is [ʃ], rounded [s]'s will be more easily perceived because their spectrums are closer to the expected spectrum (i.e. they involve less revision of templates and/or representations). This method of recovery would predict that if the hearer sets up an expectation for an [s] spectrum, then it will be easier to recover from a mismatch involving the [ʃ] of [iʃ] than one involving the [ʃ] of [uʃ] because the spectrum of the [ʃ] of [iʃ] has more in common with an [s] spectrum than does the spectrum of the [ʃ] of [uʃ] (see figure 3). If the hearer sets up an expectation for an [ ] spectrum the reverse is true. The [s] spectrum from [us] would be easier to process while the [s] spectrum from [is] would be more difficult.

The segment perception hypothesis also leads to some predictions about how the hearer might recover from a mismatch. If segment information which is spread out over the syllable is integrated in the process of forming a segment identification, then the conflict between cues in the mismatched cases will have to be resolved. One way that a resolution between conflicting cues might be reached is by comparing the relative strengths of the cues. This method of recovery also results in predictions for the relative ease of processing the fricative mismatches in this experiment. Regardless of the transition cues the fricative that will be easiest to process as an [s] is the [s] of [is]. This is because this sound is the most extremely s-like [s] of the set. When the relative strengths of the cues for the final consonant are compared this 'strong' [s] will over-ride the misleading information in the transition more quickly than will the [s] from [us]. This same type of situation prevails when a vowel with alveolar transitions is paired with an alveopalatal fricative. The [ʃ] from [uʃ] will be easier to process because it is less like an [s] than any of the other [ʃ]'s.

Thus, the two theories make opposite predictions about the ways in which rounding in the fricative will help or hinder the preception of transition mismatched tokens. The syllable perception model predicts that coarticulatory rounding will make [s] easier to perceive when [ʃ] is expected, while the segment perception approach predicts that rounding will hinder the perception of [s] when [ʃ] is expected. The effect of rounding in [ʃ] perception has the opposite pattern of predictions. In perceiving by syllables, rounding an [ʃ] should inhibit reaction time while the perceiving by segments approach predicts that rounding should facilitate reaction time.

2. Methods

The tokens used were constructed from the syllables [us, os, as, is, uʃ, oʃ, aʃ, iʃ]. These syllables were recorded in an anechoic chamber using high quality equipment. The speaker was a male native speaker of American English. They were then digitally rerecorded at a sampling rate of 15 kHz (low pass filtered at 7 kHz). The digitized forms were edited so that the vowel and fricative portions were separated. The cut was made at the point at which the periodicity of the vowel ceased. In most cases there was a small amount of frication left on the vowel but this was so low in amplitude that it could not be heard when the vowel portion was played by itself.

Figure 4 shows the spectra of the eight fricative sounds used in this experiment (each graph is the average of 10 consecutive spectra from the first half of the fricative).

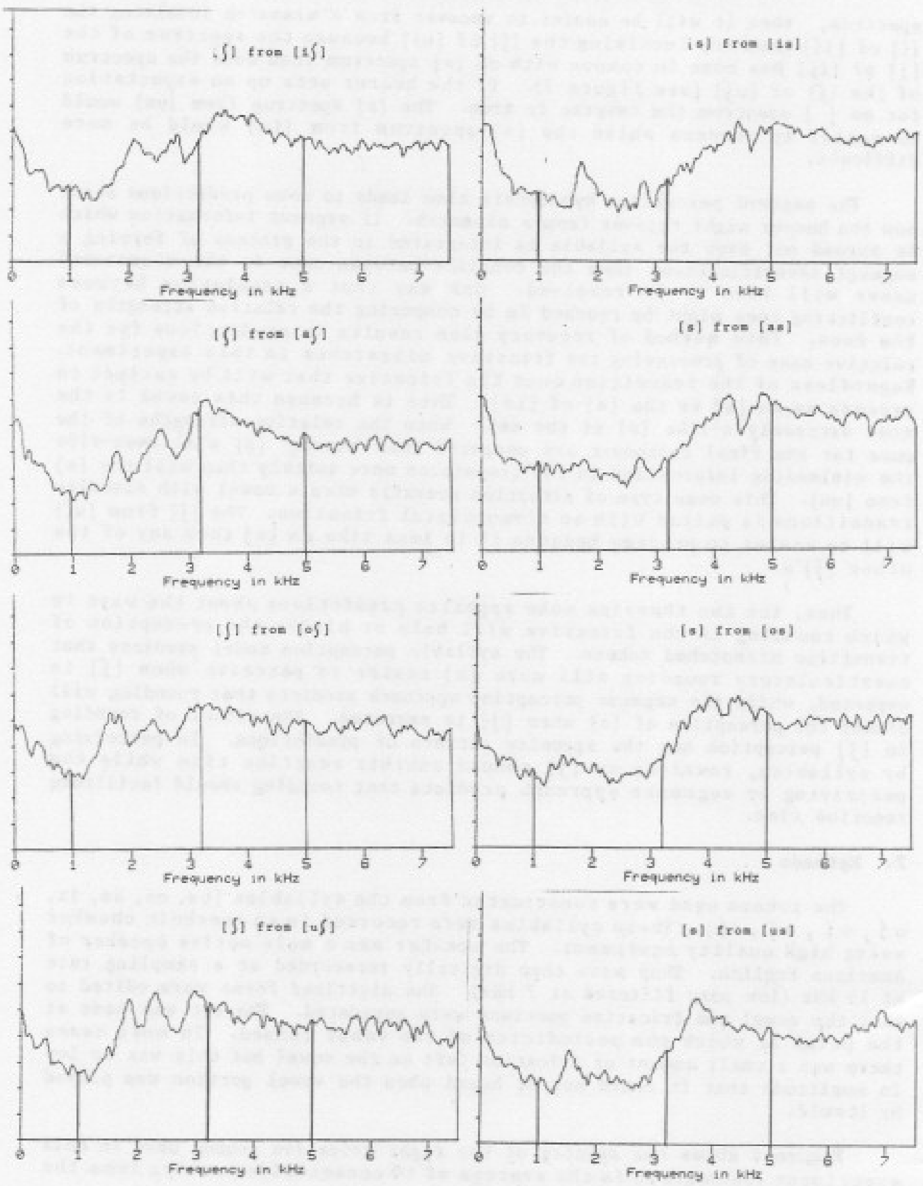


Figure 4: Spectra of the eight fricative sounds.

Using a technique that Jassem (1979) found effective in classifying fricative spectra, the spectra have been broken into regions (1000-3200, 3200-5000, and 5000-7500). Jassem found that by estimating the center of gravity in each of these three spectral regions the fricative can be correctly classified 80-90% of the time. For example, notice the middle region (3200-5000) in the [s] tokens. The center of gravity for the [s] from [is] in this region will obviously be greater than the center point of the region. As lip rounding increases the center of gravity decreases (to less than 4000). It is interesting to note that the center of gravity for the 3200-5000 region is very similar for the [s] from [us] and the [ʃ] from [iʃ]. These two are different in the first region but the similarity in the second is interesting. It makes it possible for us to consider these eight fricatives to be a type of continuum from the [s] of [is] which has a high center of gravity in region 2 and a low center of gravity in region 1, to the [ʃ] from [uʃ] which has a low center of gravity in region 2 and a high one in region 1.

The durations of the vocalic segments of the tokens were comparable to each other (intrinsic vowel length differences were retained). Likewise, the durations of the fricatives and the vowel fundamental frequencies were relatively uniform. This information is in table 1.

Table 1

	V	C	F0
is	226	215	161
iʃ	221	213	159
as	240	201	147
aʃ	238	206	155
os	238	204	157
oʃ	233	215	157
us	214	216	160
uʃ	224	211	156

Durations of the vocalic and consonantal portions of the stimulus items in milliseconds. F0 is in Hz.

Eight vowel tokens and eight fricative tokens resulted from cutting the VC syllables. In order to create the tokens which were used in the experiment each vowel token was combined with each fricative token. This is illustrated in table 2.

Table 2

The 64 tokens used in the experiment.

		vowels							
		u-s	u-ʃ	o-s	o-ʃ	a-s	a-ʃ	i-s	i-ʃ
	u-s	1	9	17	25	33	41	49	57
	u-ʃ	2	10	18	26	34	42	50	58
	o-s	3	11	19	27	35	43	51	59
fricatives	o-ʃ	4	12	20	28	36	44	52	60
	a-s	5	13	21	29	37	45	53	61
	a-ʃ	6	14	22	30	38	46	54	62
	i-s	7	15	23	31	39	47	55	63
	i-ʃ	8	16	24	32	40	48	56	64

i-s stands for the [s] from [is],
i-s stands for the [i] of [is], and so on.
Each VC combination is identified by token number (1-64).

By recombining the vowel and fricative portions in this way subcategorical mismatches are created. The two types of mismatches which are created in this particular case are (1) transition mismatches and (2) rounding coarticulation mismatches. For instance, in the first column in table 2 tokens 2, 4, 6 and 8 have transitions for [s] in the vowel but actually end in [ʃ]. Also, in the first column tokens 5-8 (and to an extent 3 and 4) are mismatched in lip rounding. They have a rounded vowel but a fricative which was originally produced with an unrounded vowel. This particular type of mismatch was central to the experiment here.

Fifteen paid subjects participated in this experiment. All subjects were native speakers of English and none reported any hearing loss. Each subject heard and responded to each of the sixty-four tokens described above four times (four blocks of 64 trials). Twenty practice items preceded the actual experimental trials. The experiment was conducted at the Linguistics Laboratory of The Ohio State University using a New England Digital Able 60 computer and the ERS experiment running package.

Subjects were seated in a quiet listening booth wearing Sennheiser HD420 headphones (the volume had been preset to a comfortable listening level). They were seated in front of a Heathkit VT52 computer terminal and responded to each token by pressing either the <s> key (for [s]) or the <h> key (for [ʃ]). Subjects responded to [ʃ] with their right hand and [s] with their left hand. One result of this arrangement was that an effect for handedness showed up in a main effect for fricative ($F(1,14)=8.923$, $p<.01$). The terminal was also used to provide subjects with reaction time feedback. Feedback during practice items included both reaction time and correct answers to the practice trials. The intertrial interval was 2 seconds.

Following Whalen (1984), only correct responses within a prescribed reaction time range (100 to 1000 ms) were included in the data analysis. The design of this experiment was such that only those tokens with transition mismatches were analyzed (this comes to 32*4 observations per subject). The overall error rate then was 11.25%.

3. Results

A three factor repeated measures analysis of variance was performed on the data collected in this experiment. The three factors were: The 'actual vowel' presented to the subject ([u,o,a,i]), the 'rounding' of the fricative presented (classed by the original vocalic context of the fricatives - [u,o,a,i]) and the 'fricative' sound actually presented ([s,ʃ]).

Figure 5 is a plot of mean reaction times as a function of the three factors. There is one graph for each of the four levels of the 'actual vowel' factor. The abscissa of each plot is used for the 'rounding' factor (four levels), and [s] identification is plotted with a dashed line while [ʃ] identification is represented by a solid line.

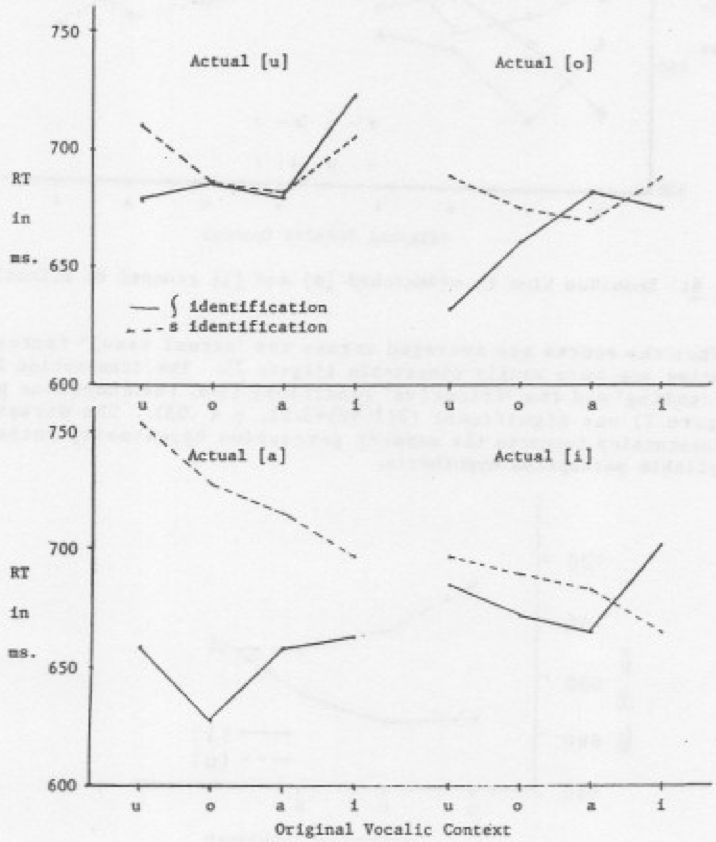


Figure 5: Reaction time to mismatched [s] and [ʃ] by actual vowel and rounding context.

In figure 6 the [ʃ] and [s] identification functions from each of the 'actual vowel' treatments are collected. The [ʃ] identifications tend to be faster when the original context of the fricative was [u], while [s] identification tended to be faster when the original context of the fricative was [i].

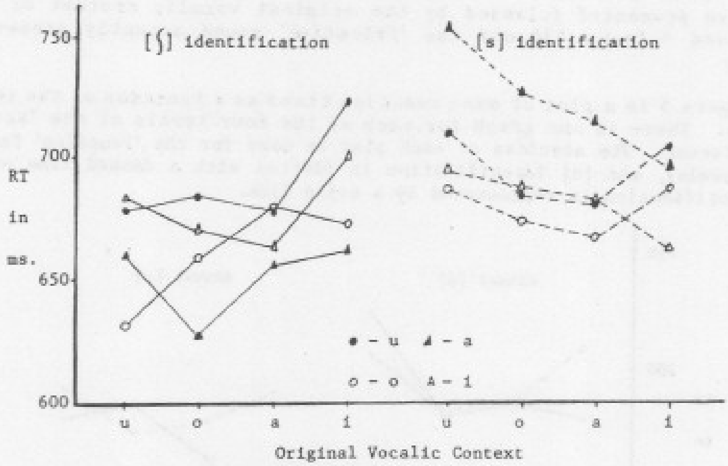


Figure 6: Reaction time to mismatched [s] and [ʃ] grouped by fricative.

When the scores are averaged across the 'actual vowel' factor these tendencies are more easily observable (figure 7). The interaction between the 'rounding' and the 'fricative' conditions (i.e. the functions plotted in figure 7) was significant ($F(3,42)=3.52$, $p < .05$). The direction of this interaction supports the segment perception hypothesis, rather than the syllable perception hypothesis.

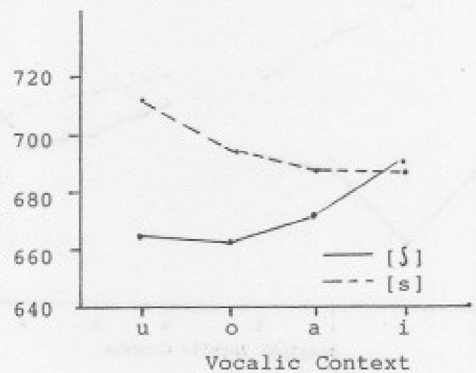


Figure 7: Reaction time to mismatched [s] and [ʃ] by rounding condition.

Two additional conditions proved to have significant results - the main effect for the 'actual vowel' treatment ($F(3,14)=2.8327$, $p<.05$) and the 'fricative' X 'actual vowel' interaction ($F(3,42)=8.169$, $p<.01$). These results are anomalous. None of the hypotheses being tested offer explanations for them. Thus, I will tentatively attribute them to some uncontrolled aspect of the tokens.

4. Conclusion

I've attempted to compare two general classes of speech perception theories. These I called segment perception and syllable perception. It is possible that there are other versions of these hypotheses which would entail different predictions from those tested in this experiment. However, for such alternatives to be useful they must make predictions which are explicit enough to be tested.

The results of this experiment indicate that the class of speech perception theories which entail perception by segments correctly characterize the nature of speech perception (at least in the case of post-vocalic fricatives). This leaves open a wide variety of questions concerning the perception of speech segments. It is still possible to posit active models or passive models, analysis by synthesis or motor approaches. Yet, one thing is suggested by this experiment: syllables in speech are perceived as the result of segment perception, not vice versa.

Acknowledgements

Thanks to Neil Johnson and Rob Fox for their comments and advice at the birth of this project. I appreciate the comments that I've received from Ilse Lehiste, Mary Beckman, Bruno Repp and Dennis Klatt. I'm responsible for my own errors. Thanks also to Debbie Stollenwerk for not talking about linguistics and to Keith Green for musical assistance.

References

Bondarko, L. V.; Zagorujko, N. G.; Kozhevnikov, V. A.; Molchanov, A. P. and Chistovich, L. A. 1970. A model of speech perception in humans. Working Papers in Linguistics. no. 6, Computer and Information Research Center, Ohio State University.

Fant, C. G. M. 1967. Auditory patterns of speech. in Wathen-Dunn, W. (ed) Models for the Perception of Speech and Visual Form. Cambridge, Mass.: MIT Press.

Jassem, W. 1979. Classification of fricative spectra using statistical discriminant functions. in Lindblom, B. and Ohman, S. (eds) Frontiers of Speech Communication Research. New York: Academic Press.

Klatt, D. H. 1979. Speech perception: a model of acoustic-phonetic analysis and lexical access. Journal of Phonetics. 7:279-312.

Liberman, A. M.; Cooper, F. S.; Shankweiler, D. P. and Studdert-Kennedy, M. 1967. Perception of the speech code. Psychological Review. 74:431-461.

Massaro, D. W. 1972. Preperceptual images, processing time and perceptual units in auditory perception. Psychology Review. 79:124-145.

Morton, J. and Broadbent, D. 1967. Passive versus active recognition models, or is your homunculus really necessary? in Wathen-Dunn, W. (ed) Models for the Perception of Speech and Visual Form. Cambridge, Mass.: MIT Press.

Pisoni, D. and Sawusch, J. R. 1975. Some stages of processing in speech perception. in Cohen, A. and Nooteboom, S. G. (eds) Structure and Process in Speech Perception. New York: Springer-Verlag.

Stevens, K. N. and Halle, M. 1967. Remarks on analysis by synthesis and distinctive features. in Wathen-Dunn, W. (ed) Models for the Perception of Speech and Visual Form. Cambridge, Mass.: MIT Press.

Whalen, D. H. 1984. Subcategorical phonetic mismatches slow phonetic judgments. Perception and Psychophysics. 35:49-64.

OSU #36: Papers from the Linguistics Laboratory 1985-1987, Sept. 1987.

Some Constraints on the Domain of Phrasal Resyllabification in French

Christiane Laeufer
Romance Languages and Literatures
Ohio State University

Abstract: This paper examines the French phrasal resyllabification process known as enchaînement 'linking' in the light of three recent proposals about connected speech phenomena and their relation to syntactic structure. It is based on a corpus illustrating linking consonants and initial consonants in minimal or near minimal pairs across different types of prosodic/ syntactic boundaries. The sentences were read five times at three different speaking rates by two native speakers. The study shows that (1) with increasing rate of speech, the domain of the rule becomes larger, irrespective of prosodic structure, and (2) there exists a specific degree of disjuncture beyond which it does not apply in a given tempo. The study thus provides evidence that, at least in French, resyllabification belongs to the phonological rules proper, and not to the rules for defining (post-lexical) phonological representation. Research on phrasal resyllabification rules in other languages and based on a larger number of speakers is, however, needed before any generalizations can be made.

0. Introduction

One of the important issues in phonological analysis is the relation between surface syntactic structure and phrasal (postcyclic) phonology. In particular, recent debate has centered on the existence and the nature of prosodic structure, and its mediating role between syntax and phonology.

In Kaisse 1985 the syntax-phonology relation is considered to be direct, with syntactic structure feeding directly into word-external phonology without any intermediary (prosodic) stage. In Nespor and Vogel 1982, and Selkirk 1984, on the other hand, this relation is seen as an indirect one, mediated by a set of prosodic rules which interpret surface syntactic representation into phonological representation. The former consider this mapping process as resulting in a single hierarchical (accentual, intonational and sentential) prosodic constituent structure. The latter sees it as resulting in two distinct sorts of hierarchical organizations, namely prosodic constituent structure as just described, and rhythmic structure, that is, the alignment of syllables with a metrical grid. The grid corresponds to a series of levels consisting of a sequence of positions (or beats) which stand for points in time and "define the recurring temporal periodicities of rhythm" (Selkirk 1984:7).

Both of the latter models also differ from Kaisse with respect to their predictions about the effect of prosodic structure on phrasal phonological rules. Nespor and Vogel claim that particular rules are restricted to particular prosodic (accentual, intonational or sentential) domains. Selkirk contends that connected speech rules (but possibly not phrasal resyllabification rules) are sensitive to rhythmic structure, and that prosodic, that is, mainly intonational structure, is important only in so far as it affects the realization of the metrical grid. More specifically, phrasal rules are said to require some degree of adjacency in time which is defined with respect to a certain number of so-called silent grid positions. These positions represent varying degrees of rhythmic disjuncture, and are introduced on the basis of syntactic structure by a rule of *Silent Demibeat Addition* (Selkirk 1984: 313ff).

One factor that plays a role in all three models is rate of speech. Kaisse distinguishes between two types of connected speech phenomena, depending on their sensitivity to tempo: so-called *rules of external sandhi* which are affected only by syntactic, lexical and stylistic factors, and so-called *fast speech rules* which are said to be sensitive only to rate of articulation and to apply across the board, that is, across any structural boundary. Nespor and Vogel (1982:234, 239) subsume rate of speech under a constellation of extra-linguistic factors also including style, constituent size and presumably others, which are assumed to bring upon *restructuring* of smaller prosodic units into larger, more inclusive ones and thereby account for a certain amount of variability in phrasing.

In Selkirk such variability follows from the fact that prosodic structure is freely assigned to the surface structure of an utterance, resulting in more than one possible prosodic representation subject to a well-formedness condition. The assumption with respect to rate of speech is that for any tempo a real-time value is assigned to individual grid positions, and the faster the tempo, the shorter the real-time duration of the grid position. With any phrasal rule there is associated a specification of the real-time adjacency it requires, say *n* msec. for rule *x*. It follows that with increasing tempo, the domain of such a rule is expected to be extended, since it will take more silent grid positions to reach *n* amount of real-time. This approach to disjuncture does not therefore require any changes in the formal representation with changes in tempo.

Surprisingly, all three models use a number of identical phonological rules such as *raddoppiamento sintattico* in Italian, French *liaison*, American English *flapping*, as illustrative and corroborative examples. The first is a resyllabification rule which accounts for the lengthening of the initial consonant of the first word in a sequence of two words, under certain phonological and syntactic conditions (e.g. *parlo[b]ene* --> *parlo[b:]ene*, 'he spoke well' (Nespor and Vogel 1982:227). The second refers to the syllabic association of an otherwise unpronounced (so-called 'mute') consonant with a following vowel-initial word, as in *petit ami* 'boyfriend' /pətit ami/ --> [pə-ti-ta-mi] vs. *petit tamis* /pətit tami/ --> [pə-ti-ta-mi] 'small strainer', where hyphens represent syllable boundaries. The third is a phonological rule which is contingent upon previous resyllabification (more specifically, ambisyllabification) of consonants with unstressed initial vowels in following words, and which reduces ambisyllabic alveolar stops to voiced flaps, as in *get a pen* /gɛt ə pɛn/ --> resyllabification: /gɛtə pɛn/ --> flapping: [gɛDəpɛn], with the ambisyllabic consonant underlined. All three rules thus have in common the fact that they involve resyllabification. Given the fact, however, that the conceptual differences between the three proposals lead to very different empirical predictions, one would not expect identical rules to concurrently support all three proposals. It seems therefore that a closer look at connected speech phenomena, based on experimental evidence, is in order to evaluate these three theories.

The present study is a first step towards such an empirically based examination of phrasal rules. It investigates the influence of the three factors (1) rate of speech, (2) degree of disjuncture and (3) prosodic structure on the domain of French enchaînement ('linking'), in light of the three proposals just summarized about connected speech phenomena and their relation to syntactic, and prosodic or rhythmic structure.

Enchaînement, that is, the phrasal resyllabification of a word-final 'nonmute' consonant with the following vowel-initial word as in petite amie 'girlfriend' [pə-ti-ta-mi], is examined here rather than liaison, because it truly exemplifies resyllabification of consonants from one syllable to another, while liaison technically speaking does not. Recent autosegmental analyses of final consonants in French posit underlying extra-syllabic final consonants, that is, consonants which are not associated with any syllable. In the course of the derivation, such consonants are either syllabically integrated into the preceding word by a morphological rule such as feminine formation, and later (in postlexical phonology) resyllabified with the following vowel-initial word (enchaînement); or they are integrated by a phrasal phonological rule (liaison) into the following word if it begins with a vowel, and remain otherwise extrasyllabic, that is, are phonetically unrealized (e.g. Encrevé 1983, Clements and Keyser 1983).

After exposing the insufficiency of previous studies on French enchaînement with respect to rate of speech, syntactic disjuncture and prosodic structure, an analysis of the French prosodic system based on recent phonetic work is sketched out in section 2. The experimental design of the experiment is outlined next with a list of the (syntactic and prosodic) contexts and the corpus, followed by the description of the data analysis with respect to a number of phonetic cues. The results are presented in section 5. They focus on the presence or absence of a pause following the relevant linking consonant, the occurrence or nonoccurrence of prosodic restructuring of the phrases on either side; the phonetic realization of the consonant; and various duration measurements which provide evidence for syllabic structure, namely the duration of the linking consonant, of the preceding vowel, and the formant transitions leading form the consonant to the vowels on either side. The discussion centers around the effects of tempo in its interrelation with syntactic and prosodic structure. It is shown that only Selkirk's model makes the right predictions, namely that (1) the domain of the French resyllabification rule becomes larger with increasing rate of articulation, irrespective of prosodic structure, and (2) that there exists in a given tempo a specific degree of syntactic disjuncture beyond which it does not apply.

1. Existing Studies on Enchaînement

Whereas there exist many studies of the various (syntactic, prosodic, stylistic, sociolinguistic, etc.) constraints on French liaison, enchaînement has received much less attention by linguists. A thorough examination of existing phonological and phonetic descriptions of French reveals merely a few references to some stylistic and potential prosodic constraints.

The stylistic constraints rest on a difference between normal connected speech and a very explicit and conscious style of speech in which enchaînement can be suspended, mainly to fulfill a disambiguating function, or as a means for contrasting and emphasizing, that is, intentionally isolating (Malmberg 1964:117ff.). Based on experimental studies by Grammont 1933 and Durand 1939 who recorded oral air pressure and "glottal tension" on oscillograms and with the help of other machinery, the difference between initial and linking consonants in monitored speech is said to correspond to a difference in articulatory tension, that is, contraction of the vocal tract muscles (in particular, the laryngeal muscles). Initial consonants are said to be articulated with increasing tension, and linking consonants with decreasing tension (see also Delattre 1940).

The rare references in the literature to the prosodic domain of enchaînement are not experimentally backed and reveal divided opinions. Older descriptions of French, such as Vidon-Varney (1933:141), suggest that enchaînement usually occurs between what she calls rhythm groups, and sometimes even between what she calls breath-groups, an opinion which is shared by Pernot (1937:337): "Que ce soit à l'interieur d'un groupe rythmique ou d'un groupe à l'autre, à condition que l'émission de voix ne cesse pas, l'enchaînement est toujours possible et même désirable". In more recent descriptions, it is suggested that enchaînement takes place only within rhythm groups (e.g., Léon 1966:118). Similarly Pulgram 1965, 1970, defines rhythm groups as the smallest units of utterance within which word boundaries are segmentally and suprasegmentally obliterated, and which are demarcated by oxytonic stress and bounded by disjuncture, thereby implying the absence of resyllabification across such units. The two prosodic domains referred to in these works are discussed in the next section.

2. The French Prosodic System

Investigators of French prosody agree that the prosodic parameters operating in the language, namely fundamental frequency (henceforth, Fo) and phrase-final lengthening, divide longer sentences into at least two hierarchical levels of prosodic units (e.g., Di Cristo 1975, Martin 1980, 1982, Vaissière 1974, 1975, 1980, Delgutte 1978).

The first level unit is the breath-group.[1] Its boundaries are signaled by a variety of phonetic cues, including those listed in (1) outlined in Vaissière (ibid.).

---

[1] There exist a wealth of different terms for the prosodic units in French, and the use of a same term by different researchers to represent a different unit and different terms to represent the same unit adds to the confusion. In the present study the terms breath-group and rhythm group are used for, respectively, the first and the second level units. They correspond roughly to Nespor and Vogel's intonation and accentual phrases and will both be subsumed under the term prosodic unit when a differentiation between them is unimportant for the discussion at hand. Furthermore, the term stressed syllable will be used to refer to the prosodic unit-final syllable.

(1) 1. a resetting of the baseline;
2. the presence of a rising contour at the beginning of the last word of a nonsentence-final breath-group, followed by a fall and terminated by a rising contour at the end of the word (= Vaissiere's P1 pattern); or
3. a rising-falling movement on the last word of a sentence-final breath-group (pattern P4); and
4. frequently, surrounding pauses.

The second level unit is the <u>prosodic word</u> or <u>rhythm group</u> (see note 1), with rising intonation at the onset of the final word which is sustained or gradually falls until the peak (sharp rise followed by lowering) on the last syllable (pattern P2). These Fo patterns and one additional pattern found exclusively within rhythm groups (pattern P3) are shown in Fig.1.[2]

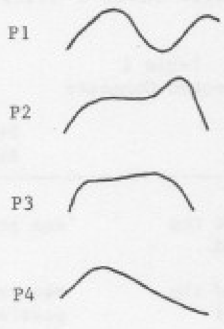


Fig.1: The four patterns characteristic of the Fo contours of French (adapted from Vaissière 1975).

These patterns are mostly co-occurrent with content words of three syllables or more. Shorter words often reveal a simpler, incomplete pattern, or share a pattern with the preceding and/or following word. The intonation curve which is always preserved, even on shorter words, is the final continuation rise of P1 and the final peak of P2.

Besides the prosodic unit-final high boundary tone (the continuation rise of pattern P1 or the final peak of P2), prosodic boundaries manifest themselves through an increased duration of the group-final syllable, compared to the surrounding syllables (e.g., Di Cristo 1975, Vaissière 1980).

---
[2]See Delgutte 1978 for an extension of the number of Fo patterns to six by differentiating between prominent and nonprominent, as well as demarcative and nondemarcative patterns.

It is also generally agreed upon that the number of prosodic units tends to decrease as the rate of elocution increases (e.g., Vaissière 1975a:251, 1980:553, 557). This presumably universal characteristic of prosodic systems is accounted for in Nespor and Vogel by a process of restructuring of two lower level units into one higher level unit. As mentioned above, in Selkirk's model no changes in the formal representation are required for different tempos, as degrees of syntactic disjuncture are encoded in the number of silent grid positions between two constituents to which real-time values are assigned for different tempos; and Kaisse does not address the question of prosodic structure, since it is assumed to play no role in defining the domain of phonological rules.

3. Experimental Design

A corpus was constructed so as to illustrate enchaînement consonants compared to initial consonants in seven minimal or near minimal pairs across different types of prosodic boundaries corresponding to different degrees of syntactic disjuncture. Table I illustrates the particular contexts chosen.

Table I
Selected Contexts

Sentence Type	Syntactic Structure	Prosodic Structure
(1)	a modifier and the following head	one prosodic unit
(2)	a modifier and the preceding head	two prosodic units, with possible restructuring at faster speaking rates
(3)	two conjoined phrases	same
(4)-(6)	a subject and the following predicate:	
	-short subject	same
	-longer subject	same
	-long subject	two prosodic units, restructuring unlikely
(7)	a detached phrase and the following clause	same

The full corpus used in the experiment is given in the appendix. It was presented to two native speakers, one male and one female, who were asked to read each sentence type five times at three different tempos (slow, "normal" and fast speech) which were to be determined by the speakers. This resulted in 210 tokens for each speaker (five tokens for each sentence at a given tempo x three tempos x 14 sentence types) which were presented to the subjects in a random order. Recordings took place in a sound treated studio.

The study was restricted to the consonant /r/, because (1) it belongs to the consonants which lengthen any tautosyllabic preceding vowel in French (especially under stress, e.g. Delattre 1951:15ff.), thereby providing evidence about syllable structure; and (2) it has been shown to exhibit a certain amount of phonetic variation depending on its distribution and syllabic position (Borel-Maisonny 1942, Straka 1965, Simon 1968, Rialland 1984).

4. Analysis

Each utterance was analyzed using broad band spectrograms made on a voice Print 700 and examined with respect to the phonetic cues listed in (2) which have been shown to be good indicators of syllabic structure in languages (e.g., Malmberg 1955, Lehiste 1960, Gårding 1967).

(2) a. the phonetic realization of /r/
 b. its duration
 c. the length of the second formant transitions to the surrounding vowels
 d. the duration of the preceding vowels
 e. the eventual presence and length of a following pause.

The duration measurements were made from the spectrograms to the nearest 10 msec., based on commonly used spectral cues to determine the boundaries of segments, such as the disappearance of the second formant (and sometimes also the first) and the appearance of a (more or less) pronounced noise pattern in the mid-frequency region to delimit (voiced or voiceless) fricative /r/ from the surrounding vowels. At times, especially in faster speaking rates, the syllable-final sonorant was fully vocalized, distinguishable from the preceding nucleus only by a change in the formant structure and the weakening or disappearance of the upper formants. The quality of the preceding vowel also plays a role: the delimitation was most difficult when syllable-final /r/ followed the vowel /a/ due to their acoustic similarity. Similarly, the intervocalic (initial and linked) sonorant was sometimes, especially in normal and fast rates of articulation, realized as a glide-like segment, separable from the nuclei on either side only on the basis of the changes in the upper formants.

The different realizations of the consonant were tabulated against the scale shown in (3).

(3) a) vocalic segment

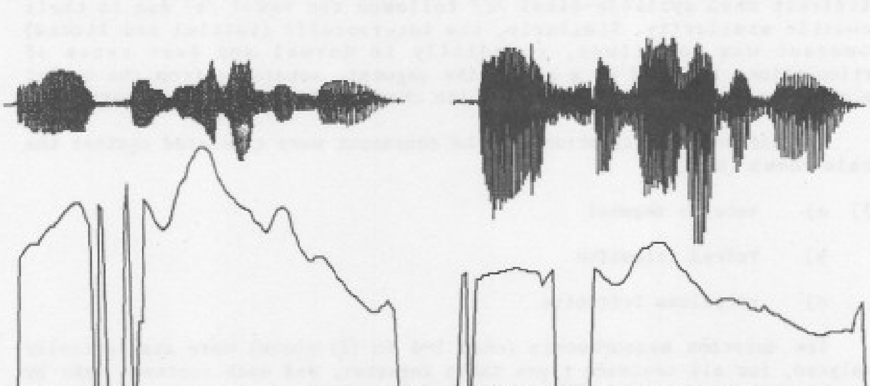
 b) voiced fricative

 c) voiceless fricative

The duration measurements (cues b-d in (2) above) were statistically analyzed, for all sentence types taken together, and each sentence type by itself, for both speakers, as well as for each speaker separately. Means (in msec.) of segment durations, of formant transition lengths, as well as the corresponding standard deviations were calculated. Two-tailed t-tests were performed and probabilities were calculated for differential means between contexts with initial, as opposed to linked /r/, and contexts with initial, as opposed to non-linked /r/ (cf. Fig. 6a-d below).

Furthermore, fundamental frequency tracings of all utterances containing linked /r/'s were obtained using the PDP11 with the Interactive Laboratory System Speech Analysis Package. The Fo contours of these tokens were interpreted in terms of Vaissière's P1 and P2 patterns. Most tokens had been read by the speakers with an unmarked intonation pattern, that is, without emphasis on a particular word, and with unmarked information focus, that is, as answers to the question "<u>What</u> <u>happens/happened?</u>".

In fact, in the contexts occurring in the present study, namely mono- or disyllabic words followed by a vowel-initial word in the next unit, it was not possible in the absence of a pause to distinguish between patterns P1 (i.e. first level prosodic unit) and P2 (i.e. second level unit) on the basis of the gross overall shape alone, since in both the final rise was followed by a lowering. The decision as to whether restructuring took place was based on (1) a comparison of the intonation contours of the various tokens of a particular sentence-type in a given tempo with respect to relative amplitude of Fo variation, (2) a comparison of the pre- and posttonic syllables with the tonic syllable in the different tokens with respect to relative Fo amplitude, and (3) a comparison of the durations of the prosodic unit-final syllable nuclei. Fig. 2 represents the intonation contours of a non-restructured and a restructured token of sentence-type 2. The crucial factor lies in the relative height of the final peak of <u>suivirent</u> which reaches 222Hz in the former, but only 156Hz in the latter. A second, generally less consistent cue is the relative duration of the final syllable, with the tonic syllable of the non-restructured token being 35 msec. longer than the same syllable in the restructured token, in this case nontonic and not prosodic group-final. Only the length of the nucleus measured from spectrograms was taken into account in the actual tabulation of relative final syllable durations. Finally, a third and even less consistent factor is the relative Fo height of pre-, tonic and posttonic syllables which is difficult to interpret due to the segments' instrinsic pitch differences and various microprosodic effects from surrounding



<u>Fig. 2</u>: Wave-form and Fo contour of a nonrestructured (left) and a restructured (right) fast speech token of sentence type 2 (<u>Ils</u> <u>nous</u> <u>suivirent</u> <u>avidement</u>) by the male speaker.

consonants. In the non-restructured token in Fig.2 the pre- and posttonic syllables are much lower in Fo amplitude than the tonic syllable. In the restructured token, the pretonic syllable has a slightly higher Fo than the tonic one, and the Fo of both is higher than the Fo of the posttonic syllable.

5. Results

5.1. Linking

The number of tokens containing linked vs. non-linked consonants was tabulated, based on (1) the absence of a following pause, (2) the absence of any laryngealization in the onset of the following vowel (ever so slight, if present),[3] and (3) the durations of /r/, the preceding vowel, and the formant transitions (see section 5.4. for discussion of exact measurements). Fig. 3 summarizes the count.

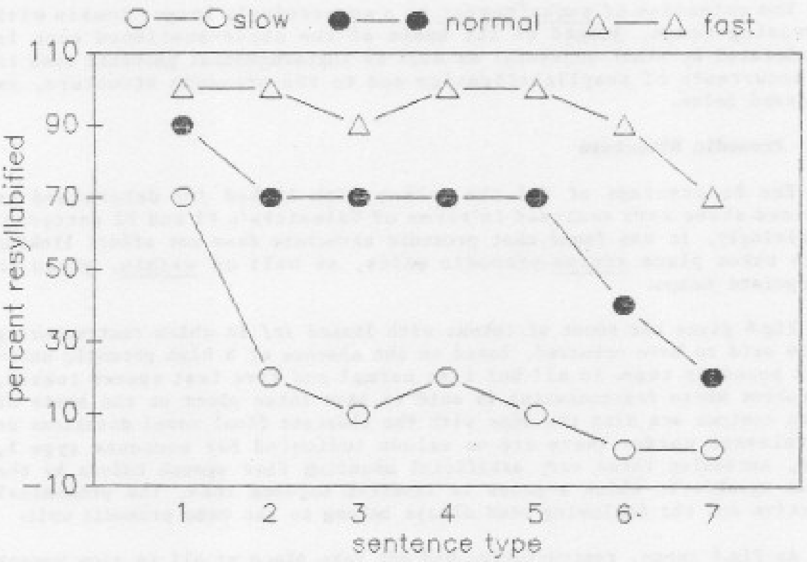


Fig. 3: Percentages of linked /r/'s in each sentence type and tempo.

As the figure suggests, the domain of resyllabification is enlarged with increasing rate of articulation to contexts with a greater degree of syntactic disjuncture. The total percentages for each speed are as follows:

---

[3]Interestingly, the corpus contained no occurrence of glottal stop, contrary to Léon's 1971 findings based on an analysis of radio programs, from which he concluded that in this particular speech style at least, this juncture marker tends to be generalized as a syntactic reinforcement at the begining of prosodic phrases.

slow speech 13%, normal speech 43%, fast speech 65%. Enchaînement is considered to have taken place in a particular sentence type, that is, in a particular syntactic and prosodic context, at a given tempo, when at least 70% (i.e. 7 out of 10) of the tokens contain a linked /r/. In slow connected speech, linking only takes place between words standing in close syntactic and prosodic connection (i.e., prosodic unit-internally) such as the head noun and its preceding modifying adjective exemplified in the first sentence type. (The few exceptions are due to some overconscious renditions by the female speaker, possibly influenced by the at times close proximity of the other member of the minimal pair.) At normal rate of speech, linking is extended to some contexts between two prosodic units which, according to Nespor and Vogel, tend to undergo restructuring into a single prosodic phrase in faster tempos. Finally, in fast speech, it applies in all the contexts tested, including across the two breath-groups in sentence types 6 and 7 in which restructuring is said to be highly unlikely (e.g. Nespor and Vogel, 1982:246).

The extension of enchaînement to a successively larger domain with increasing tempo, judged on the basis of the above-mentioned cues is corroborated by other segmental as well as suprasegmental phonetic cues to the occurrence of resyllabification and to the prosodic structure, as discussed below.

5.2. Prosodic Structure

The Fo tracings of all the tokens with linked /r/ determined as outlined above were analyzed in terms of Vaissiere's P1 and P2 patterns. Surprisingly, it was found that prosodic structure does not affect linking which takes place across prosodic units, as well as within, given an appropriate tempo.

Fig.4 gives the count of tokens with linked /r/ in which restructuring can be said to have occurred, based on the absence of a high prosodic unit-final boundary tone. In all but five normal and five fast speech tokens, the tokens where restructuring is said to have taken place on the basis of the Fo contour are also the ones with the shortest final vowel durations on the relevant words. There are no values indicated for sentence type 1, since, excluding three very artificial sounding slow speech tokens by the female speaker in which a pause is inserted between them, the prenominal adjective and the following noun always belong to the same prosodic unit.

As Fig.4 shows, restructuring did not take place at all in slow speech where the sentences (except sentence type 1) were divided by a pause following the relevant word in /r/, and it is still relatively rare at a "normal" rate of speech (14% of all tokens with linked /r/). The few occurrences of restructuring in normal speech are all found in the male speaker's tokens which were generally read at a slightly faster tempo and in a more casual style than the female speaker's renditions.

The more frequent occurrences of restructuring in fast speech (33% of all tokens with linked /r/) seem to have been determined by the syntactic context, by the length of the utterance and of the constituents involved, although here too, tempo and style play a role (it is again the male speaker who restructures more frequently) and so do presumably other factors as well.

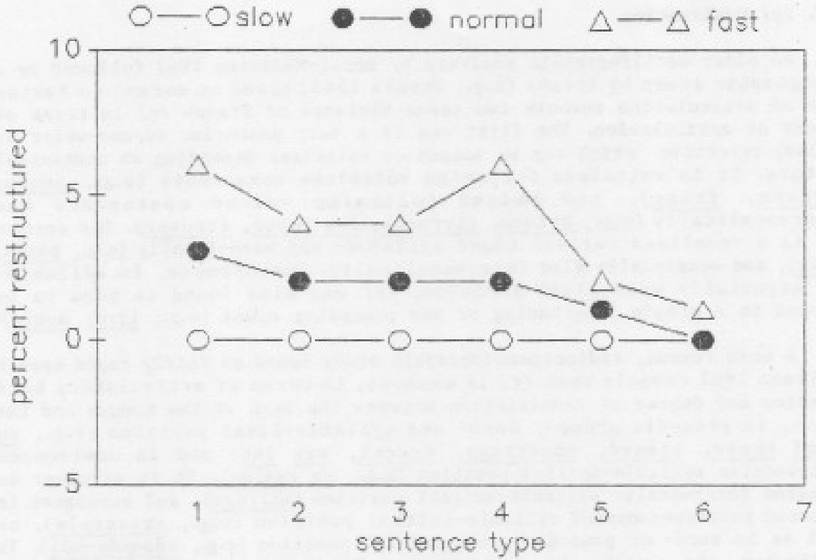


Fig. 4: Incidence of restructuring in tokens with linked /r/ for each sentence type and tempo.

The closer the syntactic context between two words separated by a prosodic phrase boundary, the more likely restructuring is, as shown by the frequency of its occurrence in sentence type 2 between the head verb and its adverbial modifier, or the two members of a conjoined phrase in sentence type 3. Length of utterance seems to be important in two ways. All other things being equal, a very short utterance potentially containing more than one prosodic unit has a greater chance of being restructured than a longer utterance with the same number of prosodic units, as a comparison of sentence types 4 and 6 shows.[4] Given constituents of similar length, such as the noun phrases in sentence types 4 and 5, restructuring is favored if the utterance as a whole is longer, with a major prosodic boundary in close proximity (sort, et in sentence type 5). Furthermore, the occurrence of restructuring in sentence type 7 contradicts the general assumption (e.g., Cinque 1977) that 'detached' phrases (in particular, so-called hanging topics which promote a previously non-topic noun phrase to topic status) are obligatorily set off as a separate prosodic phrase from the rest of the sentence.

Crucially, however, although resyllabification and restructuring are affected by the same types of factors, restructuring itself does not constitute a necessary condition for the application of resyllabification.

---

[4]Interestingly enough, in these two sentence types with monosyllabic prosodic group-final words the final high tone was sometimes delayed and realized on the following vowel-initial clitic.

5.3. /r/ Realization

An older oscillographic analysis by Borel-Maisonny 1942 followed by a kymographic study by Straka (e.g., Straka 1965) based on normal to faster rate of articulation reveals two basic variants of French /r/ in terms of manner of articulation. The first one is a very posterior (dorso-velar or uvular) fricative which can be voiced or voiceless depending on contextual factors. It is voiceless following voiceless consonants (e.g., prune, trouver, frise), and voiced following voiced consonants and intervocalically (e.g., brique, gavroche, des roses, arroser). The second one is a vocalized variant found syllable- and word-finally (e.g., perdu, clair), and ocasionally also intervocalically. Furthermore, in syllable- and especially word-final position, /r/ was also found to tend to be reduced to a simple lengthening of the preceding vowel (e.g., lire, munir).

A more recent, radiocinematographic study based on fairly rapid speech by Simon 1967 reveals that /r/ is weakest, in terms of articulatory hold duration and degree of constriction between the back of the tongue and the velum, in prosodic group-, word- and syllable-final position (e.g., au grand bazar, canard, encartage, Marcel, sur les) and in unstressed intervocalic syllable-initial position (e.g., un raglan). It is stronger in stressed intervocalic syllable-initial position (barrage), and strongest in stressed post-consonantal syllable-initial position (e.g., casserole), as well as in word- or prosodic group-initial position (e.g., réponds-moi). In particular, she reports the measurements (in msec.) listed in Table II for the contexts relevant for the present study.

Table II

Measurements (compiled from Simon 1976)

	Group-Initial	Intervocalic stressed	Intervocalic unstressed	Group-Final
articulatory hold duration (in msec.)	60	60	40	40 (often no hold at all)
degree of constriction (diameter in mm.)	10	8.5	10	13

The realizations found in the present corpus only partly support the results of these previous studies. The tabulated phonetic variants of /r/, listed against the acoustic scale described in (3) above, and according to tempo and syllabic position, are given in Fig.5a through c (for a more detailed analysis, see Laeufer, in preparation).

In slow speech the majority of the initial (61%) and final unlinked (57%) realizations of /r/ cluster at the voiceless fricative end. Presumably this articulatorily speaking rather strong realization of final /r/ is due to the monitored and conscious articulation characteristic of slow rate of speech, as well as the position of the consonant under stress,

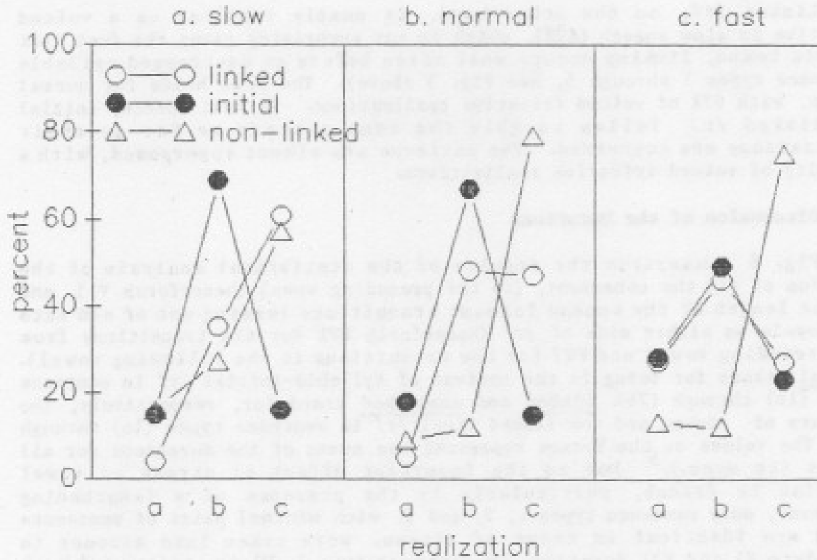


Fig. 5: Initial, linked and nonlinked phonetic variants of /r/ in slow (5a), normal (5b) and fast (5c) speech.

at the end of a prosodic unit before a pause. The few occurrences of vocalized final /r/ (16%) are found in the male speaker's tokens which are read on the average with a slightly more rapid tempo and more casually than the female speaker's. As for the high incidence of initial /r/ as a voiceless fricative (61%) it is explainable by its position under stress in most sentence types. When unstressed, as in sentence type 5 and some realizations of types 1 and 2, it is realized as a voiced fricative.

In normal speech, most of the realizations of initial /r/ are divided almost equally between the voiceless and the voiced fricative pronunciations (47 and 48% respectively). The latter pronunciation is characteristic of the male speaker's renditions and the former of the female speaker's. Most realizations of unlinked final /r/ (79%), on the other hand, are voiceless, and belong to the female speaker's renditions which, as noted above, were read at a slower pace and more consciously articulated than the male speaker's.

In fast speech, on the other hand, there is a clear difference between intervocalic syllable-initial /r/ the articulation of which tends to be located at the "voiced" end of the scale, with 46% of voiced fricative and 27% of vocalized realizations, and the few occurrences of final /r/ (again belonging to the female speaker's tokens) most which are still fairly strongly articulated (75% have voiceless realizations), as opposed to the "weak" pronunciation of prosodic group-final /r/ in fairly rapid speech reported in Simon 1967 and described above.

Linked /r/, on the other hand, is mostly realized as a voiced fricative in slow speech (69%), which is not surprising given the fact that in this tempo, linking occurs most often before an unstressed syllable (sentence types 1 through 5, see Fig. 3 above). The same holds for normal speech, with 67% of voiced fricative realizations. In fast speech, initial and linked /r/ follow roughly the same pattern as far as their realizations are concerned. The patterns are almost superposed, with a majority of voiced fricative realizations.

5.4. Discussion of the Durations

Fig. 6 summarizes the results of the statistical analysis of the duration of (1) the consonant, (2) the preceding vowel (henceforth V1), and (3) the length of the second formant transitions leading out of and into the vowels on either side of /r/ (henceforth FV1 for the transitions from the preceding vowel and FV2 for the transitions to the following vowel). Initial stands for being in the context of syllable-initial /r/ in sentence types (1b) through (7b). Linked and nonlinked stand for, respectively, the contexts of linked and non-linked final /r/ in sentence types (1a) through (7a). The values on the Y-axes represent the means of the durations for all tokens (in msec.).[5] Due to the important effect of stress on vowel duration in French, particularly in the presence of a lengthening consonant, only sentence types 1, 2, and 5, with minimal pairs of sentences which are identical in terms of stress, were taken into account to calculate V1 and FV1 duration. In sentence type 1, V1 is unstressed both before initial and linked /r/; in sentence types 2 and 5, V1 is stressed before both types of /r/'s, except for the tokens with prosodic restructuring, in which it is unstressed.

5.4.1. Initial versus non-linked final /r/

One cannot fail to notice the remarkable similarity of the patterns across the different graphs. As expected, the mean durations generally decrease from slower to faster speech due to the general compression of segments in time caused by an increasing speaking rate. As also expected, mean V1 and FV1 durations in both slow and normal speech are significantly greater before an unlinked syllable-final consonant than before an initial hetero-syllabic consonant (in fast speech, all tokens of sentence types 1, 2, and 5 contained linked /r/'s). In slower speech, the difference in the duration of V1 in sentence types 2 and 5, in which the prosodic unit-final syllable is stressed, is close to 80% (mean difference = 120.74, T-stat. = -9.404, Prob. = 0.0001). O'Shaughnessy 1981 estimates it at 130% for monosyllabic words in isolation, and Rialland 1984 at around 80% for /a/ in the pair le bar trouvé 'the discovered bar' vs. le bas r' trouvé 'the rediscovered stocking'. In normal speech the difference is 47% (mean difference = 30.50, T-Stat. = -3.074, Prob. = 0.036). The difference in V1

---

[5]There were some missing values due to (especially slow speech) pronunciations by the female speaker of some N forms with a final schwa (often heard as a strong vocalic release), in particular in bars 'bars' [ba:rə], Pierre 'Peter' [pjɛ:rə], suivirent 'followed' [sɥivi:rə]. Missing formant transition values are due to an impossibility at times to clearly locate and measure the transitions, often due to the lack of a clear steady state in /r/ and/or in the surrounding vowels.

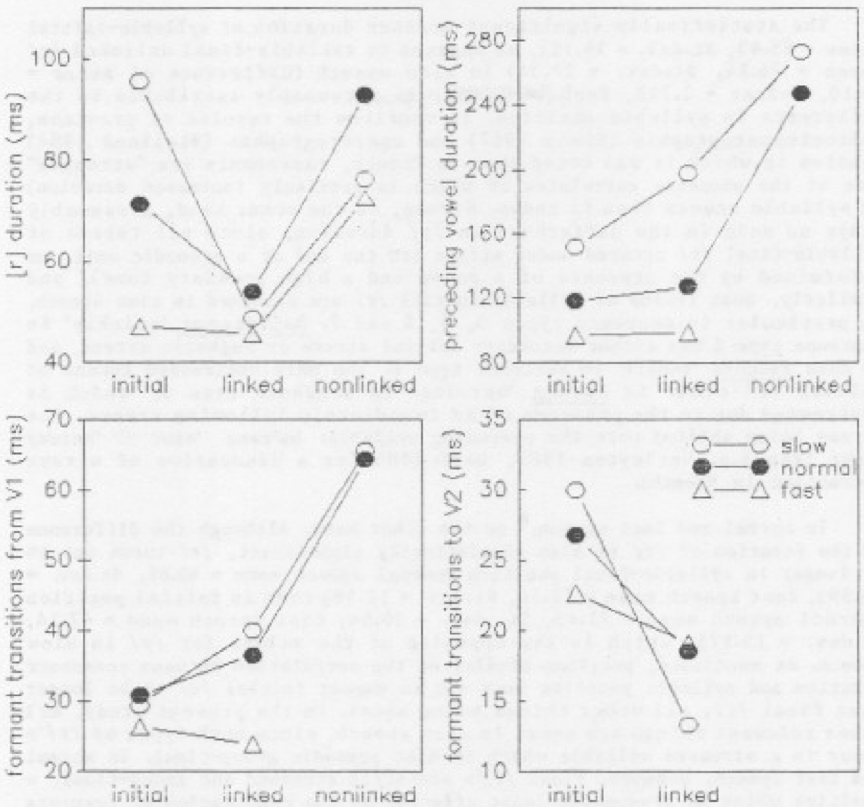


Fig. 6: Duration measurements (in ms) in initial, linked, and non-linked position for each tempo.

duration, although less pronounced, is statistically significant even in sentence type 1, where the final syllables of premier [prəmje] 'first, masc.', premiere [prəmjɛ:r] 'first, fem.' are unstressed, that is not followed by a high boundary tone: e.g., slow speech mean I = 151.00 (St.dev. = 27.16), mean N = 215.71 (St.dev. = 60.03), T-stat. = -2.668, Prob. = 0.0294.

The significant difference in the duration of FV1 is in accordance with the results of other studies which have revealed a correlation between length of formant transitions and syllabic structure (cf., e.g., Malmberg's 1955 speech synthesis experiment which revealed that for a sequence 'aga' to be percieved as 'ag-a', the necessary conditions are (1) significantly longer formant transitions from V1 to the consonant, and (2) the presence of a stop closure no shorter than 40 msec.).

The statistically significant greater duration of syllable-initial (mean = 95.43, St.dev. = 34.12), as opposed to syllable-final unlinked /r/ (mean = 76.24, St.dev. = 27.34) in slow speech (difference of means = 19.19, T-Stat = 1.718, Prob. = 0.109), is presumably ascribable to the difference in syllabic position. It confirms the results of previous, radiocinematographic (Simon 1967) and spectrographic (Rialland 1984) studies in which it was found that in French, consonants are "stronger" (one of the phonetic correlates of which is precisely increased duration) in syllable onsets than in codas. Stress, on the other hand, presumably plays no role in the difference in /r/ duration, since all tokens of syllable-final /r/ occured under stress (at the end of a prosodic unit, as determined by the presence of a pause and a high boundary tone), and similarly, most tokens of syllable-initial /r/ are stressed in slow speech, in particular in sentence types 3, 4, 6 and 7. Rapidement 'quickly' in sentence type 2 has either secondary initial stress or emphatic stress, and so does raccord 'repair' in sentence type 1. The only unstressed tokens of initial /r/ occur in hareng 'herring' in sentence type 5, which is destressed due to the presence of an immediately following stress, its stress being shifted onto the preceding syllable: ha'reng 'saur -> 'hareng 'saur (see e.g. Verluyten 1982, Dell 1984 for a discussion of stress retraction in French).

In normal and fast speech,[6] on the other hand, although the difference in the duration of /r/ is also statistically significant, /r/ turns out to be longer in syllable-final position (normal speech mean = 92.88, St.dev. = 23.59), fast speech mean = 72.50, St.dev. = 12.58) than in initial position (normal speech mean = 71.45, St. dev. = 30.84; fast speech mean = 47.14, St.dev. = 15.17), which is the opposite of the values for /r/ in slow speech. As mentioned, previous studies on the correlation between consonant duration and syllabic position lead one to expect initial /r/ to be longer than final /r/, all other things being equal. In the present study, all other relevant things are equal in slow speech, since both types of /r/'s occur in a stressed syllable which is also prosodic group-final. In normal and fast speech, however, final /r/'s are still stressed and group-final, a position which is presumably least affected by the compression of segments due to the increase in tempo, whereas initial /r/'s are stressed and occur in the group-final syllable only in the tokens in which prosodic restructuring did not take place.

3.4.2. Initial versus Linked /r/

3.4.2.1. Slow Speech

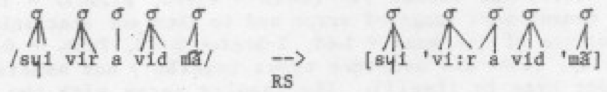
At first glance, the slow speech results seem contrary to the ones expected if linking in French corresponds to total resyllabification as is generally claimed (e.g., Delattre 1951:67). However, upon a closer examination, the results turn out to be, on the contrary, very interesting and remarkably consistent. What they in fact suggest is that in slower speech, linking does not correspond to total resyllabification whereby the

----

[6]It is doubtful that the sample means of the few observations of unlinked consonants in fast speech (5 observations, compared to 70 for initial /r/), represent the true means of the population. For the sake of the argument, we will however assume that they do come close.

consonant in question changes its affiliation from the preceding to the following syllable. Instead only partial resyllabification takes place: the consonant is associated with the following syllable, without, however, giving up its affiliation with the preceding syllable. The result is an ambisyllabic segment which belongs concurrently to both syllables. The process is shown in (4), where ' represents stress and RS stands for resyllabification.

(4) Initial syllabification   Slow speech syllabification



Supposing this to represent the facts, we would expect (1) the preceding vowel to be significantly longer before linked than before initial /r/, since the former is still tautosyllabic with the lengthening consonant, and (2) the formant transitions leading to V1 to be significantly longer before a linked consonant which occurs in the same syllable than before an initial consonant which occurs in the next syllable or word. Both of these expectations turn out in fact to be true for the three sentence types considered for V1 and FV1 duration measurements: difference of mean V1 durations = 45.46 (T-stat. = -1.926, Prob. = 0.077), difference of mean FV1 durations = 16.67 (T-stat. = 2.017, Prob. = 0,049). This is further supported by a difference in the realization of linked, as opposed to initial /r/. As was noted in section 5.3. above, the majority of the former's realizations are voiced fricatives whereas the latter's are voiceless (see Fig.5a). It is in line with characterizations of ambisyllabic segments as articulatorily as well as functionally "weak" segments, as witnessed by the contextual/assimilatory weakening they tend to undergo, especially in unstressed intervocalic context : cf., for instance, the flapping of ambisyllabic /t/ in American English lattitude, later, latter, etc.

The statistically significant difference in /r/ durations, as well as in the length of the formant transitions towards V2, with initial /r/ being significantly longer, and FV2 being significantly longer after initial /r/, is presumably due to stress. In most sentence types, initial /r/ and the following vowel are prosodic-group final in slow speech (cf. déstructurer, barons, Thierry in respectively 3, 4/6, and 7), whereas linked ambisyllabic /r/ is partly stressed (the final portion) and partly unstressed (the initial portion), and the following vowel is unstressed (structures et, bars ont, vieillard en).

Interestingly enough, there exists an older oscillographic study by M. Durand (1936) which compares masculine (i.e. liaison) and feminine (i.e. enchaînement) consonants in adjective-noun pairs like mauvais état [movɛzeta] 'bad state', mauvaise épée [movɛzepe] 'bad sword', or petit orage [pətitɔraz] 'small storm', petite orange [pətitɔrɑ̃ʒ] with respect to oral air pressure and glottal tension. Although Durand herself cautions against extrapolating too much from her experimental results,[7] she does conclude, based on the decrease of pressure on the feminine enchaînement consonant that it might retain some of its "implosive", that is, syllable-final, nature. In other words, she seems to suggest that it might be ambisyllabic, as opposed to the masculine liaison consonant, the increasing pressure of which suggests that it is truly initial. Cf. also Boudreault's (1868:45) observation that enchaînement consonants, especially sonorants, are strongly coarticulated with the preceding vowel, although perceptually, one has the feeling that they also partly belong to the following vowel.

5.4.2.2. Fast Speech

In fast speech, the difference in duration between initial (mean = 47.14, St.dev. = 15.17) and linked /r/ (mean = 45.46, St.dev. = 13.74) falls within the measurement range of error and is thus not statistically significant (difference of the mean = 1.68, T-Stat.= 0.673, Prob. = 0.502), both taken globally, (i.e. all sentence types together) and separately (i.e. each sentence type by itself). The results agree with the ones reported in Delattre 1981 who compared, among other things, word-final and word-initial /n,l,s/ in pairs like la masse agrée vs. le mât sacré, la ville imite vs. la vie limite, une avale vs. l'u nasale, etc. He notes that linking and initial consonants prove to be practically of equal length (final /n,l,s/ were 82, 76 and 135 msec.; initial ones 82, 74 and 134 msec., respectively).

Furthermore, the realizations of initial and linked /r/ are very similar, as Fig. 5c above suggests: both variants have a majority of voiced fricative realizations.

The difference in duration of the preceding vowel is likewise not statistically significant: I mean = 96.92, St.dev. = 28.73; L mean = 98.50, St.dev. = 29.93; T-Stat. = 0.312, Prob. = 0.756.

The formant transitions leading from initial /r/ to the preceding vowel (mean = 26.39, St.dev. = 11.30) are not significantly different from the ones leading from linked /r/ to that vowel (mean = 23.88, St.dev. = 9.91; difference of means = -2.51, T-Stat.= 1.288, Prob. = 0.2004). This is true globally for sentence types 1, 2 and 5, and when each of these sentence types is taken separately. Likewise, there is no difference in the formant transitions leading from initial and linked /r/ to the following vowel.

Furthermore, a comparison of the length of the transitions leading to V2 with the ones leading to V1 proves them to be very similar in duration (the difference of the means is 3.75 msec. in the case of initial consonants and 4.71 msec. in the case of linked consonants, which lies well within the range of error).

Combined, these five phonetic cues suggest therefore that in fast speech initial and linked /r/ are identical in terms of syllabic structure. In other words, linking corresponds to total resyllabification which can be represented as shown in (5).

----

[7] The results were not unambiguously clear, due in part to the experimental conditions, namely the inconvenience to the subjects brought about by the complicated and elaborate machinery.

(5) Initial syllabification — Fast speech syllabification

σ σ σ σ σ — σ σ σ σ σ

/sɥi vir a vid mɑ̃/ $\xrightarrow[RS]{}$ [sɥi vi ra vid 'mɑ̃]

With the association of the consonant with the following syllable there is a concomitant dissociation from its original syllable, which is in accordance with the generally made claim about French linking (e.g., Delattre 1951:67).

5.4.2.3. Normal Rate of Speech

In "normal" tempo, on the other hand, the global results are somewhat mixed, due to more pronounced inter-speaker variation in tempo and style. The male speaker's renditions are noticeably faster and more casual than the female speaker's whose pronunciation is more monitored and therefore also slower. It is thus useful to examine, in addition to the global results given in Fig. 6 above, the tokens of each speaker separately. These are represented in Fig. 7.

The difference in the duration of /r/ (with linked /r/ being on the average 18 msec. shorter than initial /r/) shows up as statistically significant when the results are considered globally. Considered separately, however, the difference in /r/ duration is significant only for the female speaker, whereas the male speaker's initial /r/'s are not significantly different from his linked /r/'s (Fig. 7a).

An inter-speaker difference is also noticeable in the realization of /r/. As noted in section 5.3. above, most unlinked stronger final /r/'s are due to the female speaker's slower and more monitored renditions, whereas linked /r/'s are primarily found in the male speaker's tokens with a majority of voiced fricative realizations similar to the unstressed initial /r/ renditions.

The difference in the duration of the preceding vowel, as well as in the length of the formant transitions towards V1 is not statistically significant globally in all seven sentence types taken together, as well as in sentence types 1, 2, 3 and 5 taken separately. As for the remaining three sentence types, either the difference in the duration of V1 is significant (sentence type 4 and 7, with only two observations before linking /r/ for the latter), or it is the length of the formant transitions towards V1 (sentence type 6). Taken separately, however, the V1 duration results for each speaker are as shown in Fig. 7b. Again, the difference in the means is statistically significant for the female speaker's tokens only. Similarly, for the formant transitions towards V1 which are shown in Fig. 7c.

The difference in the duration of the formant transitions towards V2 turns out to be statistically significant globally, as well as for each speaker separately, with the formant transitions following initial /r/ longer than the ones following linked /r/. It is however valid only for sentence types 3, 4, 6 and 7, where it is obviously again attributable to the presence of stress and a following prosodic group boundary in most

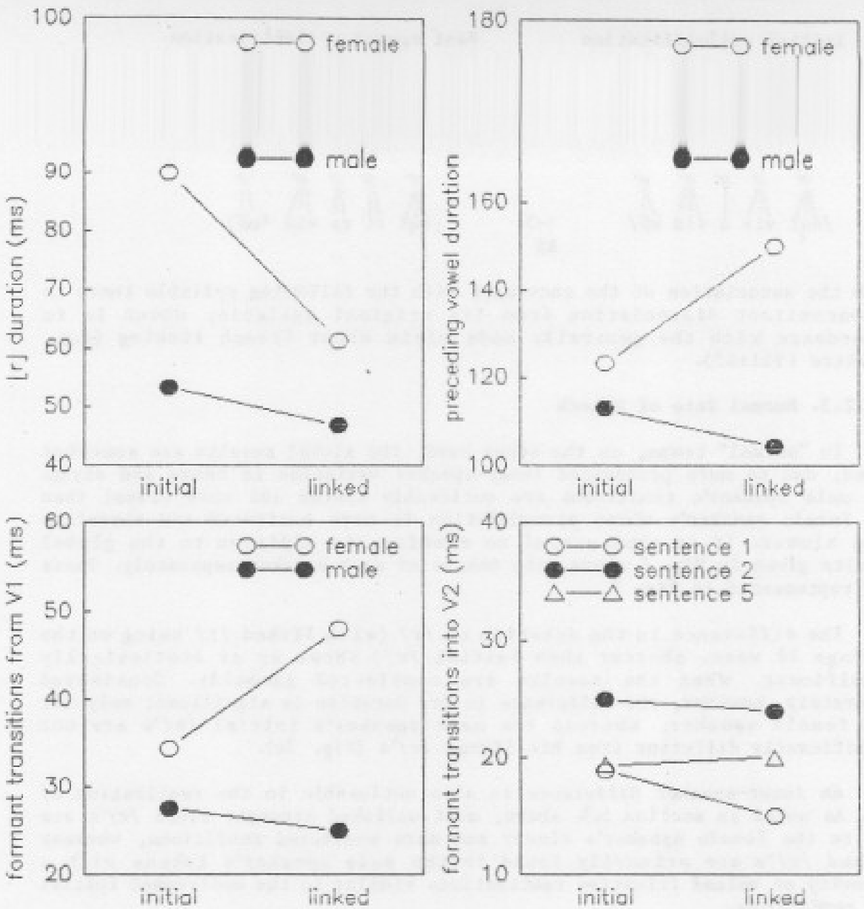


Fig. 7: Duration measurements at normal rate of speech for each speaker in initial and linked context.

tokens with initial /r/. In sentence types 1, 2, and 5, on the other hand, there is no difference in stress in the vowels following initial and linked /r/, and the difference in the duration is not statistically significant (Fig. 7d).

Finally, notice that a global comparison of the length of the formant transitions leading to the following vowel with the ones leading to the preceding vowel reveals them to be relatively similar. In the case of initial /r/ the difference amounts to 4 msec. which is well within the range of error. In the case of linked /r/ the difference is 18 msec. which exceeds the range of error. An examination of the values for each speaker separately, however, reveals this to be due to the female speaker whose transitions towards V1 exceed the ones towards V2 by 27 msec. In the male subject's renditions, on the other hand, the transitions to both vowels have quasi-equal lengths as shown in Table III.

Table III

Mean Duration of Formant Transitions (in msec.)

		FV1	FV2	FV1-FV2
F	I	34.33	27.61	6.72
	L	45.91	19.20	26.71
M	I	27.50	30.12	-2.62
	L	27.04	17.90	9.14

It is interesting to compare these results with the ones from Delattre's (1965:36ff.) investigation, by means of spectrograms and speech synthesis of so-called "internal juncture" and its perception in normal rate of speech of English, German, French and Spanish phrases of the type <u>an ice man</u> vs. <u>a nice man</u>, <u>zum einen</u> 'for one' vs. <u>zu meinen</u> 'to mean', <u>du nôtre</u> 'of ours' vs. <u>d'une autre</u> 'of another', <u>en ojo</u> 'an eye' vs. <u>enojo</u> 'anger'. He found that the main factor corresponding to a difference in syllabification between the four languages lies in "the degrees of arresting and releasing" of the formant transitions preceding and following the consonant closure, and reflecting, respectively, the closing and opening of the articulators. English and German showed maximally arresting consonant transitions before the closure standing in a three to one ratio with the releasing transitions, which is perceived as closed syllabification. In Spanish, the arresting and releasing transitions proved to be somewhat equal, and in French, the arresting transitions turned out to be very weak (up to one third of the releasing ones), and the releasing transitions maximal. The Spanish and the French state of affairs are both perceived as open syllabification. Hence, contrary to English and German, internal juncture is not distinctive in Spanish and French. In the present study, the comparison of the length of the transitions to either vowel resembles more Delattre's results from Spanish than his results from French, although the implications with respect to syllabification are the same.

The overall results in normal speech for each speaker separately are summarized in Table IV , where + stands for statistically significant and - for not significant.

Table IV

Statistical Significance of the Difference
in the Duration Means for each Speaker

	Speaker F	Speaker M
1) /r/	+	-
2) V1	+	-
3) VF1	+	-
4) VF2	-	-

As Table IV shows, the results for the female speaker are very similar to the results for slow speech. The statistical significance of the difference in the duration of /r/, V1 and in the length of the formant transitions towards V1 suggest that she preserved the same syllabic structure for linked /r/ in normal speaking rate as she had in slow speech: linking amounts to partial resyllabification of the consonant with the following vowel, with preservation of the affiliation with the preceding vowel. The male speaker's renditions, on the other hand, suggest that his linked /r/'s undergo total resyllabification, that is, dissociation from the preceding vowel and exclusive reassociation with the following vowel, which is similar to the results from fast speech discussed above.

The difference in tempo between the two speakers, signaled by a difference in the incidence of restructuring (cf. section 5.2. above), in the different lengths of the pauses between unlinked /r/ and following initial vowels (mean F = 326.00, mean M = 204.52), and in a difference in the compression of segments in faster rates of articulation, thus manifests itself also in a difference in phrasal syllabification.

6. Conclusion

Based on a bottom limit of at least 70%, linking can be said to occur in slow speech only within a prosodic unit, across the weakest type of syntactic boundary (represented in Nespor and Vogel 1982 as accentual phrase-internal contexts, and in Selkirk 1984 by a single silent grid position, as opposed to more loosely connected words, separated by prosodic boundaries in the former, and two or more silent grid positions in the latter). Crucially, in slow speech linking corresponds only to partial resyllabification, which leads to ambisyllabic, rather than fully initial linking consonants, as witnessed in particular by the duration of the preceding vowel and the formant transitions leading to it. As speaking rate increases, enchainement is extended to successively greater domains and corresponds to total resyllabification. More specifically, it is extended in normal speech to contexts with two, three and even one sentence type with four silent grid positions between the focus and the determinant words. And in fast speech it applies in all contexts tested, including across non-restructured breath-groups with five and possibly more silent grid positions between them.

Hence only the model outlined in Selkirk 1984 makes the right predictions, namely, that (1) with increasing rate of speech, the domain of the rule becomes larger, irrespective of prosodic structure, and (2) there exists a specific degree of disjuncture beyond which the rule does not apply in a given tempo.

Kaisse's model, on the other hand, does not correctly represent the facts of French enchaînement. According to this model, the rule belongs to so-called fast speech rules and should, therefore, not apply at all in slow speech, but take place across the board in normal and particularly in faster speech, irrespective of syntactic and prosodic structure.

Nor does the model proposed in Nespor and Vogel 1982 correctly represent the facts, since the domain is predicted to remain constant across different speaking rates, except for cases of restructuring which combine two prosodic units into one. As described in section 5.2. above, such restructuring took place only in a relatively small number of tokens with linked /r/, and it does not represent a conditioning factor for the application of resyllabification.

The study thus provides evidence that, at least in French, resyllabification belongs to the phonological rules proper, and not to the rules for defining (post-lexical) phonological representation, that is, the syntax-phonology mapping rules. Like purely phrasal phonological rules, it is determined by the syntactic timing (that is, the rhythmic structure) of the utterance, and thus applies to a phonological representation fully defined. A comparison with phrasal resylabification rules in other languages is however needed before any generalizations can be made.

References

Borel-Maisonny, S., 1942. Les divers aspects de l'r parisien. Le Français Moderne 10:219-231.

Boudreault, l'Abbé M., 1968. Rythme et mélodie de la phrase parlée en France et au Québec. Quebec and Paris: Klincksieck.

Cinque, G., 1977. The Movement Nature of Left Dislocation. Linguistic Inquiry 8:397-411.

Clements, G. and S. Keyser, 1983. CV Phonology. Cambridge, MA: Mit Press.

Delattre, P., 1981. Consonant Gemination in Four Languages: An Acoustic, Perceptual and Radiographic Study. In Delattre, P. (ed.), Studies in Comparative Phonetics. Heidelberg: Julius Groos Verlag. Pp. 95-133.

-----, 1966. Les dix intonations de base du français. The French Review 40:1-14.

-----, 1965. Comparing the Prosodic Features of English, German, Spanish and French. In Delattre, P., Comparing the Features of English, French, German and Spanish. Heidelberg: Julius Groos Verlag. Pp. 23-43.

-----, 1951. Principes de phonétique française à l'usage des étudiants anglo-americains. Middlebury, VT: Middlebury College.

-----, 1940. Le mot est-il une entité phonétique en français? Le Francais Moderne 8:47-56.

Dell, F., 1984. L'accentuation dans les phrases en français. In Dell, F., D. Hirst and J.-R. Vergnaud (eds.), Forme Sonore du langage. Hermann: Paris. Pp. 65-122.

Di Cristo, A. 1976. Indices prosodiques et structure constituante. Cahiers de Linguistique, d'Orientalisme et de Slavistique 7:27-40.

Delgutte, B., 1978. Technique for the perceptual investigation of Fo contours with application to French. Journal of the Acoustical Society of America, 64(5): 1319-32.

Durand, M., 1939. Etude expérimentale sur la durée des consonnes parisiennes. Bibliothèque du Français Moderne: Paris.

-----, 1936. Le genre grammatical en français parlé à Paris et dans la région parisienne. Paris: Bibliothèque du Français Moderne.

Encrevé, P., 1983. La liaison sans enchaînement. Actes de la Recherche en Sciences Sociales 46. Paris: Minuit.

Gårding, E., 1967. Internal Juncture in Swedish. Lund University Working Papers 6. Lund: Gleerup.

Grammont, M., 1933. Traite de phonétique. Paris: Delagrave.

Kaisse, E.M., 1985. Connected Speech. The Interaction of Syntax and Phonology. New York: Academic Press.

Laeufer, C., in preparation. The Influence of Speaking Tempo and Position on the Realization of French /r/.

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

Léon, P.R., 1971. La joncture externe en français: nature et fonctions linguistiques et expressives. Studia Phonetica 4:57-66.

-----, 1966. Prononciation du français standard. Paris: Didier.

Malmberg, B., 1964. Juncture and Syllable Division. In Abercrombie, D. et al. (eds.). In Honour of D. Jones. London: Longmans. Pp. 116-119.

-----, 1955. The Phonetic Basis for Syllable Division. Studica Linguistica 9:80-7.

Martin, P., 1982. Phonetic realizations of prosodic contours in French. Speech Communication 1:283-94.

-----, 1980. Sur les principes d'une théorie syntaxique de l'intonation. Studia Phonetica 17: 234-271.

Nespor, M. and I. Vogel, 1982. Prosodic Domains of External Sandhi Rules. In Van der Hulst, H. and N. Smith (eds.) The Structure of Phonological Representations (Part I), 225-256. Dordrecht: Foris.

O'Shaughnesssy, D., 1981. A Study of French Vowel and Consonant Durations. Journal of Phonetics 9:385-406.

Pernot, N., 1937. La Liaison en français, liaison et enchaînement. The Modern Language Journal 22:333-8.

Pickett, J.M., 1980. The Sounds of Speech Communication. Baltimore: University Park Press.

Pulgram, E., 1970. Syllable, Word, Nexus, Cursus. The Hague: Mouton.

-----, 1965. Prosodic Systems: French. Lingua 13:125-44.

Rialland, A., 1984. Schwa et syllable en français. In Wetzels, L. and Sezer, E. (eds.). Studies in Compensatory Lengthening. Dordrecht: Foris.

Rossi, M., 1980. Le français, langue sans accent? In Fonagy I. and P.R. León, L'accent en français contemporain. Studia Phonetica 15: 13-51.

Rossi, M., Di Cristo, A., Hirst, D., Martin, P. and Nishinuma, Y., 1981. L'intonation: de l'acousique à la sémantique. Paris: Klincksieck.

Selkirk, E.O., 1984. Phonology and Syntax: The Relation between Sound and Structure. Cambridge, Ma: The MIT Press.

Simon, P., 1967. Les consonnes françaises. Mouvements et positions articulatoires à la lumière de la radio-cinématographie. Paris: Klincksieck.

Straka, G., 1965. Contribution à l'histoire de la consonne r en français. Neuphilologische Mitteilungen 66:572-606.

Vaissière, J., 1983. Language-Independent Prosodic Features. In Cutler, A. and D.R. Ladd (eds.), Prosody: Models and Measurements. Berlin: Springer Verlag. Pp. 53-66.

-----, 1980. La Structuration acoustique de la phrase française. Annali della Scola normale superiore di Pisa, Ser. III, 10:529-560.

-----, 1975. Further note on French Prosody. Res. Lab. Electr., Quarterly Progress Report, MIT, 115:251-61.

-----, 1974. On French Prosody. Res. Lab. Electr., Quarterly Progress Report, MIT, 114:212-23.

Verluyten, S.P., 1982. Investigations on French Prosodics and Metrics. Ph.D. Dissertation, University of Antwerp.

Vidon-Varney, J., 1933. Pronunciation of French. Ann Arbor: Edwards Brothers.

Appendix

Corpus used in the experiment

1 a. Nous avons atteint la premiere accore à midi.
'We reached the first reef at noon.'

b. Nous avons fini le premier raccord à midi.
'We finished the first splicing at noon.'

2 a. Ils nous suivirent avidement.
'They followed us eagerly.'

b. Il nous suivit rapidement.
'He followed us quickly.'

3 a. Il faut des structures et des phrases, pour pouvoir se faire une opinion.
'Structures and phrases are needed to be able to form an opinion.'

b. Il faut déstructurer des phrases, pour pouvoir se faire une opinion.
'It is necessary to unstructure phrases in order to be able to form an opinion.'

4 a. Les bars ont perdu des clients.
'The bars have lost clients.'

b. Les barons perdaient des clients.
'The barons were losing clients.'

5 a. Un vieillard en sort, et beaucoup plus tard, une femme.
'An old man comes out (of it), and much later, a woman.

b. Un vieil hareng saur est bien meilleur mariné.
'An old sour herring is much better marinated.'

6 a. Et tous les nouveaux bars ont perdu des clients.
'And all the new bars have lost clients.'

b. Et tous les nouveaux barons perdaient des clients.
'And all the new barons were losing clients.'

7 a. Quant a Jean-Pierre, Irene ne l'a pas vu.
'As for Jean-Pierre, Irene has not seen him.'

b. Quant a Thierry, Reine ne l'a pas vu.
'As for Thierry, Reine has not seen him.'

OSU #36: Papers from the Linguistics Laboratory 1985-1987, Sept. 1987.

**Phonemic Quantity, Stress,
and the Half-Long Vowel in Finnish**

Riitta Välimaa-Blum
Linguistics, Ohio State University

Abstract: An experiment was conducted to compare the duration of the final vowels in Finnish disyllabic words which have short stressed initial syllables - CVCV(V) - with that in words with long stressed initial syllables - CVVCV(V), CVCCV(V), CVVCCV(V). It was found that in the first group the final short and long vowels were systematically longer than in the second group. This is taken to suggest that stress interacts with phonemic quantity: on the initial syllable in the CVCV(V)-words there is a conflict between the phonemic duration of the vowel and the use of duration as a stress cue. This is resolved by spreading the stress-induced duration to the second syllable. Thus even if the domain of stress is a syllable, its realization is not independent of the rest of the phonological structure but it operates in concert with it.

1. Introduction

The representation of time has long been a problem in the phonological description of speech. One argument, for example, concerns whether long segments are 'long' or whether they are clusters of two short segments. There is no single solution to this question, but the decision seems to depend on language-particular facts. Apart from phonemic quantity distinctions, there is also allophonic variation in terms of duration. The English lengthening of vowels before voiced consonants is a well-known example of this kind of alternation. This lengthening is not predicted from 'universal phonetics' but is a language-specific phenomenon. This kind of allophonic behavior can be present even in a language which has distinctive quantity. For example, Finnish has a 'half-long' vowel which occurs as the second vowel in CVCV-structures. It is about 1.5 times longer than the first, stressed vowel. It is also significantly longer than corresponding vowels in words with long initial syllables. Also long final vowels are significantly longer after short initial syllables than after corresponding long syllables, and geminates, too, tend to be longer after short stressed vowels than after long stressed vowels. It seems that these lengthening phenomena relate to the interaction of phonemic vowel quantity and stress on the first syllable. If so, then Finnish would seem to illustrate why Jakobson and Trubetzkoy suggested that no language can have both distinctive stress and distinctive quantity.

Stress serves a demarcative function in Finnish. Primary word stress is on the first syllable of the word and after that there are secondary stresses on roughly each odd-numbered non-final syllable (Karlsson, 1982, 150). Vowels and consonants have a phonemic quantity opposition and this opposition is present in both stressed and unstresssed syllables. Thus a stressed syllable may contain either a short or a long vowel. A syllable may have any one of the following structures:

(1) a. (C)V b. (C)VV c. (C)VC d. (C)VVC e. (C)VCC

I will call (1a) a short syllable and all the others long. On the average a long vowel is twice as long as the corresponding short vowel (Wiik & Lehiste, 1968; Lehtonen, 1970) and stressed vowels are longer than corresponding unstressed vowels. There is one exception to this last generalization, however: the half-long vowel (Wiik & Lehiste, 1968; Lehtonen, 1970), which is phonemically short. Thus, in the following word the second, unstressed [a] is significantly longer than the first, stressed [a]:

(2) kala 'fish'

Wiik and Lehiste also note that "the long unstressed vowels are longer when preceded by a short stressed vowel than when preceded by a long stressed vowel" (1968, 570). Actually the final vowel is shorter after any stressed long syllable, not just one containing a long vowel (Lehtonen, 1970, 126). Karlsson states that Finnish has an allophonic rule that lengthens a short vowel after an initial CV-syllable (1982, 151). He gives the following rule to account for this (ibid., 152):

/V/ ---> [longer] / # CV.(C)___
Note: as compared to the context $\# \, CV \begin{Bmatrix} V \\ C \end{Bmatrix} . \, (C)$ ___

The note is added because, he explains, we cannot state durational relations in absolute terms. None of the above authors associates the two lengthenings, i.e., those of short and long vowels, to the same underlying cause. Nor is any explanation offered for the lengthening of the long vowel. The geminate lengthening is also usually seen as a separate, unexplained phenomenon (Karlsson 1982); note though Lehtonen (1970, 152).

The vowel phenomenon is not a case of final lengthening because the final vowels, short and long, are lengthened only after a short initial syllable and not after corresponding long syllables. That is, the final vowel in CVCV(V)-words is longer than the final vowel in, e.g., CVVCV(V), CVCCV(V) and CVVCCV(V)-words. We can thus note two facts here:

(a) the second, unstressed vowel in CVCV words is longer than the stressed first vowel even if stressed vowels are generally longer than corresponding unstressed vowels.

(b) both long and short second vowels are on the average considerably longer after a short first syllable than after a long initial syllable.

Wiik and Lehiste (1968) explain the half-long vowel as follows: the second vowel is lengthened in order to create a certain ratio between the vowels thereby signalling the phonemic quantities of these two and at the same time establishing a particular word structure, i.e., CVCV. This takes place within a two-syllable sequence which is taken to be the fundamental prosodic unit in Finnish. Also Karlsson notes that the realization of vowel quantity depends on the structure of the whole word as is implicit in his rule above (1982, 72). Lehtonen presents similar ideas (1970). If it is true that the second vowel is lengthened in order to create a certain ratio which is needed for the identification of the vowel durations, then the lengthening to half-long serves a distinctive function at the disyllabic level. This proposal, however, does not include any explanation for why the ratio between the vowels in the CVCV-words without the half-long vowel could not serve this putative distinctive function.

Kohler has proposed a solution similar to the one above which would solve the problem mentioned: the syntagmatic relations between the vowels in CVCV and CVVCVV words must be different in order not to neutralize the two structures under various "higher-order timing levels of tempo, rhythm, and intonation" (1986: 269). This would be, according to Kohler, the motivation for the half-long vowel in Finnish with the resulting differences in the V1/V2-ratios in CVCV and CVVCVV words. However, this suggestion leaves unexplained why also the long vowel in CVCVV words gets lengthened. There is no corresponding word for CVCVV which might possibly have the same ratio between the vowels to form the analogous pair to the CVCV and CVVCVV pair. Also, as we will see below, this study does not provide unequivocal support for the proposal that the syntagmatic relations between the vowels indeed remain the same under various "higher-order timing levels."

Lehiste (personal communication) has offered another possible explanation which would solve the problems found above: there is a tendency to make all the basic prosodic units -- i.e., the disyllabic sequences -- isochronous. However, the data given in Lehtonen (1970, 126) do not support this view. The differences between the various disyllabic structures are considerable. Also, since the long final vowel in CVCVV is lengthened it would seem to go against this tendency. If the final vowel in CVCV structures is lengthened in order to create structures isochronous with, e.g., CVVCV, then the lengthening of the final VV in CVCVV is counterproductive.

Lehtonen explains the presence of the half-long vowel by proposing that the domain of stress in Finnish is a sequence of two moras (1970, 151). Word stress would have "some standard minimum duration or amount of energy whose domain is the two-mora group" (ibid.). This would explain the half-long vowel: "the articulation of segmental sounds should...be synchronized with this imagined energy pulse" (ibid.). And since the second mora in a CVCV-word falls on the second syllable, the half-long vowel is created. Lehtonen's explanation is very similar to mine (below) but for the notion that the domain of stress is two moras. First, it is not obvious to me why the second mora should become lengthened. Furthermore, since diphthongs do not behave like other two-mora sequences (in the following words the second vowel in the second word is half-long but not in the first: [tuo] 'bring!' vs. [ta.o] 'forge!'; see below) it seems that something other than a specific domain is more important. Lehtonen mentions that quantity distinctions need to be maintained, and I propose that this is the major factor in play. Stress simply makes segments longer but it does not have any specific two-mora domain.

The central fact here seems to be that the half-long and the lengthened long vowel both occur after a short initial syllable and not after long initial syllables; also geminates get longer after short stressed vowels than after long stressed vowels. Important, too, is that unstressed vowels are in general shorter than corresponding stressed vowels except in CVCV-words which again involves the short initial syllable.

If we consider what the phonetic correlates of stress are, an alternative to the above explanations may be found. The usual phonetic cues to stress are intensity, pitch and duration. A particular language may use these to a different degree each so that even though they all are actively employed, maybe only one gives the decisive cue (e.g., Lehiste, 1970, 138). Since Finnish has a quantity opposition we might not expect it to use the duration cue very much in signalling stress. But clearly it does use duration since stressed syllables are longer than the unstressed ones. If we assume that duration is one of the important stress cues in Finnish then we may understand the behavior of the vowels after stressed CV-syllables.

I propose that the exceptional length of the second syllable vowel after stressed CV-syllables results from a conflict between the phonemic quantity opposition at the segmental level and the use of duration as a stress cue at the suprasegmental level. Since stress would lengthen the vowel of a CV-syllable, this might create an unwarranted neutralization. To resolve this conflict between the use of duration at two different but simultaneous levels the lengthening as a signal of stress spreads over to the second syllable in a CVCV(V) word. Consequently, the final vowel is longer than it would be if the first syllable were long. This same explanation extends to the geminates.

Estonian, which also has initial stress and quantity opposition (on the first syllable), has a similar property as shown by Lehiste:

> ...the duration of the vowel on the second syllable, which is not independently contrastive, is inversely proportional to the quantity of the first syllable...This is especially clear in case of short first syllables, which are followed by unstressed syllables whose vowel is usually about 1.5 times as long as the short vowel of the stressed syllable. (1970, 50)

Also Hungarian, where the relevant characteristics are the same as in Finnish and Estonian, may have this same phenomenon. Fónagy reports on earlier research showing that unstressed syllables may be louder and longer than preceding stressed ones. The examples he gives are cases where the initial stressed syllable is short (1966, 234). The fact that these three languages share the lengthening may be suggestive of a general interaction between stress and phonemic quantity and not just a genetic characteristic of these particular Finno-Ugric languages.

This paper reports an experiment done in order to provide evidence in support of the prosposed interpretation of the half-long vowel. The main focus of the experiment was on the following:

(a) the durations of the final vowels after short and long stressed syllables will be compared

(b) the ratios between the vowels in disyllabic sequences with different word structures will be compared

A secondary aim was to see if also the geminates become longer under the various rates and discourse conditions as predicted by earlier findings. Essentially the experiment replicated some earlier studies but new variables like speech rate and sentence position/sentence stress were included. These are especially important if we want to determine whether the ratios can indeed serve a distinctive function; if they can, they should remain largely the same under all discourse conditions. I also expected that the final vowels are significantly longer after stressed short syllables than after stressed long syllables. Thus I did not expect my findings to be different from the earlier ones in this respect, but I wanted to see that the facts will be the same regardless of speech rate and various discourse contexts and that they will be the same for both short and long final vowels.

2. Methods

Lehtonen in the above study looked at words in naturally occurring sentences which were read in isolation and where the test-words were in neutral sentence stress positions. I replicated his study with some modifications. I had the words under consideration in two sentence positions, in three sentence stress conditions and in two speech rates. The sentence positions were initial and medial; the sentence stress conditions were initial neutral declarative sentence stress, initial contrastive stress, and medial neutral stress following contrastive stress; the speech rates were fast and normal. The complete corpus can be found in the appendix. About two thirds of the sentences (set I) were given a context in parentheses immediately before each sentence; the subjects were instructed to read aloud only the test sentence. The rest of the sentences (set II) had no context given. Each list of the test sentences was read six times in normal speech rate and six times "as fast as they could". Each reading of the list used a different random order of the test sentences; each set was printed on a separate sheet. The speech rates alternated in a random order. The test words contained only vowels and voiceless plosives; the word after the test word always began with a voiceless plosive. The CVCC syllable-type is excluded from the comparisons. Also, since the final consonant has no effect on the duration of the other segments in the word (Lehtonen, 1970), the words to be examined have open final syllables.

In the first part, the final vowels of the words were studied. In set I only vowels of the same vowel quality were compared, in set II it is always [a], short or long, that occurs at the end of a word. The comparisons shown in Table I were made.

Table I

Set I				Set II			
CVCV –	CVVCV	[kato] –	[kaato]	CVCV –	CVCVV	[tupa] –	[tupaa]
	CVCCV		[katto]				
	CVVCCV		[taatto]	CVCV –	CVVCVV	[tupa] –	[piikaa]
					CVCCVV		[takkaa]
CVCV –	CVVCCV	[piki] –	[piikki]		CVVCCVV		[taakkaa]
			[tuutti]				
	CVCCV		[tutti]	CVCVV –	CVVCVV	[tupaa] –	[piikaa]
					CVCCVV		[takkaa]
					CVVCCVV		[taakkaa]

For the geminates only three word pairs were examined: [katto] - [taatto], [tutti] - [tuutti] and [takkaa] - [taakkaa]. The measurements were made from the wave-form using the sfm wave-form editor on the New England Digital computer. The beginning and end of each vowel was measured from voice onset to subsequent voice offset. The consonant duration was measured from the end of voicing at the consonant implosion to the beginning of voicing for the next vowel.

The second part consisted of obtaining the V1/V2-ratios of each test word in each experimental condition in order to see, first of all, if the ratios are constant. Also it will have to be established if any particular discourse condition has a systematic effect on the ratios.

Three subjects were used, two females and one male. One of the females had lived all her life in Helsinki, the other most of her adult life also in Helsinki; the male subject grew up in central Finland but had lived over ten years in Helsinki. One female had been in the USA about half a year, the other female nine years, the male for 2.5 years. All subjects had university degrees ranging from BA to PhD.

3. Results

In comparing the average durations of the vowels which occurred after the initial CV-syllable with those occurring after CVV, CVC or CVVC-syllables it was found that the former were significantly longer across both the three sentence stress conditions and the two speech rates. This lengthening after a stressed CV was present in both short and long vowels. Figures 1 through 3 give the averages of the second syllable vowels in milliseconds, first in all conditions combined, then by rate and by emphatic conditions. The sets to be compared are:

(a) [kato] vs. [kaato], [katto], [taatto]

(b) [tupaa] vs. [piikaa], [takkaa], [taakkaa]

(c) [piki] vs. [piikki], [tutti], [tuutti].

The consonants preceding the vowels in (b) and (c) are different, but they are always voiceless plosives so that there should not be any great differential effects (the vowels are measured from the voice onset). The P-values in the figure captions were obtained from a two-tailed t-test, where the mean vowel duration after CV-syllables was compared with the combined mean value of the final vowels in the other three word structures in each comparison set.

We can note in figure 1 that the final vowel averages in the CVCV(V) structures are significantly different from the other three structures. The male speaker was a fast speaker even in his normal rate which shows in his mean values being smaller than those of the female speakers. In figure 2 we compare the averages by speech rate. The difference in rate did not have any effect on the durational differences, although, as expected, the averages were shorter in fast rate than in normal rate. In figure 3, likewise the various emphatic conditions do not seem to have any influence on the differences.

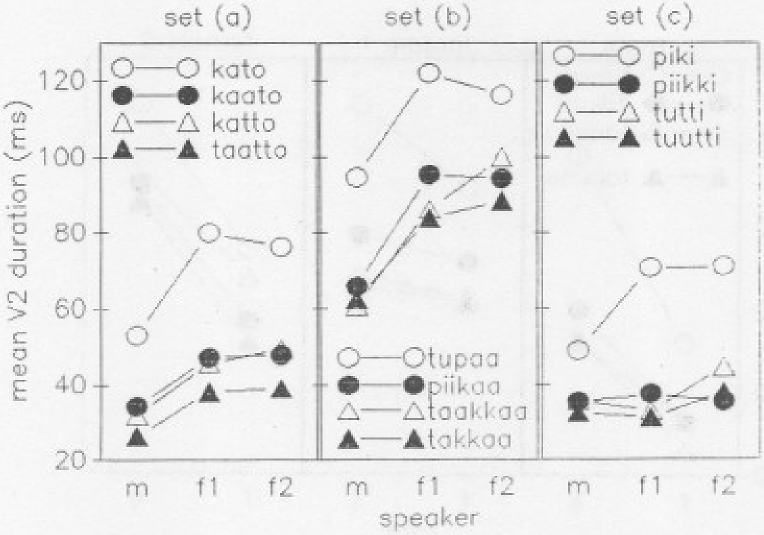


Fig. 1: Mean 2nd vowel durations for all conditions combined. Probability that V2 in CV__ context is the same as in other 3 words in each comparison set: $P < 0.0001$. Speakers: m = male, f1 = female 1, f2 = female 2.

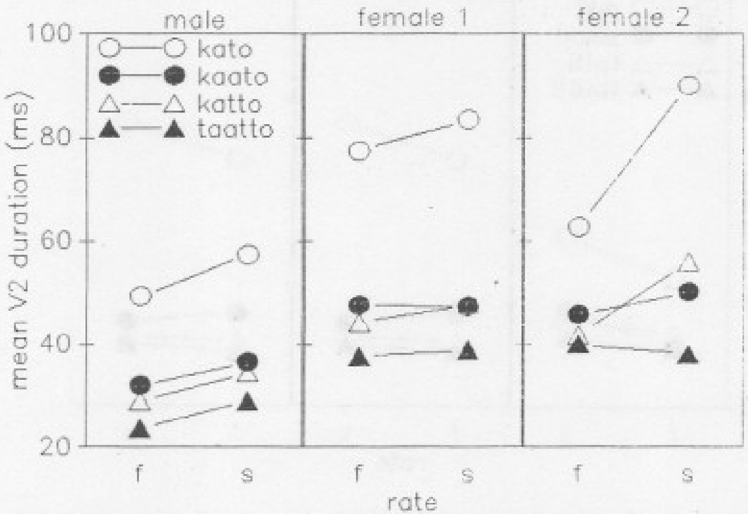


Fig. 2a: Mean second vowel durations for words in comparison set (a), for each speech rate (slow or fast) for each speaker. Probability that the vowel in kato is the same as in the other three words $P<0.0001$ for both speeds for all three subjects.

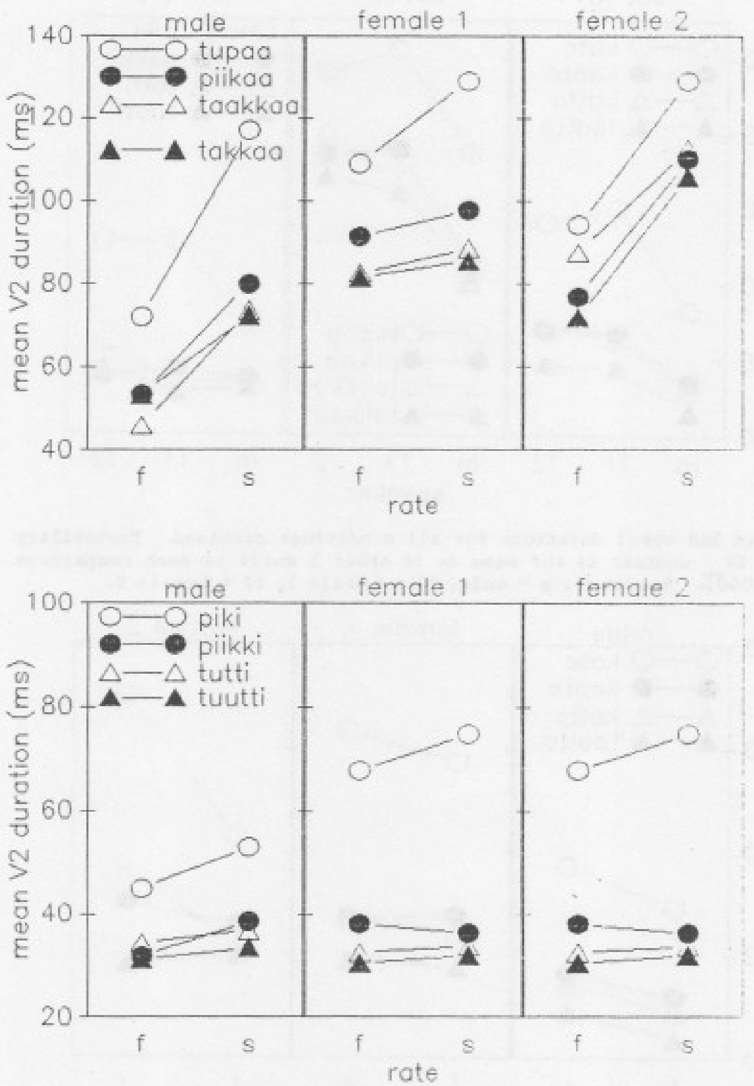


Fig. 2b-c: Mean second vowel durations for words in comparison sets (b) and (c), for each speech rate (slow or fast) for each speaker. Probability that the vowels in the CV__ context are the same as in the other 3 words in each set is P<0.0001 for both speeds for all 3 subjects, except P<0.005 for f2's tupaa at fast rate.

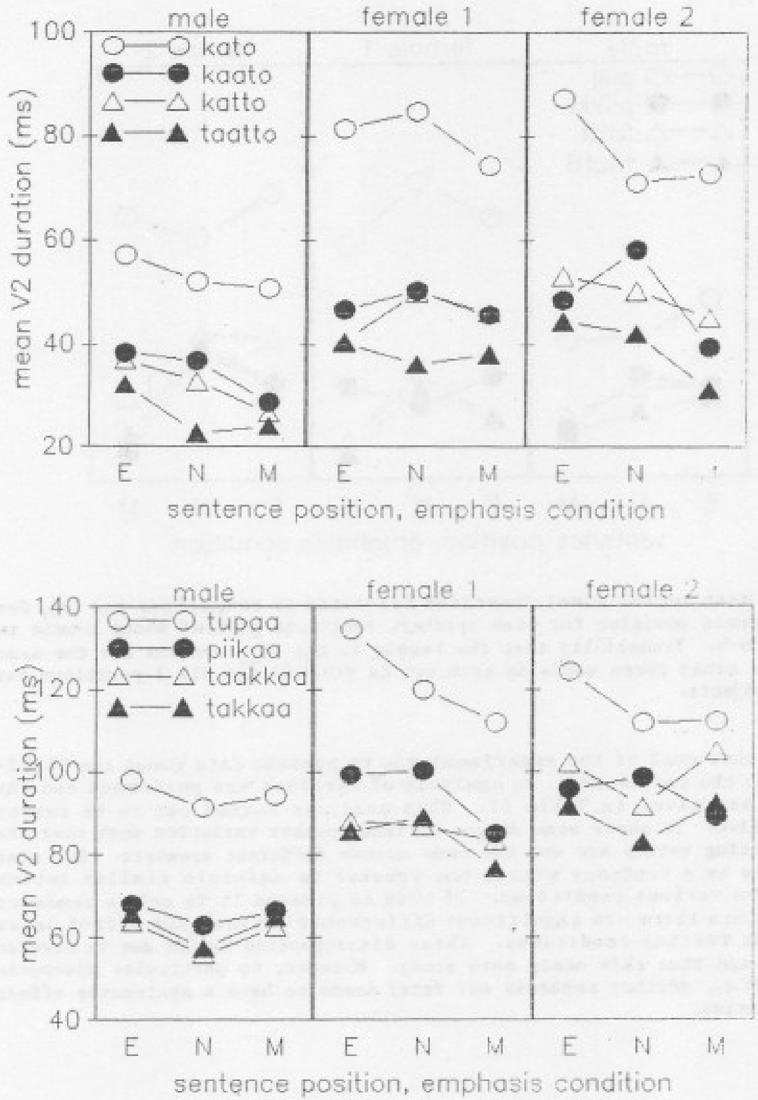


Fig. 3a-b: Mean second vowel durations for words in comparison sets (b) and (c), for each sentence position for each speaker. Positions plotted along x-axis: N = Neutral, initial sentence position, E = Emphatic, initial sentence position, M = Medial, post-emphatic position. Probability that the vowels in the CV__ context are the same as in the other three words in each set is P<0.001 for all 3 positions for all 3 subjects, except P<.02 for male and female 2 in set (b).

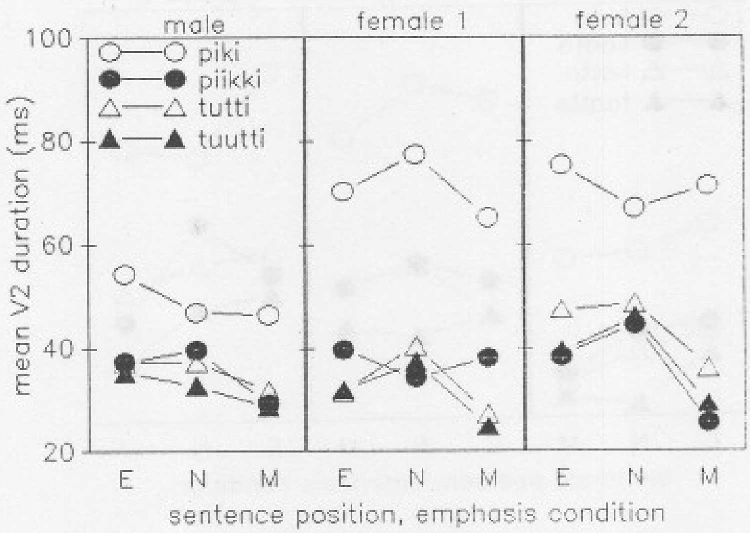


Fig. 3c: Mean second vowel durations for words in comparison set (c), for each sentence position for each speaker. Positions plotted along x-axis as in Fig. 3a-b. Probability that the vowels in the CV̄ context are the same as in the other three words in each set is P<0.0001 for all 3 positions for all 3 subjects.

Another goal of the experiment was to present data about the V1/V2-ratios in the test words. An analysis of variance was performed and the results are given in Table II. This analysis turned out to be rather inconclusive. It shows some degree of interspeaker variation such that the corresponding ratios are not the same across different speakers. But also there may be a tendency within one speaker to maintain similar ratios across the various conditions. If this is present it is only a tendency though since there are significant differences between the ratios under different testing conditions. These discrepancies may be due to diverse factors, and thus this needs more study. However, no particular discourse context (i.e., neither emphasis nor rate) seems to have a systematic effect on the ratios.

Table II.

V1/V2-Ratios by rate and emphatic conditions

a. Male

word		Rate normal	fast	P	Sentence Position Neutral	Emphatic	Medial	P
kato	x	1.08	.88	.002	.95	1.09	.90	.07
	sd	.16	.20		.24	.20	.14	
kaato	s	3.34	2.84	.41	2.97	2.68	3.54	.49
	sd	1.78	1.63		1.96	.64	2.06	
katto	x	2.10	2.20	.84	1.63	1.84	2.86	.08
	sd	.70	1.99		.75	.06	2.04	
taatto	x	4.28	3.71	.48	4.30	2.96	4.68	.17
	sd	1.81	2.63		2.22	.73	2.95	
tupa	x	.90	.91	.90	.86	.97	.87	.54
	sd	.21	.26		.16	.27	.25	
tupaa	x	.50	.63	.01	.53	.62	.53	.24
	sd	.12	.15		.19	.14	.10	
piikaa	x	1.16	1.22	.52	1.03	1.24	1.30	.02
	sd	.21	.27		.19	.18	.28	
takkaa	x	.90	1.17	.02	1.03	1.12	1.11	.77
	sd	.15	.44		.36	.45	.39	
taakkaa	x	1.47	1.56	.28	1.40	1.55	1.58	.16
	sd	.22	.27		.17	.23	.28	

b. Female 1

word		Rate normal	fast	P	Sentence Position Neutral	Emphatic	Medial	P
kato	x	.96	.97	.73	.94	1.03	.93	.007
	sd	.11	.08		.09	.09	.06	
kaato	x	3.35	2.78	.001	2.93	3.28	2.85	.15
	sd	.58	.39		.52	.74	.27	
katto	x	1.91	1.87	.07	1.71	2.14	1.80	.01
	sd	.31	.46		.27	.47	.32	
taatto	x	3.75	3.30	.008	3.56	3.64	3.50	.24
	sd	.49	.46		.62	.45	.46	
tupa	x	.76	.09	.06	.79	.76	.83	.39
	sd	.11	.08		.11	.11	.10	
tupaa	x	.56	.63	.03	.59	.52	.63	.02
	sd	.08	.09		.09	.07	.07	
piikaa	x	1.32	1.25	.14	1.24	1.34	1.31	.28
	sd	.15	.14		.19	.10	.11	
takkaa	x	.98	.99	.83	.92	1.02	1.01	.11
	sd	.13	.10		.10	.09	.13	
taakkaa	x	1.61	1.56	.33	1.56	1.61	1.60	.72
	sd	.14	.17		.11	.18	.16	

Table II. (cont.)

c. Female 2

word		Rate: normal	fast	P	Sentence Position: Neutral	Emphatic	Medial	P
kato	x	.76	.85	.02	.76	.85	.81	.23
	sd	.10	.12		.08	.10	.14	
kaato	x	3.07	2.26	.03	1.81	2.86	3.15	.01
	sd	1.05	1.04		.35	.96	1.29	
katto	x	1.49	1.52	.78	1.49	1.47	1.55	.86
	sd	.22	.47		.56	.16	.30	
taatto	x	3.53	2.82	.23	2.37	3.20	4.06	.05
	sd	.79	2.34		1.04	1.33	2.26	
tupa	x	.71	.69	.66	.70	.76	.62	.01
	sd	.08	.14		.11	.13	.08	
tupaa	x	.51	.52	.71	.47	.55	.52	.08
	sd	.07	.09		.06	.07	.08	
piikaa	x	1.07	1.17	.11	1.00	1.18	1.15	.03
	sd	.15	.17		.11	.14	.19	
takkaa	x	.66	.76	.05	.69	.79	.61	.02
	sd	.12	.15		.14	.13	.08	
taakkaa	x	1.12	1.07	.51	1.08	1.27	1.02	.01
	sd	.18	.21		.17	.25	.13	

A secondary aim of the study was to look at the geminate durations. They were generally longer after short stressed vowels than after long stressed vowels. These differences, however, were not as dramatic as those found in vowels. In some instances the differences were tendencies not seen in the averages. But the overall means do not take into account token variability in speech rate in that there is not a match between words produced in a certain rate in one set with words produced in the same rate in the other set. If we want to consider this variability in rate we can take the duration of the V1 as an indicator of the overall rate and plot the geminate duration against it; now differences become observable. In figure 4, the X-axes have been adjusted so that the minimum values of the two types are aligned. If we compare cases where the means are significantly different (e.g., *katto* versus *taatto* for female 1 in figure 4a) and where they are not significantly different (e.g., the same words for female 2 in figure 4b), we can see that the separation of the regression lines is the same in both.

4. Discussion

My proposal is that it is the conflict between the use of duration as a stress cue and the need to maintain phonemic quantity distinctions that is responsible for the lengthening of short and long vowels after short stressed syllables. Thus we could modify Karlsson's rule as follows to include also long vowels and to note that stress is relevant, not just the syllable structure:

/V(V)/ ---> [V(V):] / (C)V́.(C)___

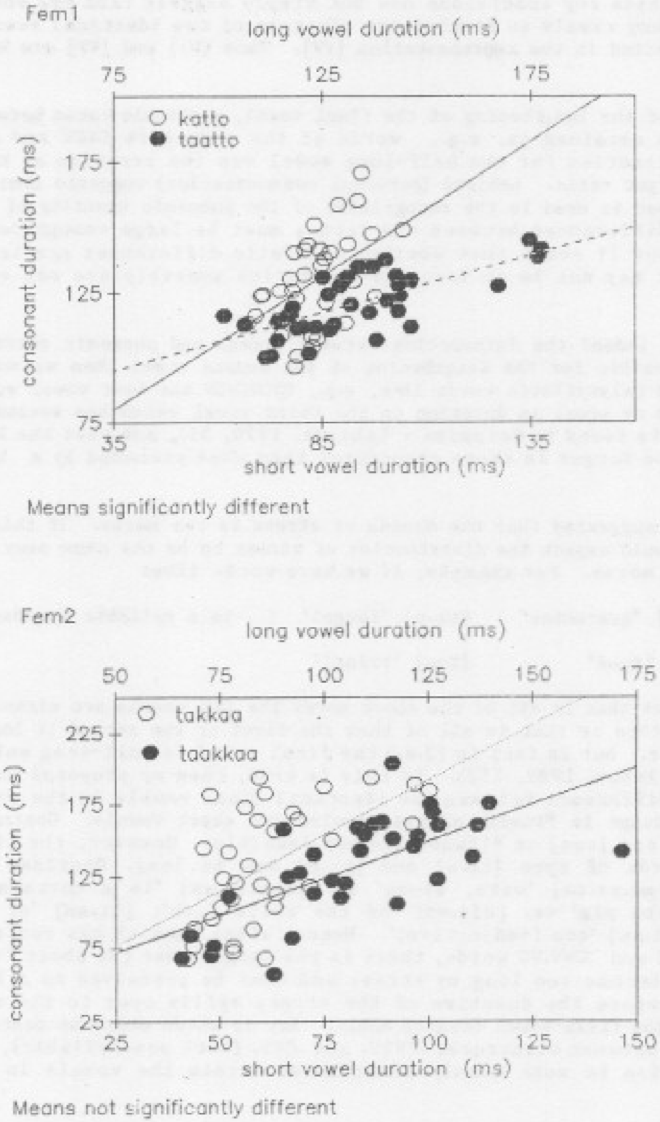


Fig. 4: Geminate consonant duration plotted against preceding long or short vowel duration in all tokens of katto and taatto produced by female 1 (top) and female 2 (bottom). Dashed and dotted lines are regression curves fitted to the two sets of data points.

We need not state any conditions now but simply suggest that the vowel lengthens. Long vowels in Finnish are clusters of two identical vowels which is reflected in the representation [VV]. Thus [V:] and [VV] are kept distinct.

Because of the lengthening of the final vowel, a certain ratio between the vowels is obtained in, e.g., words of the structure CVCV and one proposed explanation for the half-long vowel was the creation of this particular target ratio. Lehiste (personal communication) suggests that if the ratio indeed is used in the recognition of the phonemic quantity of the vowels, the differences between the ratios must be large enough to be perceived. But it seems that whether the ratio differences are large enough or not may not be an issue as the ratios possibly are not even constant.

If it is indeed the interaction between stress and phonemic quantity that is responsible for the lengthening of the second vowel then we would expect that in polysyllabic words like, e.g., CVCVCVCV the last vowel would be longer than or equal in duration to the third vowel which has secondary stress (this is found in Estonian - Lehiste, 1970, 51), and that the last vowel would be longer in these structures than when preceded by a long syllable.

Lehtonen suggested that the domain of stress is two moras. If this is so, then we would expect the distribution of stress to be the same over any two adjacent moras. For example, if we have words like:

(3) a. [ta.e] 'guarantee' [ta.o] 'forge!' (. is a syllable boundary)

b. [tie] 'road' [tuo] 'bring!'

we would expect that in all of the above words the two vowels are either of the same duration or that in all of them the first or the second is longer than the other. But in fact in (3.a.) the final vowel is half-long unlike in (3.b.) (Karlsson, 1982, 152). If this is true, then my proposal would explain the difference between the identical final vowels in the words above. Diphthongs in Finnish always involve two short vowels. Contrasts like *[tuuo] and [tuo] or *[tuoo] are not possible. However, the first vowel in words of type [ta.o] and [ta.e] may be long. Consider the following: [mu.as.sa] 'with, along' vs. [muu.as.sa] 'in a certain...'; [si.an] 'of the pig' vs. [sii.an] 'of the white fish'; [li.an] 'of the dirt' vs. [lii.an] 'too (+adjective)'. Hence, since Finnish has contrast between CV.VC and CVV.VC words, there is the danger that the short first vowel might become too long by stress and thus be perceived as a long vowel. Therefore the duration of the stress spills over to the next syllable and the first vowel remains short. But in words where no contrast can be found between structures .CVVV. and .CVV. (both monosyllabic), the stress-duration is more evenly distributed across the vowels in the syllable.

It has been proposed that in some Finnish dialects the occurrence of the half-long vowel has threatened the CVCV/CVCVV contrast and that this would be the source of the so called "general gemination" of the medial consonant in the CVCVV structures: [sanoo] 'he says' becomes [sannoo] (Karlsson, 1982, 153). But if it is true that long vowels are longer after initial CV-syllables than after long syllables then there should not be any danger of losing a contrast: the final vowels in CVCV and CVCVV are both lengthened, not only in CVCV. In this study all three speakers had a significant difference (P < .00001) between the final vowels in [tupa] and [tupaa]. For one speaker, Female 1, the final vowels in [tupa] vs. [piikaa], [takkaa], [taakkaa] were not significantly different (P < .95) while in [tupa] and [tupaa] they were (P < .00001). This was not the case for the other speakers: they had significant differences also in these instances. By these data we may suggest that the general gemination is caused by the same phenomenon as the half-long vowel in other dialects but in this one only the lengthening falls on the immediately following consonant.

The lengthened vowels and geminates seem to indicate that Finnish is struggling with the property that caused Jakobson and Trubetzkoy to claim that in no language can we have simultaneously free stress(accent) and free quantity. If a language were to have both of these properties then it could have on the same syllable both contrastive stress à la English <dígest>/<digést> and contrastive quantity à la Finnish <lika>/<liika> 'dirt/excess'. If we had an imaginary word like [latáaja] where there is both contrastive stress and quantity on the second syllable it would be hard to tell what caused the long vowel. If we have a short stressed vowel in place of the long, [latája], then this short vowel would be lengthened by the stress and again we would not know what we have there on that syllable. The introduction of a lengthened mora after a short stressed vowel illustrates one strategy to solve this problem without neutralization. If a language were to develop this kind of conflict involving the expression of distinctive stress and distinctive quantity we would expect it to get rid of one of these. And this is what has actually happened in languages which have acquired both of these as phonemic realities (Anderson, 1984).

5. Summary

Eight different word structures were studied, CVCV(V), CVVCV(V), CVCCV(V) and CVVCCV(V), in two different sentence positions, under three different sentence stress conditions and in two different speech rates. The results show that vowels, short and long, are longer after stressed short syllables than after long stressed syllables. Geminates show the same tendency after short stressed vowels. I interpret these facts to be the resolution of a conflict between the phonemic shortness of the vowel at the segmental level and the use of duration as a stress cue at the suprasegmental level: in order to avoid making the first vowel too long in which case it might neutralize with the long vowel, the duration associated with stress spills over to the following mora. This study did not lend unequivocal support to the use of ratios between vowels in the identification of their phonemic quantities. My proposal unifies some phenomena in Finnish which are usually considered to be independent. The half-long vowel is not unique because it behaves exactly like the extra-long long vowel. Also, the lengthened geminate is found after identical contexts: only after a short stressed vowel. These lengthenings relate to the realization of time in the stream of speech and they show that we cannot produce simultaneously something that must take only a little time and something that would make it longer.

Acknowledgement

I wish to thank Mary Beckman for her generous help during the preparation of this paper.

References

Anderson, Stephen R. 1984. "A Metrical Interpretation of Some Traditional Claims about Quantity and Stress." In Mark Aronoff and Richard Oehrle (eds.), Language Sound Structure. MIT Press.

Fónagy, Iván. 1966. "Electro-physiological and acoustic correlates of stress and stress perception." Journal of Speech and Hearing Research 9:231-244.

Karlsson, F. 1982. Suomen kielen äänne- ja muoto-oppi (Finnish Phonology and Morphology), WSOY, Helsinki.

Kohler, Klaus J. 1986. "Invariance and Variability in Speech Timing: From Utterance to Segment in German." In Joseph S. Perkell and Dennis Klatt (eds.) Invariance and Variability in Speech Processes. Hillsdale, NJ: Lawrence-Earlbaum Associates, pp. 268-299.

Lehiste, I. 1970. Suprasegmentals. MIT Press.

Lehtonen, J. 1970. Aspects of Quantity in Standard Finnish, Jyväskylä.

Wiik, K. and I. Lehiste. 1968. "Vowel Quantity in Finnish Disyllabic Words." Congressus secundus internationalis fennougristarum, Helsinki.

Appendix

the corpus of utterances

SET I

Kato tuhosi kauran. failure destroyed oats	'Bad year destroyed the oat harvest.'
KAto§ kauran tuhosi.	'It was the bad year (and not something else) that destroyed the oak harvest.'
TUhosi kato kauran.	'It was indeed the bad year that...'
Kaato tuhosi kannan. felling destroyed base	'Hunting destroyed the base.'
KAAto kannan tuhosi.	'It was hunting that destroyed the base.'
TUhosi kaato kannan.	'Hunting did indeed destroy the base.'
Katto tummui palossa. ceiling darkened fire-in	'The ceiling got dark in the fire.'
KATto palossa tummui.	'It was the ceiling that got dark in the fire.'
TUMmui katto palossa.	'The ceiling did indeed get dark in the fire.'
Piki putosi katolle. tar fell roof-on	'The tar fell on the roof.'
PIki katolle putosi.	'It was the tar that fell on the roof.'
PUtosi piki katolle.	'The tar did indeed fall on the roof.'
Piikki putosi katolle. awl fell roof-on	'The awl fell on the roof.'
PIIKki katolle putosi.	'It was the awl that fell on the roof.'
PUtosi piikki katolle.	'The awl did indeed fall on the roof.'

---

§Capitalization indicates emphasis

Taatto käveli kovaa. grandpa walked fast	'Grandpa walked fast.'
TAATto kovaa käveli.	'It was grandpa who walked fast.'
KAveli taatto kovaa.	'Grandpa did indeed walk fast.'
Tutti putosi pojalta. pacifier fell boy-from	'The pacifier fell from the boy.'
TUTti pojalta putosi.	'It was the pacifier that fell from the boy.'
PUtosi tutti pojalta.	'The pacifier did indeed fall from the boy.'
Tuutti putosi pojalta.	'The (icecream)cone fell from the boy.'
TUUTti pojalta putosi.	'It was the cone that fell from the boy.'
PUtosi tuutti pojalta.	'The cone did indeed fall from the boy.'

SET II

Tupa pilkotti puiden takaa. cottage showed trees behind	'The cottage showed behind the trees.'
TUpa puiden takaa pilkotti.	'It was the cottage (and not something else) that showed behind the trees.'
PILkotti tupa puiden takaa.	'The cottage did indeed show behind the trees.'
Tupaa pikotti puiden takaa. cottage-partit. showed trees behind	'Part of the cottage showed behind the trees.'
TUpaa puiden takaa pilkotti.	'It was part of the cottage ...'
PILkotti tupaa puiden takaa.	'Part of the cottage did indeed...'
Piikaa kehuttiin kovasti. maid-partit. praised a lot	'The maid was praised a lot.'
PIikaa kovasti kehuttiin.	'It was the maid who was praised a lot.'
KEhuttiin piikaa kovasti.	'The maid was indeed praised a lot.'

Takkaa kehuttiin kovasti. fireplace-partit. praised a lot	'The fireplace was praised a lot.'
TAKkaa kovasti kehuttiin.	'It was the fireplace that was praised a lot.'
KEhuttiin takkaa kovasti.	'The fireplace was indeed praised a lot.'
Taakkaa kevennettiin kahdesti. load-partit. lightened twice	'The load was made lighter twice.'
TAAKkaa tuolla kevennettiin. over-there	'It was the load over there that was lightened.'
KEvennettiin taakkaa kahdesti.	'The load was indeed lightened twice.'

OSU #36: Papers from the Linguistics Laboratory 1985-1987, Sept. 1987.

Phonetic Characteristics of Levantine Arabic Geminates with Differing Morpheme and Syllable Structures

Ann M. Miller
Linguistics, Ohio State University

Abstract: Phonologists distinguish two types of geminates: tautomorphemic (belonging to a single morpheme and composed of one set of phonological features) and heteromorphemic (belonging to separate morphemes and composed of two identical sets of phonological features). If there is a phonetic difference between geminates due to the fact that the second type has two separate sets of features while the first type has only one set, it would likely be cued by differences in duration or by movement during the duration of the consonant. Arabic provides a means for comparing these types of consonants since it has both occurring in a variety of morphological affiliations as well as in contrasting surface phonological positions. A pilot study and a larger experiment were performed with speakers of Levantine Arabic to compare these conditions. No significant differences in duration occurred, and several occurrences of both types of geminates had apparent releases during their durations. These results show that no phonetic difference was found between heteromorphemic and tautomorphemic geminates in Levantine Arabic. On the other hand, a phonetic difference was found linked to syllable structure: tautosyllabic and heterosyllabic geminates had different mean durations as compared to their single counterparts. Since the structure that has been posited by phonologists (association of phonological features with C and V slots) does not capture these differences, it is therefore necessary to refer to syllable structure (= association of phonetic features with syllable slots) in order to represent these phonetic differences.

1. Introduction

Phonologists have recently distinguished two types of geminates based on evidence that geminates with different phonological structures behave differently in certain phonological and morphological situations (cf. Hayes 1986, McCarthy 1986, and the references listed in these articles). Tautomorphemic geminates occur within a single morpheme and have one set of phonological features linked to two consonant slots. They are also known as "monomorphemic," "underlying," or "true" geminates. Heteromorphemic geminates are formed by concatenation of morphemes and have two identical sets of phonological features linked to two consonant slots. They are also known as "derived" or "fake" geminates.

One of the differences between these types of geminates has been pointed out clearly by McCarthy (1986:210-219), who discusses a number of phonological and morphological situations in which tautomorphemic geminates act like one segment rather than like two. He argues that, therefore, the two consonant slots of tautomorphemic geminates are associated with only one underlying set of phonological features (= "melodic segment") rather than with two as the two consonant slots of heteromorphemic geminates are.

Other differences have been articulated clearly by Hayes (1986:326-33), who discusses differences between these types of geminates in a number of phonological situations. He shows that tautomorphemic geminates are never split by epenthesis (they "have integrity"), while heteromorphemic geminates can be split by epenthesis (they do not "have integrity") if they are not formed by total assimilation. Also, tautomorphemic geminates are never affected by phonological rules whose descriptions otherwise fit them (and so would be expected to apply to them) if the rules specify that the affected segment be associated with only one consonant slot (they "respect inalterability"), while heteromorphemic geminates in this situation are affected by such phonological rules (they do not "respect inalterability"). Furthermore, Hayes points out that heteromorphemic geminates derived by total assimilation have the surface structure of tautomorphemic geminates and exhibit their characteristics. He notes that none of these characteristics is surprising when it is accepted that these geminates which act differently have different phonological structures. In fact, the differences in behavior fall out from the differences in structure.

In order to capture the differences in phonological structure, these two types of geminates are often schematized as in (1) and (2). (1) shows that tautomorphemic geminates and heteromorphemic geminates derived by assimilation are represented as consisting of one set of phonological features linked to two consonant slots. (2) shows that heteromorphemic geminates not derived by assimilation are represented as consisting of two sets of phonological features, each linked to a separate consonant slot.

(1) Phonological Representation of Tautomorphemic Geminates and Heteromorphemic Geminates Derived by Assimilation

CC

[F] ([F] = any set of phonological features which comprise a segment in a language, often abbreviated with the phonetic symbol; e.g., [t])

(2) Phonological Representation of Heteromorphemic Geminates

CC

[F][F]

Older phonetic studies have provided evidence that there are some possible phonetic correlates of these phonological differences, although not couched in these phonological terms and not specifically addressing this issue. Some conflicting results have been obtained, but the fact that positive results have been obtained by some studies shows that phonetic differences may exist between different types of geminates--although perhaps in specific conditions that have yet to be identified.

Conflicting results have been obtained for movement characteristics during the duration of heteromorphemic and tautomorphemic geminates. Several studies have found heteromorphemic geminates to have two peaks or phases in their amplitude curves (Stetson 1951: heteromorphemic geminates in English) and in their amplitude curves and tongue shapes (Delattre 1971: English, German, French, and Spanish), but tautomorphemic geminates to generally have only one peak or phase in these (Hegedűs 1959: Hungarian had only one peak; Delattre 1971: Spanish word-initially and German medially and word-finally had only one phase while Spanish medially had one phase in less than half the tokens but two phases in more than half). Other studies, on the other hand, have found heteromorphemic and tautomorphemic geminates not to differ with respect to average number of peaks in their air pressure curves, articulatory pressure curves, or EMG traces. Evidence for this is found in studies which have shown both types of geminates to generally have at least two such peaks but the corresponding single consonants to generally have only one such peak (Poirot in Rousselot 1897-1901 and 1901-1908:1087: tautomorphemic geminates vs. single consonants; and Lehiste, Morton, and Tatham 1973: heteromorphemic and tautomorphemic [two lengths] geminates compared to single consonants and to each other in Estonian). It should be noted, though, that contrary to Delattre (1971), Lehiste, Morton, and Tatham (1973) found the slightly greater number of EMG peaks in English heteromorphemic geminates not to be significantly greater than the number of peaks in corresponding single consonants. It should also be noted that Rousselot (1891, 1897-1901 and 1901-1908:351 and 1087, and 1913:77) found only two tautomorphemic geminate consonants which showed two peaks in the air pressure and articulatory pressure curves in Swedish and none in Gallo-Roman. However, Stetson (1951:61) believed that Rousselot's results could not be relied on because his equipment was not very sensitive.

Conflicting results have also been obtained for the duration characteristics of heteromorphemic and tautomorphemic geminates. A number of studies have found both types of geminates to be about one and one-half to two times as long as the corresponding single consonants for the shortest length distinction within a language (geminate to single consonant in languages with only one length distinction, and short geminate to single consonant in languages with more than one length distinction) (Rousselot 1891: tautomorphemic geminates in Gallo-Roman; Josselyn 1901:222ff.: tautomorphemic geminates in Italian; Stetson 1951: heteromorphemic geminates in English; Delattre 1971: tautomorphemic geminates in German and Spanish, and heteromorphemic geminates in these languages, English, and French; and Lehiste, Morton, and Tatham 1973: short tautomorphemic geminates in Estonian). A recent phonetic study (Hankamer and Lahiri 1986) also found no significant difference in length between tautomorphemic geminates and heteromorphemic geminates in Turkish and Bengali. All these results point to the conclusion that heteromorphemic and tautomorphemic

---

[1] My classification of one of the sets of French tokens--the conditional present--as containing heteromorphemic rather than tautomorphemic geminates depends on the morphological analysis of the tokens mourrait, courrait, and acquerrait as containing an infixed -r- rather than a stem which ends in -rr-. The latter is the traditional account (cf. Judge and Healey 1983:221, 259n.2; Ollivier 1978:165-66), while the former is based on formal linguistic analysis (cf. Judge and Healey 1983:216, 259n.2). I have chosen to follow the linguistic analysis. Following the traditional analysis would change the conclusions made here since half the French tokens with two phases in Delattre's study would then contain tautomorphemic geminates.

geminates do not differ in their duration characteristics since they both differ by the same amount in duration from their single counterparts. However, Lehiste, Morton, and Tatham (1973) obtained results contradictory to this in their study of English, in which English heteromorphemic geminates did not differ significantly in length from their single counterparts. Also, the two older phonetic studies discussed here which investigated both heteromorphemic and tautomorphemic geminates in the same study showed the tautomorphemic geminates to sometimes be significantly different in duration from the corresponding heteromorphemic geminates in the language (Delattre 1971: comparisons of both types of geminates to the corresponding single consonants showed this for Spanish word-initially and medially, and German word-finally, although not for German medially; Lehiste, Morton, and Tatham 1973: comparisons of heteromorphemic geminates to long tautomorphemic geminates in Estonian showed this for one of the two sets of tokens).

The evidence described above of possible phonetic differences between different types of geminates is in contrast to assumptions currently made by phonologists, who assume that, despite the different phonological structures, there is no phonetic difference between different types of geminates. For example, McCarthy (1986:250) states that "we usually find that languages make no phonetic distinction between hetero- and tautomorphemic geminates, despite their different melodic representation." It has been suggested that the presence vs. the absence of distinguishing phonetic characteristics of heteromorphemic and tautomorphemic geminates may be language-specific (Lehiste, Morton, and Tatham 1973:147) since a native speaker of Estonian and a native speaker of English each carried her native-language patterns of duration and number of peaks in EMG traces over to the other language. However, the existence of contradictory results for English (by Stetson 1951 and Delattre 1971, as described above) and the negative results of Hankamer and Lahiri (1986) suggest that there may also be other reasons for differences in phonetic characteristics of geminate consonants. These may be differences between speakers within a language or differences in language structure according to the number and type of contrasts (e.g., single, short tautomorphemic, long tautomorphemic, and heteromorphemic consonants in Estonian vs. only single and heteromorphemic consonants in English) which exist in the consonants of the languages. Much more study is needed before possible phonetic differences due to variation between speakers or language types can be pinpointed accurately. But the existence of the contradictory results described here shows that the existence or non-existence of possible phonetic differences between different types of geminates has not yet been determined.

Since duration and movement are the major characteristics for which differing results have been obtained in studies of geminates, this study investigated these characteristics for various types of geminates in yet another language--Arabic--by investigating the Levantine dialect. Arabic is a good candidate for such study because it has not only tautomorphemic and heteromorphemic geminate consonants[2] but also tautosyllabic and

---

[2]Neither traditional analyses nor recent linguistic analyses of Arabic (cf. Brame 1970; McCarthy 1979, 1981) distinguish by terminology these types of geminates. These analyses refer to all of the lengthened consonants with the same term--as doubled or geminate consonants--and also represent all of

heterosyllabic geminate consonants. Geminates in Levantine Arabic thus occur in a variety of morphological affiliations as well as in contrasting surface phonological positions. This made possible an investigation for phonetic similarities and differences in terms of both morphological and phonological structure. In terms of morphological structure, Arabic has tautomorphemic geminate consonants word medially (e.g., [kət$təb]) and finally (e.g., [fətt]), and heteromorphemic geminate consonants across morpheme boundaries (e.g., [t+t+əbəʕ]) and word boundaries (e.g., [nəbət#təxt] and [rufæːt#təxt]), including some which are assimilated (e.g., [d+d+əθər] < /d+t+əθər/ and [t#təxt] < /ʔIl#təxt/). In terms of surface phonological structure, the tautosyllabic and heterosyllabic geminate consonants in Arabic both include some tautomorphemic and some heteromorphemic geminates. Tautosyllabic geminates include tautomorphemic [fətt] as well as heteromorphemic [t+t+əbəʕ] and [t+təxt]. Heterosyllabic geminates include tautomorphemic [kət$təb] and heteromorphemic [nəbət#təxt] and [rufæːt#təxt]. This second contrast is especially interesting since it seems likely that if there is a regular phonetic difference between different types of geminates, it would more likely reflect differences at the surface phonological level than at the abstract morphological level.

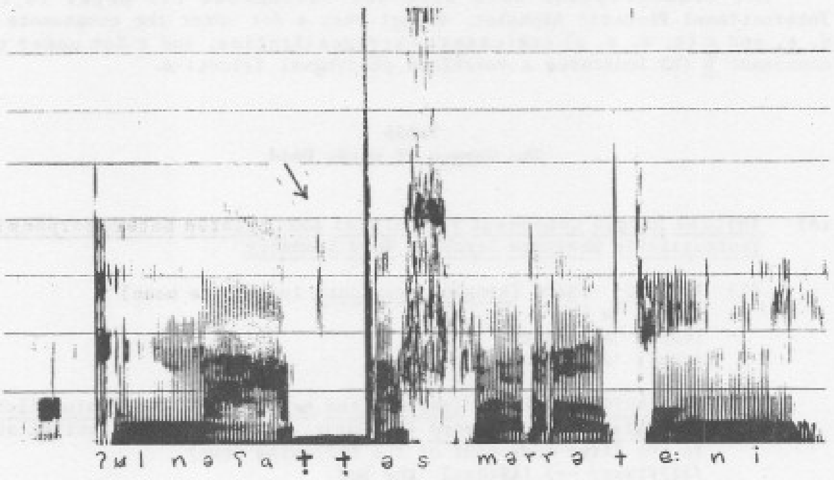
This investigation examined these two sets of contrasting types of geminate consonants in Levantine Arabic by comparing them to minimally contrasting short consonants. It was expected that if there is a phonetic distinction between tautomorphemic and heteromorphemic geminate consonants or between tautosyllabic and heterosyllabic geminate consonants, it would likely be cued by durational differences or by the absence (which would suggest no movement) vs. the presence (which would suggest movement),

---

them as being composed of two identical consonants (the recent analyses including some justification in terms of their overall frameworks). However, these lengthened consonants are structurally two different types of consonants on morpho-syntactic grounds since the first do not cross morpho-syntactic boundaries whereas the second do. Therefore, the first are treated here as tautomorphemic geminates, and the second are treated as heteromorphemic geminates. Although I have argued that McCarthy's (1979, 1981) analysis of Arabic is inelegant in its treatment of the morphological differences in Arabic, thereby obscuring lexical similarities of the various binyanim (see Miller 1986), it should be noted that a strength of McCarthy's analysis is that the structure it proposes for Arabic does implicitly recognize the morphological difference between heteromorphemic and tautomorphemic geminates by representing them differently from each other at the autosegmental level. For the geminates treated in this paper, for example, it places the Form VIII reflexive infix -t- on a separate autosegmental (= morphological) tier from the root consonants (which it sometimes assimilates to, thus forming a geminate consonant) and has the infix and the root consonants associate to different consonant slots in the prosodic template, while it places the geminate consonants in Form II and in geminate roots on the same tier by representing them as composed of one melodic element which associates with two consonant slots in the prosodic template. However, even though McCarthy's (1979, 1981) analysis represents these types of consonants as different structurally at the autosegmental level, the terminology is misleading since it terms both types of geminates as simply geminates and does not label them as different in terms of their CV structures.

respectively, of a release spike during the duration of the consonants since the first member of each pair is <u>one</u> unit at that level of structure (<u>tauto</u>morphemic and <u>tauto</u>syllabic), while the second member of each pair is composed of <u>two</u> units (<u>hetero</u>morphemic and <u>hetero</u>syllabic).

A pilot study found that the difference between tautomorphemic geminates and heteromorphemic geminates in a small corpus was not cued systematically by either of these characteristics. This suggested that there may not be a phonetic difference between these types of geminate consonants. The results also suggested the possibility that tautomorphemic geminate consonants may not be as closely joined as the phonological studies claim since one instance of a medial tautomorphemic geminate showed an apparent release spike in the middle of the closure duration--which may be evidence of rearticulation[3]--as did three instances of heteromorphemic geminates across word boundaries. Figure 1 shows the spike which occurred in the middle of the tautomorphemic geminate duration.



<u>Figure</u> <u>1</u>. An apparent release spike--which may be evidence of rearticulation--in the middle of the tautomorphemic geminate [ṭṭ] in [ʕaṭṭəs].

The larger study reported here confirmed the results of the pilot study. Furthermore, the results suggested that the factor which influenced the durations of the geminates (whether tautomorphemic or heteromorphemic) was the type of phonological boundary crossed rather than the type of

---

[3]The possibility that this could also show movement of saliva without rearticulation was suggested to me by Arthur Abramson during the presentation of the results of the pilot study as a poster at the 111th Meeting of the Acoustical Society of America, May 14, 1986. Since this is certainly possible, further investigation is needed in order to determine the cause of the spikes.

morpho-syntactic boundary. This was suggested by the fact that the heterosyllabic geminates (which include both tautomorphemic and heteromorphemic geminates) had similar durations relative to their single counterparts (approximately 2.5:1), while the tautosyllabic geminates (which include both tautomorphemic and assimilated heteromorphemic geminates) had a shorter duration relative to their single counterparts (approximately 2:1).

2. Method

The corpus comprised the three dental consonants which occur in all the positions to be tested: [t, d, and ṭ]. The words exemplified six sets of contrast types, with one word type chosen for each relevant consonant in each condition, as shown in Table I below, where X = tested consonant.

The transcription used here and throughout the paper is the International Phonetic Alphabet, except that a dot under the consonants t, d, s, and z (ṭ, ḍ, ṣ, ẓ) indicates pharyngealization, and a dot under the consonant h (ḥ) indicates a voiceless pharyngeal fricative.

Table I
The Corpus of Words Used

(A) Initial Single Consonant vs. Initial Assimilated Heteromorphemic, Tautosyllabic Geminate Across A Word Boundary

(1) Type #C- = XəCC (single consonant; indefinite noun)
[təxt] 'a bed'
[dərs] 'a lesson'
[ṭəbl] 'a drumbeat'

(2) Type ART#C- = X#XəCC (assimilated heteromorphemic, tautosyllabic geminate; across a word boundary; definite article assimilated to the first consonant of the following noun)[4]
/ʔIl#təxt/ --> [t#təxt] 'the bed'
/ʔIl#dərs/ --> [d#dərs] 'the lesson'
/ʔIl#ṭəbl/ --> [ṭ#ṭəbl] 'the drumbeat'

(B) Medial Single Consonant vs. Medial Tautomorphemic, Heterosyllabic Geminate Across a Syllable Boundary

(3) Type -C- = CəXəC (single consonant; Form I verb)
[kətəb] 'to write'
[wədəʕ] 'to give up'
[bəṭəl] 'to become invalid'

---

[4]When the definite article /ʔIl/ occurs in speech, [ʔI] is generally not pronounced when the preceding word begins with a vowel. In this study it was never pronounced, so it is not included in the phonetic representation.

(4) Type -C$C- = CəX$XəC (tautomorphemic, heterosyllabic geminate; Form II verb = 'causative of Form I')
[kət$təb] 'to make someone write'
[wəd$dəʕ] 'to take leave of someone'
[bəṭ$ṭəl] 'to invalidate'

(C) Initial Single Consonant vs. Initial Assimilated Heteromorphemic, Tautosyllabic Geminate Across a Morpheme Boundary

(5) Type #C- = XəCVC (Form I verb)
[təbIʕ] 'to follow'
[dəθər] 'to be forgotten'
[ṭələʕ] 'to come into view'

(6) Type #C+REFL+ = X+X+əCəC (assimilated heteromorphemic, tautosyllabic geminate; across a morpheme boundary; Form VIII verb = 'reflexive of Form I'; reflexive morpheme /-t-/ assimilated to the first consonant of the verb)[5]
/t+t+əbəʕ/ --> [t+t+əbəʕ] 'to follow'
/d+t+əθər/ --> [d+d+əθər] 'to cover oneself'
/ṭ+t+ələʕ/ --> [ṭ+ṭ+ələʕ] 'to be well informed about'

(D) Initial Single Consonant vs. Initial Heteromorphemic, Heterosyllabic Geminate Across a Word Boundary

(7) Type #C- = Cə[CəC#XəCC] (single consonant; noun; the phonetic material in brackets matching that in brackets in Condition 8 as closely as possible--since verbs end in a consonant, it was not possible to have a schwa precede the tested consonant in Condition 7 as it does in Condition 8--so that the difference of interest between Conditions 7 and 8 is the length of the initial C of the second word)
[ǰənəb#təxt] 'He warded off a bed.'
[zəḥəm#dərs] 'He jostled a lesson.'
[ḥəbəs#ṭəbl] 'He apprehended a drumbeat.'

(8) Type C#C- = [CəCəX#XəCC] (heteromorphemic, heterosyllabic geminate; across a word boundary; Form I verb#noun)
[nəbət#təxt] 'He sprouted a bed.'
[ḥəməd#dərs] 'He praised a lesson.'
[bəsəṭ#ṭəbl] 'He offered a drumbeat.'

(E) Final Single Consonant vs. Final Tautomorphemic, Tautosyllabic Geminate Across No Word Boundary

(9) Type -C# = CV(C)[CV:X#I1#CəCC] (single consonant; noun#definite

---

[5]This contrast occurs only for [d], [t], and [ṭ] because these are the only consonants which the reflexive morpheme -t- assimilates to totally when it is infixed after the first consonant of the verb. Hence, this study investigated only these three consonants.

article#noun; the phonetic material in brackets matching that in brackets in Condition 10 as closely as possible--including matching the stress on the first word in 10 with a long vowel in 9--so that the difference of interest is the length of the final consonant of the first noun)[6]

[rufæ:t#Il#kəlb] 'the remains of the dog'
*[ʔIl#ḥa:d#Il#xərb] 'the graves of the war'
*[ʔIl#ɣa:ṭ#Il#kəlb] 'the noises of the dog'

(10) Type -CC# = [CəXX#Il#CəCC] (tautomorphemic, tautosyllabic geminate; noun#definite article#noun)

[fətt#Il#kəlb] 'the weakening of the dog'
[ḥədd#Il#xərb] 'the limiting of the war'
[ɣəṭṭ#Il#kəlb] 'the dipping of the dog'

(F) Final Single Consonant vs. Heteromorphemic, Heterosyllabic Geminate Across a Word Boundary

(9) Type -C# = [CV(C)CV:X]#Il#CəCC (single consonant; noun#definite article#noun; the phonetic material in brackets matching that in Condition 11)

[rufæ:t#Il#kəlb] 'the remains of the dog'
*[ʔIl#ḥa:d#Il#xərb] 'the graves of the war'
*[ʔIl#ɣa:ṭ#Il#kəlb] 'the noises of the dog'

(11) Type -C#C = [CV(C)CV:X]#XəCC (heteromorphemic, heterosyllabic geminate; across a word boundary; noun#indefinite noun)[7]

[rufæ:t#təxt] 'remains of a bed'
*[ʔIl#ḥa:d#dərb] 'the graves of a pass'
*[ʔIl#ɣa:ṭ#ṭəbl] 'the noises of a drumbeat'

Five tokens of each word were written on separate index cards, grouped into the two categories A + B + C, and D + E + F, and randomized within categories by shuffling. One native speaker of Levantine Arabic from the pilot study participated in this experiment. Before beginning the experiment, he was asked to read through the words on the cards to make sure he could read them, and he received the same instructions as in the pilot study, written in Arabic and reproduced in (3) in English.

(3) "Place each of the following words in the blank space in the following sentence: [ʔulnə______ mərrə tæ:ni] 'We said ______ again.' Read the sentence as if you were speaking with a good friend."

When the native speaker had finished reading, the experimenter told him again, for emphasis, to be sure to say each written word as he would in his usual speech at home. Then he sat in an anechoic chamber to carry out the

---

[6]Two phrases in which the first definite article is ungrammatical--marked by *--were included in order to facilitate the phonetic matching.

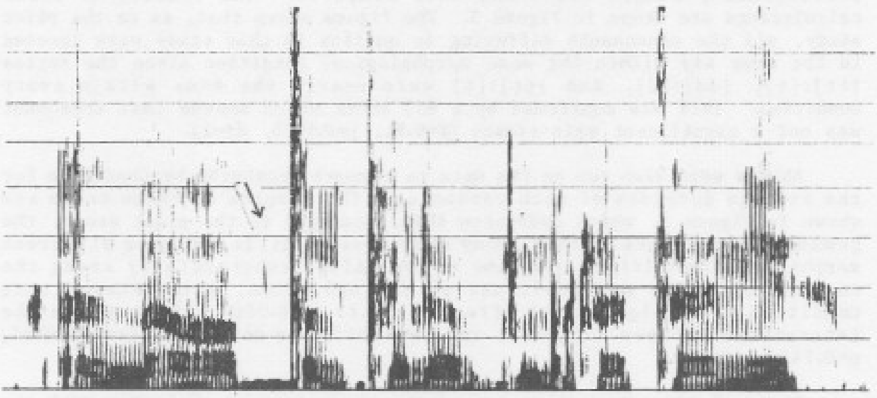
[7]As in Condition 9 (see Note 6), two phrases in which the first definite article is ungrammatical--marked by *--were included in order to facilitate the phonetic matching.

experiment and was recorded as he read the sentences. A spectrogram was made of each sentence, and the length of each tested consonant was measured to the nearest half millimeter.

3. Results and Discussion

Spectrograms showed that, as in the pilot study, release spikes were sometimes present within the duration of both the tautomorphemic and the heteromorphemic geminates. Not surprisingly, they were found in two types of unassimilated heteromorphemic geminates: one consonant geminated initially across a word boundary (Type C#C-, [bəsəṭ#ṭəbl]), and one consonant geminated finally across a word boundary (Type -C#C, [ʔIl#ɣa:ṭ#ṭəbl]). However, they were also found in two types of assimilated heteromorphemic geminates (which are said to have a tautomorphemic type of structure and therefore not to be separable): one consonant geminated with the assimilated definite article (Type ART#C-, [d#dərs]), and one consonant geminated initially with the assimilated reflexive morpheme (Type #C+REFL+, [ṭ+ṭ+ələʕ]). Furthermore, they were found in one type of tautomorphemic geminate: two final tautomorphemic geminates (Type -CC#, two instances of [ɣəṭṭ]). The spectrogram in Figure 2 shows one of the tokens with a release spike present within the duration of the tautomorphemic geminate.

All of these spikes appear to indicate some sort of rearticulation within the duration of the consonant--whether a heteromorphemic or a tautomorphemic geminate. These results confirm the spectrographic results of the pilot study that spikes did not distinguish the heteromorphemic geminates from the tautomorphemic geminates since they occurred during the



<u>Figure</u> <u>2</u>. An apparent release spike--which may be evidence of rearticulation--within the duration of the tautomorphemic geminate consonant [ṭṭ] in [ɣaṭṭ ʔIl kələb].

closure durations of both. This suggests that if there is a phonetic difference between the two types of geminate consonants it must be something other than the presence or absence of a release within the consonant duration.

The instances of apparent rearticulation in the heteromorphemic geminates across a word boundary were not surprising because phonological studies have shown that the individual consonants of heteromorphemic geminates can be separated. However, the instances in the final tautomorphemic geminates and in the heteromorphemic geminates assimilated with the definite article and with the reflexive morpheme were surprising since phonological studies have given evidence that the individual consonants in these types of geminates are never separated. These results also suggest the possibility that the individual consonants of tautomorphemic geminates and assimilated heteromorphemic geminates are less closely joined than the phonological studies suggest since the consonant duration can be interrupted in the middle by a spike, like heteromorphemic geminates, whose individual consonants are treated phonologically as not joined to each other at all. As in the pilot study, this hypothesis cannot be taken as conclusive since a spike occurred only two times in the middle of both a tautomorphemic geminate and an assimilated heteromorphemic geminate in this study and since this study did not find any spikes in the medial tautomorphemic geminates (Type -C$C-) as the pilot study did. However, this is an interesting occurrence worthy of further investigation.

To examine whether there were differences in duration between heteromorphemic and tautomorphemic geminates, the average duration of the five tokens of the geminate consonant and the five tokens of the single consonant for each word pair was calculated. Then the ratios of the average duration of each geminate consonant to the average duration of the corresponding single consonant were computed. The results of these calculations are shown in Figure 3. The figure shows that, as in the pilot study, all the consonants differing in quality in this study were treated in the same way within the same morphological condition since the ratios [tt]:[t], [dd]:[d], and [ṭṭ]:[ṭ] were nearly the same within every condition. This was confirmed by a 6X3 ANOVA which showed that Consonant was not a significant main effect ($F=0.91$, $p=0.4070$, $df=2$).

ANOVAs were also run on the data to compare Consonant by Condition for the average duration of each consonant. The results of these tests are shown in Figure 3, which indicates that, contrary to the pilot study, the geminate consonants in this study were treated differently in different morphological conditions since the ratios differ substantially among the conditions. This was confirmed by the 6X3 ANOVA, which showed that Condition was a significant effect ($F=11.73$, $p<0.0001$, $df=5$), while the interaction between Consonant and Condition was not significant ($F=0.67$, $p=0.7462$, $df=10$).

Figure 3 also shows that two of the three sets of geminates with the same morphological structure--unassimilated and assimilated heteromorphemic geminates--were treated in the same way since the ratios for the former (C#C- and -C#C) and for the latter (#C+REFL+ and ART#C-) were similar. On the other hand, Figure 3 shows that the two types of geminates with tautomorphemic structure were treated differently from each other since the ratios for -CC# and -C$C- differ substantially.

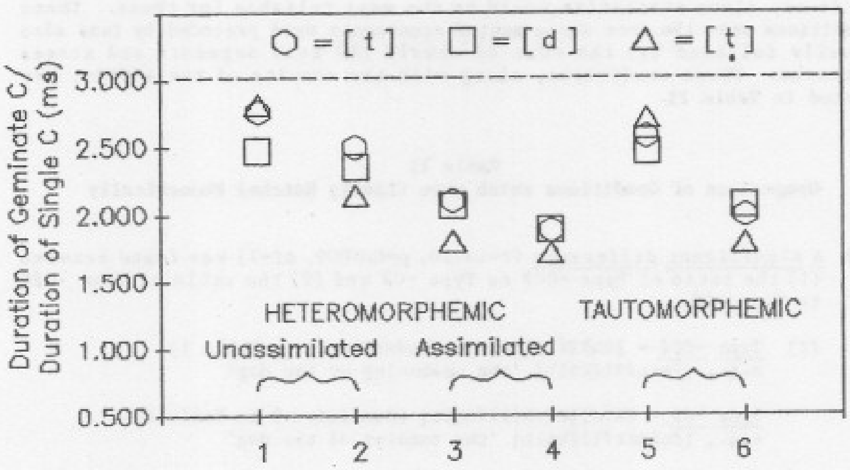


Figure 3. Ratios of the average duration of the geminate consonants to the average duration of the corresponding single consonants, arranged to show the values for heteromorphemic geminates on the left and for tautomorphemic geminates on the right.
1 = C#C-/#C- 3 = #C+REFL+/#C- 5 = -C$C-/-C-
2 = -C#C/-C# 4 = ART#C-/#C- 6 = -CC#/-C#

Therefore, while this analysis shows that all three of the consonants tested were treated somehow in the same way, it does not show what this was. It shows that (1) the unassimilated heteromorphemic geminates in this study were all treated the same--as predicted by phonological theory, (2) the assimilated heteromorphemic geminates were treated the same as each other but differently from the unassimilated geminates--also as predicted by phonological theory, and (3) the tautomorphemic geminates were not all treated the same: the medial tautomorphemic geminates were treated like the unassimilated heteromorphemic geminates, while the final tautomorphemic geminates were treated like the assimilated heteromorphemic geminates.

The random pattern just described in result (3) is not as predicted by phonological theory, which says that tautomorphemic geminates and assimilated heteromorphemic geminates pattern together as opposed to unassimilated heteromorphemic geminates. Phonological theory does not say that there are different types of tautomorphemic geminates, each of which patterns with a different type of heteromorphemic geminate, as this study found. These results therefore seem to indicate another type of pattern which has not been identified by phonological theory--a pattern of phonetic difference between geminates with some sort of different structure, but not between the classes of heteromorphemic geminates and tautomorphemic geminates.

To verify the significance of the differences shown in Figure 3, four post-hoc ANOVAs were done to compare the four conditions in this experiment whose environments were matched phonetically as closely as it was possible to do so, since statistics would be the most reliable for these. These conditions were the ones whose tested consonants were preceded by (and also usually followed by) the same or nearly the same segments and stress patterns. These conditions, along with the results of the ANOVAs, are listed in Table II.

Table II
Comparison of Conditions which were Closely Matched Phonetically

(A) A significant difference (F=14.20, p=0.0009, df=1) was found between (1) the ratio of Type -CC# to Type -C# and (2) the ratio of Type -C#C to Type -C#.

(1) Type -CC# = [CəXX#Il#CəCC] (Condition 10 in Table I)
e.g., [fətt#Il#kəlb] 'the weakening of the dog'

Type -C# = CV(C)[CV:X#Il#CəCC] (Condition 9 in Table I)
e.g., [rufæ:t#Il#kəlb] 'the remains of the dog'

The ratio of the duration of final geminates to the duration of final single consonants was computed from these types. In this contrast, the phonetic material immediately preceding each pair of tested consonants (in brackets) was matched as closely as possible by matching the stress on the first word in Type -CC# with a long vowel in Type -C#. The phonetic material following each pair of tested consonants (also in brackets) was matched exactly.

(2) Type -C#C = [CV(C)CV:X]#XəCC (Condition 11 in Table I)
e.g., [rufæ:t#təxt] 'remains of a bed'

Type -C# = [CV(C)CV:X]#Il#CəCC (Condition 9 in Table I)
e.g., [rufæ:t#Il#kəlb] 'the remains of the dog'

The ratio of the duration of geminates across a word boundary to the duration of final single consonants was computed from these types. In this contrast, the phonetic material preceding each pair of tested consonants (in brackets) was matched exactly, while the material following the tested consonants (not in brackets) differed.

(B) A significant difference (F=14.18, p=0.0010, df=1) was found between (1) the ratio of Type C#C- to Type #C- and (2) the ratio of Type #C+REFL+ to Type #C-.

(1) Type C#C- = [CəCəX#XəCC] (Condition 8 in Table I)
e.g., [nəbət#təxt] 'He sprouted a bed.'

Type #C- = Cə[CəC#XəCC] (Condition 7 in Table I)
e.g., [ǰənəb#təxt] 'He warded off a bed.'

The ratio of the duration of geminates across a word boundary to the duration of initial single consonants was computed from these types. In this contrast, the phonetic material in brackets in both of these conditions was matched except that the tested single consonants could not be preceded by a schwa, as the tested geminate consonants were, because verbs in Levantine Arabic end in consonants. Both types of tested consonants were, however, preceded by unstressed syllables.

(2) Type #C+REFL+ = X+X+əCəC (Condition 6 in Table I)
e.g., /ṭ+t+əlǝʕ/ --> [ṭ+ṭ+əlǝʕ] 'to be well informed about'

Type #C- = XəCVC (Condition 5 in Table I)
e.g., [ṭəlǝʕ] 'to come into view'

The ratio of the duration of initial geminates assimilated across a morpheme boundary to the duration of initial single consonants was computed from these types. In this contrast, both the preceding and the following phonetic material was matched exactly, except that the second vowel after one of the tested single consonants ([t]) was [I], while all the other vowels were [ə].

(C) A significant difference (F=19.88, p=0.0002, df=1) was found between (1) the ratio of Type C#C- to Type #C- and (2) the ratio of Type ART#C- to Type #C-.

(1) Type C#C- = [CəCəX#XəCC] (Condition 8 in Table I)
e.g., [nəbət#təxt] 'He sprouted a bed.'

Type #C- = Cə[CəC#XəCC] (Condition 7 in Table I)
e.g., [ǰənəb#təxt] 'He warded off a bed.'

The ratio of the duration of geminates across a word boundary to the duration of initial single consonants was computed from these types. In this contrast, the phonetic material in brackets in both of these conditions was matched except that the tested single consonants could not be preceded by a schwa, as the tested geminate consonants were, because verbs in Levantine Arabic end in consonants. Both types of tested consonants were, however, preceded by unstressed syllables.

(2) Type ART#C- = X#XəCC (Condition 2 in Table I)
e.g., /ʔIl#təxt/ --> [t#təxt] 'the bed'

Type #C- = XəCC (Condition 1 in Table I)
e.g., [təxt] 'a bed'

The ratio of the duration of initial geminates assimilated across a word boundary to the duration of initial single consonants was computed from these types. In this contrast, the phonetic material both preceding and following the tested consonants was matched exactly.

(D) <u>No significant difference</u> (F=2.46, p=0.1302, df=1) was found between (1) the ratio of Type #C+REFL+ to Type #C- and (2) the ratio of Type ART#C- to Type #C-.

(1) <u>Type</u> <u>#C+REFL+</u> = X+X+əCəC (Condition 6 in Table I)
e.g., /ṭ+t+ələʕ/ --> [ṭ+ṭ+ələʕ] 'to be well informed about'

<u>Type</u> <u>#C-</u> = XəCVC (Condition 5 in Table I)
e.g., [ṭələʕ] 'to come into view'

The ratio of the duration of initial geminates assimilated across a morpheme boundary to the duration of initial single consonants was computed from these types. In this contrast, the phonetic material both preceding and following the tested consonants was matched exactly, except that the second vowel after one of the tested single consonants ([t]) was [I], while all the other vowels were [ə].

(2) <u>Type</u> <u>ART#C-</u> = X#XəCC (Condition 2 in Table I)
e.g., /ʔIl#təxt/ --> [t#təxt] 'the bed'

<u>Type</u> <u>#C-</u> = XəCC (Condition 1 in Table I)
e.g., [təxt] 'a bed'

The ratio of the duration of initial geminates assimilated across a word boundary to the duration of initial single consonants was computed from these types. In this contrast, the phonetic material both preceding and following the tested consonants was matched exactly.

The statistics given in Table II support the conclusions which were drawn by inspection of Figure 3. They confirm that the differences between three of the higher and lower CC:C ratios (shown by the higher peaks and the lower peaks) in Figure 3 are significant, while the differences between two of the lower peaks are not significant. Although the statistical significance or insignificance of the other differences shown in Figure 3 cannot be verified directly since any other comparisons involve material that is not matched phonetically, it can be assumed by analogy to the tested differences that the other large differences are probably significant, while the small differences are probably not significant. This assumption should, of course, be tested by further study.

In addition, a t-test run on the two geminates of different type whose environments were matched completely (-CC# and -C#C) provides further evidence that the lengths of these geminates were significantly different. It showed that for types -CC# (tautomorphemic, tautosyllabic) and -C#C (heteromorphemic, heterosyllabic) the absolute values of their mean durations--and not just their ratios--differed significantly. Figure 4 shows this clearly since each consonant of the type -C#C is substantially longer than the same consonant of the type -CC#. The other geminates in this study could not be accurately tested this way since none of their environments were matched totally.

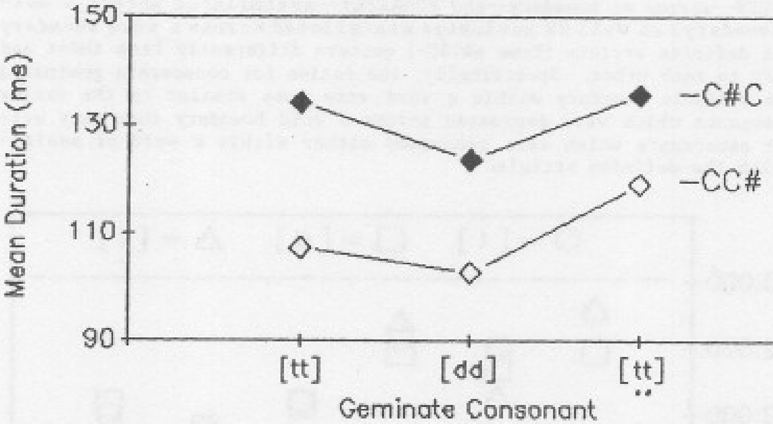


Figure 4. Mean durations of geminates in exactly the same environments (-CC# and -C#C), showing that the absolute values of their mean durations--and not just their ratios--differed significantly.

If the durational differences shown in Figure 3 are reliable, then, how do we account for the fact that the tautomorphemic geminates do not exhibit the same behavior for this characteristic--contrary to the predictions of current phonological theory, while the unassimilated and assimilated heteromorphemic geminates do--as predicted by current theory? Furthermore, why does one of the tautomorphemic geminates (-CC#) pattern with the assimilated heteromorphemic geminates--as current theory predicts--and the other tautomorphemic geminate (-C$C-) pattern with the unassimilated heteromorphemic geminates--the opposite of what is predicted? It is clear that something besides the morpheme structure is involved in determining the pattern of durational differences.

What the two different groups have in common within themselves and different from each other is syllable structure. The unassimilated heteromorphemic geminates and the tautomorphemic geminate which patterns with them are heterosyllabic; that is, the geminates are part of two syllables. On the other hand, the assimilated heteromorphemic geminates and the tautomorphemic geminate which patterns with them are tautosyllabic--they are part of only one syllable.

Figure 5 shows clearly the type and extent of phonetic difference that was found between the different types of geminates in this study. It gives a comparison of the means for the ratios of all the geminate consonants to their single counterparts, and it shows that the conditions which pattern together--three on the left and three on the right--do not have morpheme structure in common. That is, geminates within a word which cross only a syllable boundary (Type -C$C-) pattern with geminates which cross a word boundary but are not formed by assimilation with the definite article (Types C#C- and -C#C), while the other types of geminates within a word (Types -CC#--across no boundary--and #C+REFL+--assimilated across a morpheme boundary) as well as geminates assimilated across a word boundary with the definite article (Type ART#C-) pattern differently from these and similarly to each other. Specifically, the ratios for consonants geminated across a syllable boundary within a word were more similar to the ratios for consonants which were geminated across a word boundary than they were to other consonants which were geminated either within a word or assimilated with the definite article.

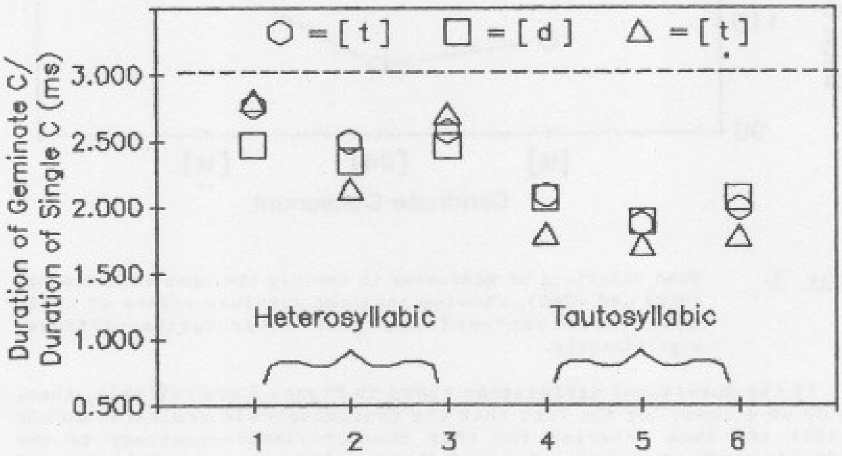


Figure 5. Ratios of the average duration of the geminate consonants to the average duration of the corresponding single consonants, as in Figure 3, but here arranged to show the values for heterosyllabic geminates on the left and tautosyllabic geminates on the right.

1 = C#C-/#C- 3 = -C$C-/-C- 5 = ART#C-/#C-
2 = -C#C/-C# 4 = #C+REFL+/#C- 6 = -CC#/-C#

This makes it obvious that the factor which conditioned the phonetic difference found in this study was not morpho-syntactic. Rather, it was phonological (more specifically, prosodic) since the three conditions which pattern together on the left in Figure 5 are heterosyllabic while the three which pattern together on the right are tautosyllabic. Specifically, crossing a syllable boundary--whether within a word or across words--resulted in a longer geminate (approximately 2.5 times the length of the single counterpart) than not crossing a syllable boundary did (approximately 2.0 times the length of the single counterpart).

Figure 5 also shows that durational differences did not distinguish tautomorphemic geminates from heteromorphemic geminates in this study since the former (Types -C$C- and -CC#) had different average durations in comparison to their single counterparts. The consonants of Type -C$C- were approximately 2.5 times the length of their single counterparts--like the other consonants which were geminated across a syllable boundary, but the consonants of Type -CC# were approximately 2.0 times the length of their short counterparts--like the other consonants which were geminated in contexts other than across a syllable boundary.

4. Conclusions

This investigation has found some evidence that for this Levantine Arabic speaker there is no significant phonetic difference between heteromorphemic and tautomorphemic geminate consonants in terms of either duration or distinctive release within the duration. This agrees with the results obtained by the previous study which specifically tested such geminates in Turkish and Bengali--Hankamer and Lahiri (1986). It also agrees with the previous phonological studies which claim that there are no phonetic differences between these types of geminates. Furthermore, the absence of a difference in these types of geminates in this investigation suggests that what have previously been termed "long consonants" in phonetic studies because they do not cross any morpho-syntactic boundary (here called "tautomorphemic geminates") should more properly be termed "geminates" as is done in phonological studies since several of them in this study had spikes on their spectrograms in the middle of their durations, as did several of the consonant clusters (here called "heteromorphemic geminates"). If confirmed by further research that these spikes are evidence of rearticulation, this would provide evidence for a phonetic argument that all of the lengthened consonants investigated here are clusters of two like segments--geminates of some sort, based on the previous studies which found rearticulation to often be a characteristic--or even the defining characteristic--of geminate consonants.

Furthermore, this investigation has found that for this speaker of Levantine Arabic there is a significant difference between heterosyllabic and tautosyllabic geminates in terms of ratio to their single counterparts, and--by implication--duration (verified for one of the sets of geminates in the study). Although to my knowledge this type of geminate structure has not been tested before, these results agree with the general claims of previous phonetic studies that there are phonetic characteristics which differentiate different types of geminates. Furthermore, these results suggest that in order to investigate whether there are phonetic differences between different types of geminates--and to discover what they are--geminates must be investigated at a different level of structure from that which they have been so far. The level of investigation must be syllable (prosodic) structure, rather than either morpheme structure or word structure--since, as described in Section 1, the phonetic differences between geminates of different morpheme or word structure claimed by some of the previous phonetic studies were not found consistently by all the studies or all the time by any of the studies. This finding confirms the hypothesis of this study that whatever phonetic differences exist consistently between geminates are likely manifested at the surface level of structure rather than at any abstract level.

This investigation has also added to our knowledge of the details of the characteristics of geminates in the areas of consonant duration and movement during the duration. The results on duration generally agreed with the results of most of the previous studies. For this speaker of Levantine Arabic all types of geminate consonants were about twice as long as the corresponding single consonants. Therefore, this study did not find evidence that heteromorphemic and tautomorphemic geminates ever differed significantly in duration, as suggested by some of the data from two previous studies (Delattre 1971; Lehiste, Morton, and Tatham 1973). However, this investigation provided new information on heterosyllabic and tautosyllabic geminates, finding that they differed in duration. The former were approximately 2.5 times the length of their single counterparts, while the latter were approximately 2.0 times the length of their single counterparts, whether or not these different syllabic types crossed morpho-syntactic boundaries.

Furthermore, since the investigation reported here found apparent release spikes on some of the spectrograms of both heteromorphemic and tautomorphemic geminates, these results on movement during consonant duration agreed with the results of the group of previous studies which suggested that there is some movement during the duration of both types of geminates (since these studies found both heteromorphemic and tautomorphemic geminates to generally have two or more peaks in their air pressure curves, articulatory pressure curves, or EMG traces) (see Section 1). The results of this study therefore did not support the group of previous studies which suggested that heteromorphemic geminates generally exhibit some movement during their durations (since they generally had two peaks in their amplitude curves and tongue shapes), while tautomorphemic geminates do not exhibit movement during their durations (since they had only one peak in these curves) (see Section 1). Furthermore, since this study suggests that both of these types of geminates may have the characteristic of movement during their durations, it shows that terming them both geminates (rather than long consonants vs. geminate consonants as many of the previous phonetic studies did) may have a phonetic basis as well as a phonological basis, since they can both apparently be interrupted during their durations. It also suggests that the individual consonants of tautomorphemic geminates and assimilated heteromorphemic geminates may be less closely joined than the phonological studies suggest since the consonant duration can be interrupted in the middle by a spike, like heteromorphemic geminates, whose individual consonants are treated phonologically as not joined to each other at all.

However, since the apparent release spikes found here occurred in only a few instances--far less frequently than the indications of movement occurred in previous studies--these can be only tentative conclusions. It may be that the consonants investigated here differ from those investigated previously or that the previous tests used more powerful measures and so could record finer movements than this test could. Other types of tests would be needed on Levantine Arabic in order to decide among these possibilities. Further research needs to be done with more speakers; all the consonants of Arabic; air pressure, amplitude, and EMG recordings; and x-ray recordings of tongue shape.

Acknowledgements

A poster based on the pilot study for this paper was presented at the 111th Meeting of the Acoustical Society of America, May 14, 1986, under the title "Long Consonants vs. Geminate Consonants in Arabic." I would like to thank Ilse Lehiste for helpful suggestions on the pilot study, and Mary Beckman for helpful suggestions on the poster presentation, the expanded study reported here, and several drafts of this paper. I would also like to thank Gina Lee for help with computer preparation of the graphs for both the pilot study and this study.

References

Brame, M. 1970. Arabic Phonology: Implications for Phonological Theory and Historical Semitic. Doctoral Dissertation. MIT. Cambridge, Mass.

Delattre, Pierre. 1971. Consonant Gemination in Four Languages: An Acoustic, Perceptual, and Radiographic Study. Part I and Part II. International Review of Applied Linguistics. Vol. IX.1: 31-52, and Vol. IX.2: 97-113.

Hankamer, Jorge, and Aditi Lahiri. 1986. The Timing of Underlying and Derived Geminates. Paper presented at the 61st Annual Meeting of the Linguistic Society of America. New York, N.Y. December 27-30.

Hayes, Bruce. 1986. Inalterability in CV Phonology. Language. Vol. 62.2: 321-51.

Hegedüs, L. 1959. Beitrag zur Frage der Geminaten. Zeitschrift für Phonetik 12: 68-106.

Josselyn, F. 1901. Études expérimentales de la phonétique italienne. La Parole. Vol II: 227-52.

Judge, Anne, and F.G. Healey. 1983. A Reference Grammar of Modern French. London: Edward Arnold Ltd.

Lehiste, Ilse, Katherine Morton, and M.A.A. Tatham. 1973. An Instrumental Study of Consonant Gemination. Journal of Phonetics 1: 131-48.

McCarthy, John J. 1979. Formal Problems in Semitic Phonology and Morphology. Doctoral Dissertation. MIT. Cambridge, Mass.

_____. 1981. A Prosodic Theory of Nonconcatenative Morphology. Linguistic Inquiry. Vol. 12.3: 373-418.

_____. 1986. OCP Effects: Gemination and Antigemination. Linguistic Inquiry. Vol. 17.2: 207-63.

Miller, Ann M. 1986. Morphological Idiosyncracies in Classical Arabic. ESCOL '86: Proceedings of the Third Eastern States Conference on Linguistics. Columbus, Oh.: The Ohio State University Linguistics Dept. Pp. 393-404.

Ollivier, Jacqueline. 1978. Grammaire française. New York, N.Y.: Harcourt Brace Jovanovich, Inc.

Rousselot, P.J. 1891. Les modifications phonétiques du langage étudiées dans le patois d'une famille de Cellefrouin. Paris: Charente.

____. 1897-1901 and 1901-1908. Principes de phonétique expérimentale. Vol. I and II. Paris: H. Welter.

Rousselot, P.J., and F. Laclotte. 1913. Précis de la prononciation française. 2nd ed. Paris: H. Welter.

Sievers, G.E. 1876. Grundzüge der Lautphysiologie. Leipzig.

____. 1893. Grundzüge der Phonetik. 4th ed. Leipzig: Breitkopf and Härtel.

Stetson, R.H. 1951. Motor Phonetics: A Study of Speech Movements in Action. Amsterdam: North-Holland Publishing Company.