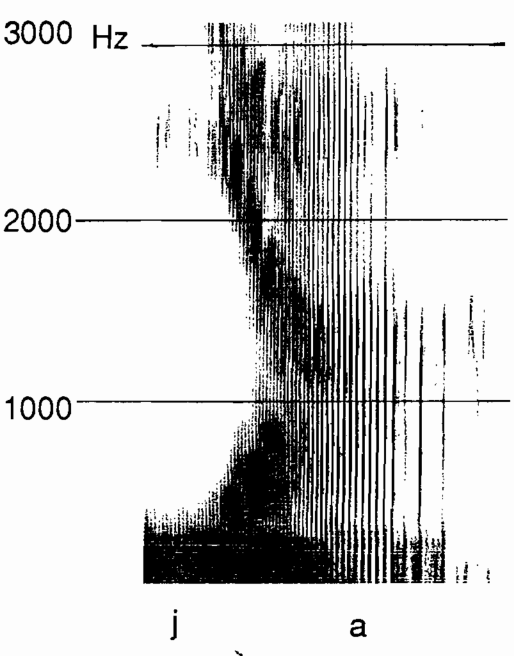
Vote for Phonetics

Up with Formant One

Down with Formant Two!



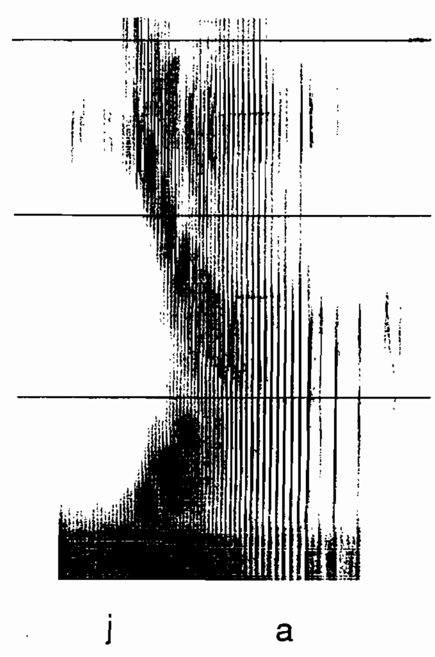
UCLA

Working Papers

in Phonetics

63

October 1986



UCLA Working Papers in Phonetics 63

October 1986

Cover of this issue inspired by a bon mot from Peri Bhaskararao.

DURATIONAL CORRELATES OF QUANTITY AND SENTENCE STRESS: A CROSS-LANGUAGE STUDY OF SWEDISH, FINNISH AND CZECH

Olle Engstrand

1 Introduction

Several languages of the world make use of the lexically sense-discriminative feature of quantity (relative length). The common phonetic denominator of quantity is generally taken to be a regular pattern of opposing segment durations. However, the feature is known to have multiple correlates (e.g. Malmberg 1944, Durand 1946:162). For example, vowel quality co-varies with length in several languages. Out of the 317 languages included in the UCLA Phonological Segment Inventory Database (UPSID; Maddieson 1984), 62 (19.6 %) display length contrasts associated with quality differences. Such a length-quality link is essential to the pronunciation of, e.g., Swedish, German and English, but less prominent in, e.g., Czech, Serbo-Croatian and Finnish (1, 2). Thus, Lehtonen (1970:87) claims that native speakers of Finnish and Czech, as opposed to native speakers of Swedish, English and German, do not use a quality difference between vowels as a cue to the phonemic length of the vowel segment (3).

Does this difference imply that lexically sense-discriminative durational alteration in these respective languages should be considered to reflect the operation of two substantially different kinds of distinctive features? While Lehtonen's claim may not be fully warranted experimentally, he is certainly not alone in drawing attention to the distinction between a "prosodic" feature of length and an "inherent" property of vowel quality as a principal aspect of the tenseness vs. laxness feature (Jones 1934, Jakobson et al. 1969, Jakobson and Halle 1969:59). Jakobson and Halle (op. cit.), in their discussion of the physical correlates of tenseness and laxness, state that "the tense vowels are necessarily lengthened in comparison with the corresponding lax phonemes. Tense vowels have the duration needed for the production of the most clear-cut, optimal vowels" (p. 58), whereas "the prosodic length of a vowel is inferred from the contrast of long and ceteris paribus short vowels in a syllabic sequence" (p. 59). Thus, this featural dichotomy bears directly upon the question of predictability vs. autonomy of length in the phonological description of language (4).

If we intuitively dichotomize Swedish vs. Finnish and Czech in these terms, by what means could such an intuition be given quantitative support? The experimental approach taken in this paper is to observe the sensitivity of durational relationships to systematic variation of sentence stress. The underlying assumption is that a relative durational invariance under sentence stress transformation would support a prosodic feature of length in the above sense of Jakobson and Halle. On the other hand, if segmental timing turns out to depend critically on stress, the mere absence of temporal stability would point in a different direction. In particular, the Swedish vowel inventory with its symmetrically organized quality pairs would in such a case be more conveniently described in terms of tenseness and laxness.

Effects of stress on segment duration have been reported to be smaller in Finnish and Czech than in English (Lehiste 1970, Lehtonen 1970, Niemi 1984, Rigault 1972, Ondráčková 1980). Lehtonen (op. cit:145 ff.), with data based on one Finnish speaker, found a moderate decrease in vowel and consonant duration as a function of a shift from what he calls strong sentence stress to weak secondary sentence stress. Ondráčková (op. cit.), also observing a single speaker, shows that Czech segment duration is determined to a considerably greater extent by phonological length than by stress. For Swedish, on the other hand, data in Elert (1964), Bannert (1979), Engstrand (1983) and Engstrand and Nordstrand (1984) suggest that the absence of stress tends to weaken or neutralize durational differences between long and short vowels.

The purpose of the present experiment is to observe the behavior of the durational correlate of quantity under sentence stress shift using some fairly comparable data from Swedish, Finnish and Czech (5). More specifically, we test the null hypothesis that durational data for all three languages will consistently reflect the quantity opposition under such a contextual transformation. To do this, we have recorded a speech material comprising sets of words that are minimally contrasting with respect to quantity. The test words appear in sentences, the stress pattern of which has been systematically manipulated. In the next few sections, we shall describe the experimental procedures (Sec. 2) and the resulting data (Sec. 3). The final section of the paper (Sec. 4) contains some concluding remarks.

2 Methods and procedures

2.1 Speech material and subjects

The present data corpus includes all possible combinations of long and short segments that may occur in lexically stressed syllables in Swedish, Finnish and Czech. This amounts to two combinations (VC and V:C) for Czech and Swedish and all four possible combinations (VC, V:C, VC:, V:C:) for Finnish (6).

The test words are given in Tables I-III. They were embedded in carrier phrases and spoken under two different stress conditions. Variations in sentence stress patterns were obtained by asking the subjects to pronounce the utterances in an alternating topic vs. comment mode: either the test word or one of the carrier words was identified as the semantic focus of the utterance (cf. Bruce 1977). All utterances were randomized and spoken several times as specified below for each subject.

The Finnish speech material is given in Table I. Note that Finnish orthography consistently indicates both vowel and consonant length by gemination. The test words were all embedded in the carrier phrase 'Lausu ___ hitaasti (I say ___ slowly). Sentence stress was either on the test word or on the carrier 'hitaasti'.

Table I. The Finnish test words in orthography.

Finnish word	English equivalent
Malinen	(a possible family name)
maalinen	adjective derivation from 'maalata' (paint)
mallinen	adjective derivation from 'malli' (model)
maallinen	terrestrial

The Finnish subjects are three females and one male. They are all native speakers of slightly differing variants of educated Standard Finnish:

Subj. MLJ (female, b. 1944) grew up in Ulvila close to the city of Pori ot the west coast of Finland. She has been living in Sweden since 1976. However, she regularly spends her summers as well as one winter month in Finland. She has a background as a high-school teacher of Finnish and now teaches Finnish at the university level.

Subj. LP (female, b. 1959) grew up close to the city of Turku on the south-west coast of Finland. At the time of the recording, she had spent 15 months in Sweden as an undergraduate student of Swedish. She speaks Swedish fluently with a slight Finnish accent.

Subj. SK (female, b. 1914) grew up in Helsinki. She is perfectly bilingual in Finnish and the variant of educated Standard Swedish spoken in Finland. She has been living in Sweden since 1949. She holds a doctorate in Finnish and has written papers on various aspects of Finnish grammar and phonology. Up to her retirement, she taught Finnish at the university level.

Subj. RJ (male, b. 1941) grew up and is resident in the city of Tampere, situated in the southern part of Finland, north-west of Helsinki, where he has spent about ten years. He is a businessman by profession. He speaks Swedish fairly fluently with a marked Finnish accent.

The Czech speech material is given in Table II. In Czech orthography, vowel quantity is indicated by means of the acute accent symbol.

Table II. The Czech test words in orthography.

Czech word	English equivalent
víla	fairy
vila	villa

These words were embedded in the carrier phrase '___ je české slovo' (___ is a Czech word) /subj. KS/ and 'já říkám že ___ je české slovo' (I say that ___ is a Czech word) /the remaining subjects/. Sentence stress shift was elicited as described above with the test word and the carrier phrase word 'české' (Czech) alternatingly in focus position.

The Czech subjects are two males and two females, all native speakers of slightly differing variants of educated Standard Czech:

Subj. KS (male, b. 1948) grew up in Prague. At the time of the recording he had spent a few months in Sweden lecturing Czech at the university level. He has an elementary knowledge of Swedish.

Subj. HKP (female, b. 1948) grew up in Prague. She has spent the past 15 years in the United States as a lecturer of Czech language and literature at the university level. She is a member of a Czech speaking community and speaks Czech at home. She visits Czechoslovakia frequently. During those visits, native speakers of Czech generally do not notice any foreign influence on her pronunciation. A non-native accent is noticeable in her English, which she otherwise speaks fluently.

Subject KV (female, b. 1956) and subj. MA (male, b. 1948) are a married couple. They have resided in the United States for the past four years. Subj. MA teaches Czech at the university level; subj. KV does this on a less regular basis. Subj. MA grew up in Moravia and now speaks Standard Czech with a slight Moravian accent. On the whole, they both speak Czech more often than English. Their English is fluent but colored by a non-native accent.

The Swedish test words are shown in Table III (7). In Swedish orthography, the complementary length pattern (cf. footnote 5) is generally notated by means of consonant gemination.

Table III. The Swedish test words in orthography.

Swedish word	English equivalent
vila	rest
villa	villa

These words were embedded in the carrier phrase 'Sin ___ fick

Pelle betala för' which is a topicalized version of a sentence meaning 'Pelle had to pay for his ___'. Sentence-focal stress was assigned to yield the following four versions:

1 Sin 'vila fick Pelle betala för.
2 Sin 'villa fick Pelle betala för.
3 'Sin vila fick 'Pelle betala för.
4 'Sin villa fick 'Pelle betala för.

The double stress assignment was made in order to efficiently remove stress from the test words in (3) and (4). However, subj. LA preferred a version with a single sentence stress; therefore, only the word 'betala' (pay) is in sentence-focal position for this subject.

With one exception (TF, see below), the Swedish subjects were described in Engstrand (1983). They are all males, ranging between 25 and 54 years of age at the time of the recording. They are all speakers of educated Central Standard Swedish. Five out of the six subjects are native speakers, whereas the remaining subject (FF) has a German background. At twelve years of age, he moved to Sweden and now remains bilingual in Swedish and German; at the time of the recording he was 54 years of age. Unlike the other subjects, he has a slight Southern Swedish accent.

Subj. TF (male, b. 1957) is a native speaker of Central Standard Swedish. He grew up in Stockholm, where he graduated from Business School. At the time of the recording, he had spent five weeks in the United States.

2.2 Experimental procedures

The recordings of the Swedish subject TF and the Czech subjects HKP, KV and MA were made at the UCLA Phonetics Lab using a Tascam 122 cassette tape recorder. The remainder of the subjects were recorded at the Uppsala University Phonetics Lab with a Telefunken Magnetophone M/12 reel-to-reel tape recorder. All subjects were seated comfortably in sound-proof surroundings about 20 cm from the microphone. The operator and the recording equipment were outside the recording room. The operator monitored the speech signal over head-phones and VU-meter.

The test utterances were either read from cards (the Swedish subjects except TF, all Finnish subjects, and the Czech subjects except KS) or from lists (the remainder of the subjects). First, the experimenter read a number code corresponding to the word to be spoken by the subject; the subject then read the word embedded in the appropriate carrier phrase; the experimenter read the number code corresponding to the next word, and so forth.

Spectrograms were made of the entire set of utterances by means of a Kay Electric Digital Sonagraph (subjects TF, HKP, KV and MA) and a Voice Identification, Inc., Series 700 Sound Spectrograph (the remaining subjects). The Kay Electric Digital Sonagraph time scale was expanded in order to enhance the time resolution for durational measurements. In those measurements, 1 mm on the spectrogram paper corresponds to 4.03 msec. On the spectrograms produced with the Voice Identification Sound Spectrograph, 1 mm corresponds to 7.84 msec. Durations were measured to the nearest 0.5 mm relative to the selected segment boundaries. Those boundaries were drawn at acoustic discontinuities exemplified in the below spectrogram (Fig. 1). Let us now turn to the results of the durational measurements.

3 Results

The measurement data are presented numerically and graphically. The diagrams (Figures 2-5) are introduced along with the body of text. The numerical tables form Appendices, which have been organized by language: App. 1 for Swedish, App. 2 for Czech and App. 3 for Finnish; the Tables comprised in the Appendices are numbered A1:I, A1:II, A2-I, ..., A3-VI.

Swedish

We shall first present means and standard deviations for the vowel and consonant durations in 'vila' and 'villa' as spoken by the six Swedish subjects under two stress conditions. Those data are listed in Table A1- I and the means are displayed graphically in Fig. 2. Table IV below summarizes t-values pertaining to differences between means (for the Swedish as well as the Czech and Finnish data); differences meeting the 0.5 % criterion for statistical significance (one-tailed t-test) are indicated by asterisks.

Under the stressed speaking condition, all six subjects display the expected complementary VC durational distribution; i.e., the long vowel in 'vila' precedes a relatively short consonant, and the short vowel in 'villa' precedes a relatively long consonant. In the unstressed version, the difference between durational means for the long vs. the short vowels fails to reach the 0.5 % level of statistical significance in three out of six subjects; the corresponding difference between the long and the short consonants also fails to satisfy this condition in three subjects. In particular, two subjects distinguish neither vowel nor consonant duration in the quantity pair 'vila' vs. 'villa'. On the other hand, effects of stress are always significant in the vowels and generally in consonants after short vowels.

Czech

The durational data pertaining to Czech are given in Table A2-I and Fig. 3. Under both stress conditions, the long i-vowel has a significantly greater duration than the short i-vowel. An accompanying consonant lengthening effect is significant in one

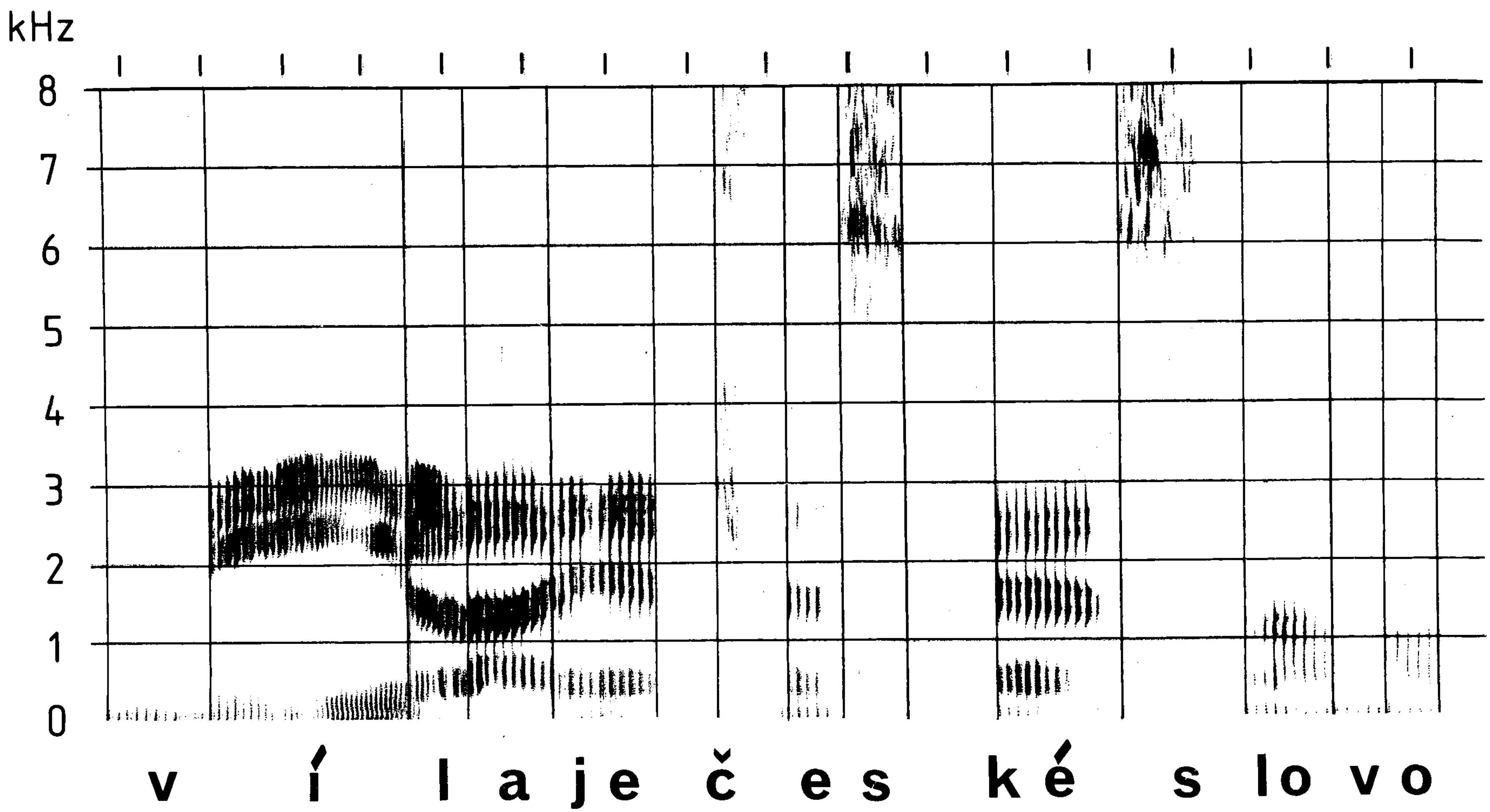


Fig. 1. Broad-band spectrogram illustrating the Czech sentence Víla je české slovo /vi:la je tʃeske: slovo/ pronounced by subj. KS with the word víla in sentence focus. Vertical lines have been drawn at acoustic discontinuities roughly corresponding to the accompanying phonemic transcription. The time distance between tic marks (top) is 100 msec.

Table IV. t-values (cells) and number of degrees of freedom (Col. 1) for a one-tailed t-test of statistical significance for differences between durational means in vowels and consonants spoken by the Swedish, Czech and Finnish subjects (Col. 2). The variables are quantity (Cols. 3 and 4) and stress (Cols. 5 and 6). Durational differences satisfying the 0.5 % criterion for statistical significance are indicated by asterisks.

VARIABLE: 1	2	QUANTITY 3		QUANTITY 4		STRESS 5		STRESS 6	
		'V:C/'VC(:)		V:C/VC(:)		'V:C/V:C		'VC(:)/VC(:)	
		vow	cons	vow	cons	vow	cons	vow	cons
df= 18	OE (Sw)	17.44 *	-15.13 *	-0.89 (n.s.)	-0.44 (n.s.)	26.43 *	3.79 *	8.04 *	20.00 *
df= 26	CN (Sw)	11.63 *	-20.14 *	1.06 (n.s.)	-0.25 (n.s.)	15.00 *	0.84 (n.s.)	9.76 *	12.78 *
df= 8	TF (Sw)	17.22 *	-9.5 *	3.15 (n.s.)	-4.16 *	18.19 *	0.63 (n.s.)	4.29 *	8.73 *
df= 22	FF (Sw)	16.81 *	-17.85 *	6.00 *	-2.83 (n.s.)	5.50 *	4.03 *	6.19 *	9.93 *
df= 32	GI (Sw)	6.94 *	-8.85 *	5.06 *	-5.16 *	4.83 *	-0.79 (n.s.)	4.53 *	2.89 (n.s.)
df= 30	LA (Sw)	22.59 *	-26.29 *	7.73 *	-3.05 *	21.70 *	7.61 *	10.30 *	18.87 *
df= 12	KS (Cz)	11.71 *	0.92 (n.s.)	7.74 *	1.14 (n.s.)	4.06 *	2.80 (n.s.)	2.95 (n.s.)	1.65 (n.s.)
df= 10	HKP (Cz)	16.13 *	2.10 (n.s.)	26.85 *	-0.77 (n.s.)	6.57 *	6.79 *	4.86 *	0.94 (n.s.)
df= 12	KV (Cz)	30.14 *	2.12 (n.s.)	8.49 *	0.73 (n.s.)	9.68 *	9.98 *	1.40 (n.s.)	7.69 *
df= 10	MA (Cz)	14.35 *	6.86 *	11.26 *	-1.36 (n.s.)	9.83 *	13.69 *	4.78 *	5.20 *
df= 12	RJ (Fi)	21.78 *	-11.04 *	9.07 *	-9.80 *	6.69 *	0.42 (n.s.)	0.98 (n.s.)	0.00 (n.s.)
df= 16	MLJ (Fi)	14.77 *	-15.73 *	10.43 *	-8.80 *	-0.51 (n.s.)	-0.29 (n.s.)	-2.68 (n.s.)	2.20 (n.s.)
df= 8	LP (Fi)	13.44 *	-11.13 *	12.63 *	-7.44 *	7.15 *	-0.77 (n.s.)	2.35 (n.s.)	9.49 *
df= 18	SK (Fi)	28.87 *	-27.23 *	11.14 *	-9.72 *	5.48 *	0.00 (n.s.)	5.92 *	2.83 (n.s.)

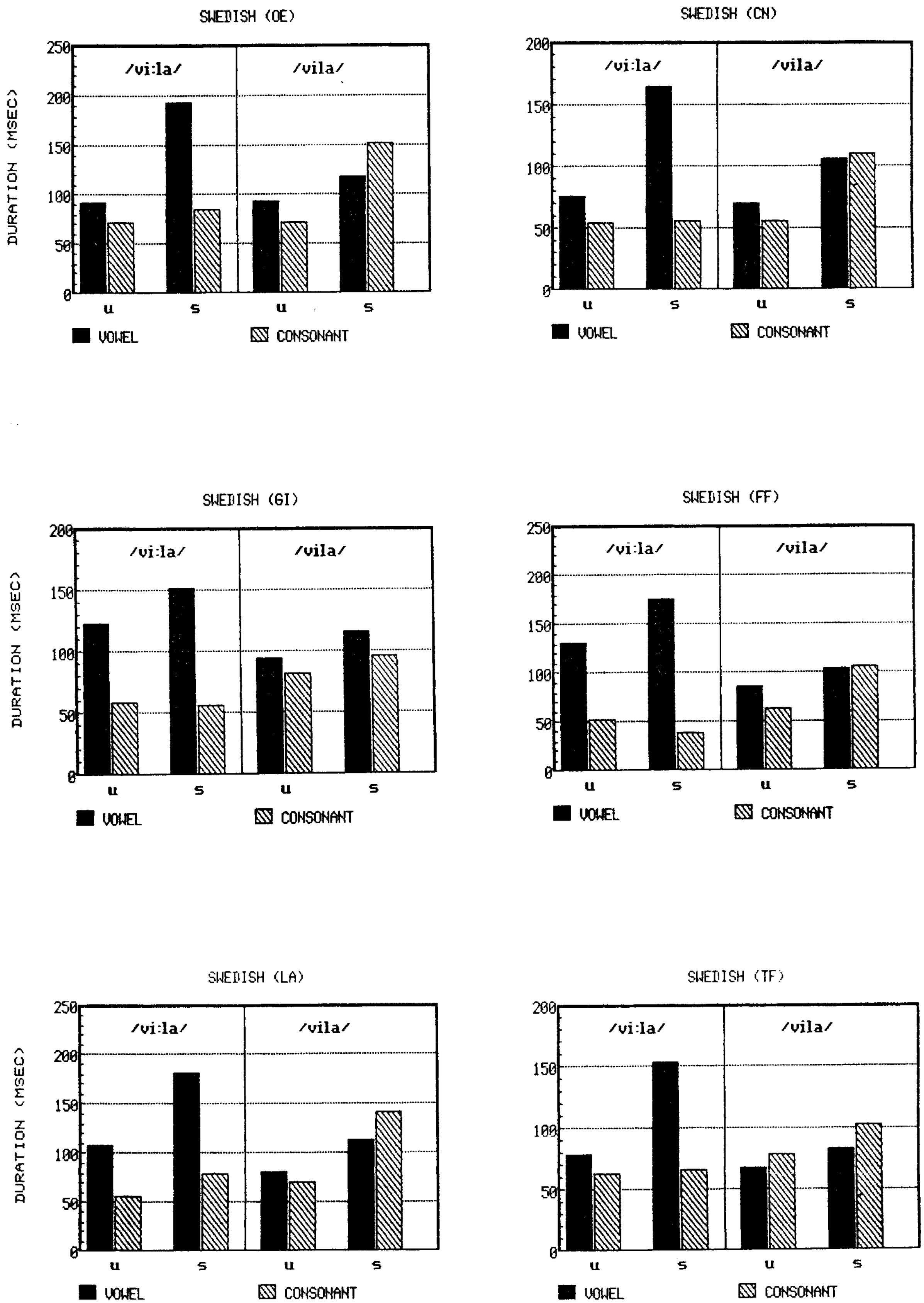


Fig. 2. Mean vowel and consonant durations for the Swedish words <u>vila</u> /vi:la/ and <u>villa</u> /vila/ as pronounced by six subjects under two different sentence stress conditions; 'u' and 's' below the diagrams stand for the unstressed and stressed condition, respectively.

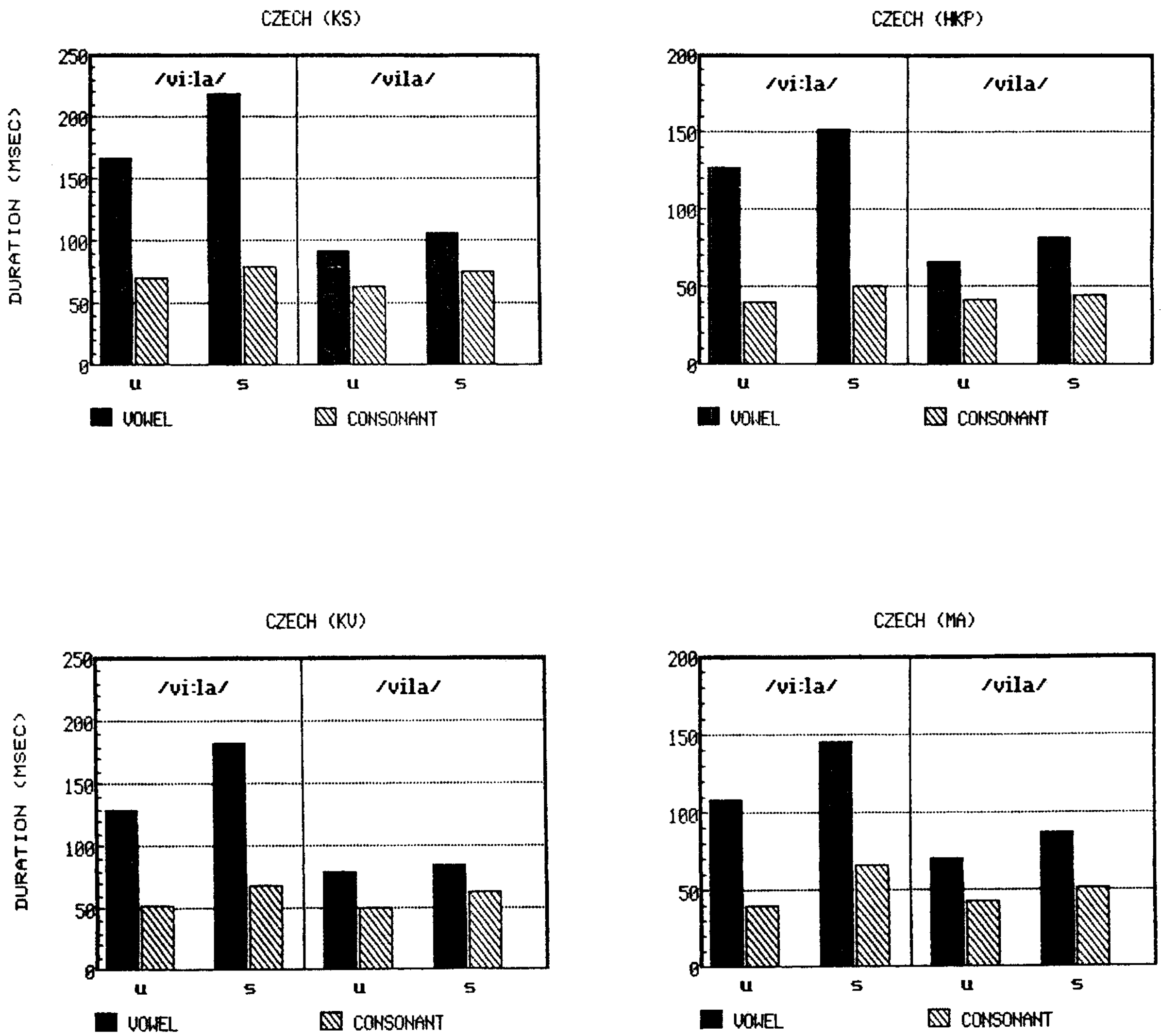


Fig. 3. Mean vowel and consonant durations for the Czech words víla /vi:la/ and vila /vila/ as pronounced by four subjects under two different sentence stress conditions; 'u' and 's' below the diagrams stand for the unstressed and stressed condition, respectively.

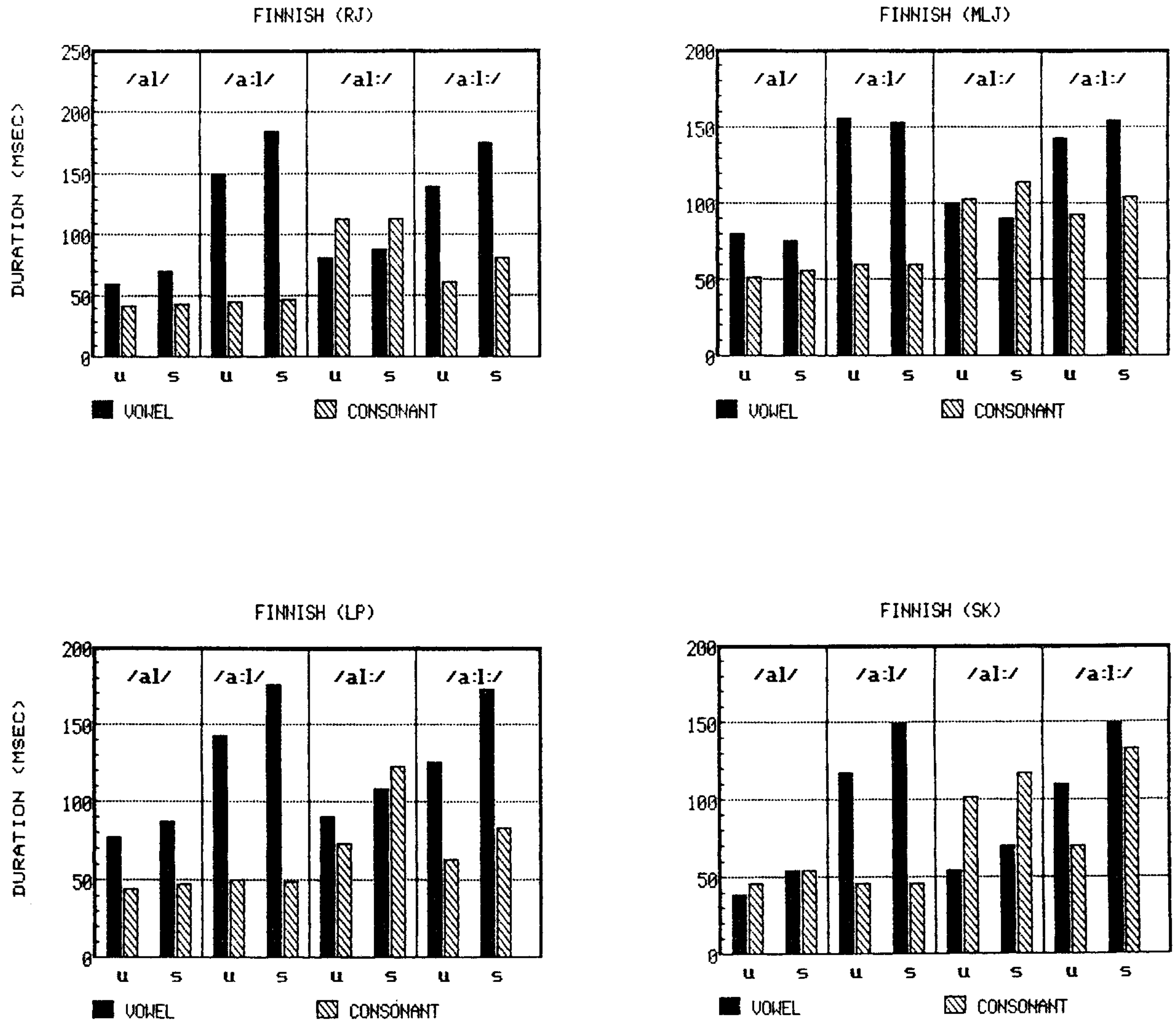


Fig. 4. Mean vowel and consonant durations for the Finnish words Malinen /malinen/, maalinen /ma:linen/, mallinen /mal:inen/ and maallinen /ma:l:inen/ as pronounced by four subjects under two different sentence stress conditions; 'u' and 's' below the diagrams stand for the unstressed and stressed condition, respectively.

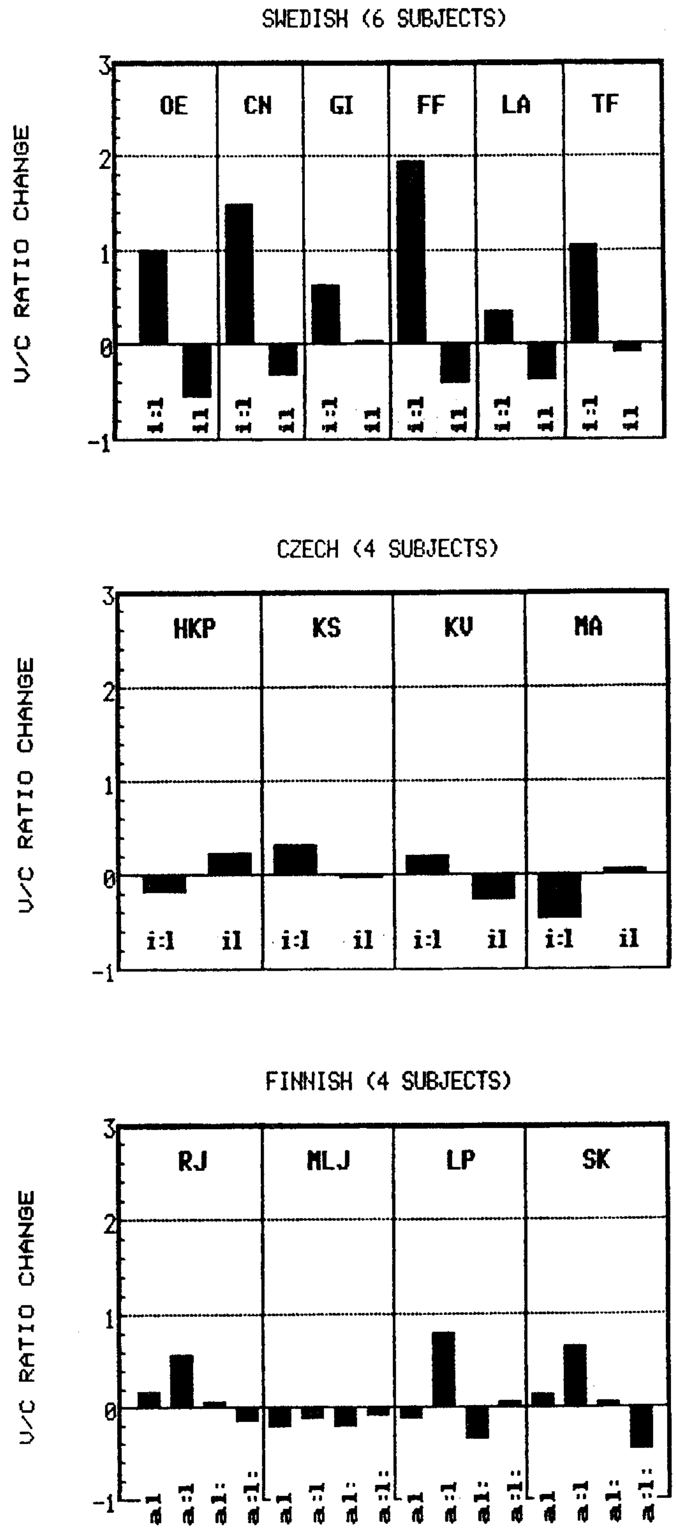


Fig. 5. Changes in ratio between vowel and consonant durations as a function of sentence stress on the Swedish, Czech and Finnish test words.

case for the stressed version; in the remaining cases, and for the unstressed version, the effect on the consonants is not significant. Consequently, there is no complementary connection between V and C durations. Stress alteration leads to significant durational effects mainly in the V:C sequence; in the VC sequences, stress-conditioned lengthening is weaker and frequently non-significant at the 0.5 % risk level. On the whole, durational effects of stress tend to affect the entire VC sequence rather than either one of the segments.

Finnish

The results pertaining to Finnish are shown in Tables A3-I,II,III,IV and Fig. 4. The durational data are fairly straightforward, consistently reflecting quantity regularities such that phonologically long and short vowels and consonants are kept significantly distinct under both stress conditions. Effects of stress are relatively modest and frequently non- significant at the 0.5 % level (exemplified in Table IV for the words 'maalinen' and 'mallinen').

Durational V/C ratios

Vowel-consonant ratios (V/C) are shown as a function of speaker, quantity and stress in Tables A1-II, A2-II and A3-V for Swedish, Czech and Finnish, respectively. In these tables, V/C=1 means that the vowel and the consonant are equally long; V/C<1 means that the consonant is longer than the vowel, and V/C>1 means that the vowel is longer than the consonant.

As expected from the above data, the Swedish ratios display a considerable variation as a function of stress such that the V:/C vs. V/C ratios differ more under the stressed than under the unstressed speaking condition (Table A1-II). In particular, the V:/C ratios are strongly affected by stress shift. By comparison, stress effects are relatively weak in the Czech and the Finnish data (Tables A2-II and A3- V, respectively). On the other hand, the durational effect of quantity is more consistently preserved in the latter two languages than in Swedish. In Finnish, the members of the apparently exceptional VC vs. V:C: pairs are clearly distinguished by the duration of the entire vowel-consonant sequence (cf. Tables A3-I,II,III,IV). The stress-induced ratio changes are depicted for all three languages in Fig. 5, based on Tables A1-III, A2-III and A3-VI.

4 Concluding remarks

In the present study, we have put to trial the assumption that durational data for Swedish, Finnish and Czech would consistently reflect the prosodic feature of quantity across variation in sentence stress. Relative duration was thus conceived as an independent variable in the sense of possessing sufficient integrity to resist destructive influences of secondary context. We now find, however, that the experiment did not fully corroborate this assumption as far as the Swedish data are concerned. Rather, the Swedish speakers largely failed to exhibit a consistent durational dichotomy under the given stress conditions. In this respect, then, the Swedish linguistic behavior was singular among the examined languages.

As pointed out above (Sec. 1), most Swedish vowel pairs said to differ in length also differ in quality. Do spectral criteria provide a more reliable boundary in the Swedish sound space than the durational ones (8)? Data in Engstrand (1983), involving most of the present speakers, and Engstrand and Nordstrand (1984) suggest that sentence stress alteration does preserve a relatively invariant spectral correlate. Also, Engstrand (1983) showed in a combined X-ray and acoustic study of the long (tense) vowels /i a u/ that articulatory as well as spectral configurations remained relatively invariant under changes in speech rate, whereas temporal patterns changed considerably. Those results parallel data for American English (Gay 1979; but cf. Lindblom 1963, 1964). Moreover, Engstrand (1983) found that stress shift produced a significant spectral sharpening effect on the acuteness, compactness and gravity of the respective tense vowels. Those effects were interpreted as a means of giving perceptual salience to the inherent qualities of those vowels. On the other hand, the corresponding spectral effect of stress was frequently not significant in the lax vowels. These results lead to the hypothesis that spectral sharpening is restricted to the tense vowels in Swedish. Such a state of affairs would, of course, further enhance the perceptual efficiency of the alleged tensity feature.

The data emerging from the present experiment are difficult to interpret in some respects. For example, two of the Swedish speakers (GI and LA) were peculiar in displaying completely consistent durational differences between the vowel and consonant segments in the unstressed versions of 'vila' and 'villa'. This behavior might be conditioned by factors such as a tense vs. lax alteration (Jakobson and Halle 1969) or a failure to control sentence stress experimentally. The data available offer no reliable means to solve this problem; thus, the present experiment highlights the need for refined models of interarticulatory programming and phonetic parameter interdependencies (9).

FOOTNOTES

1. See, for example, the following sources. For Swedish: Malmberg (1944:37 f.), Elert (1964:12 ff.); for German: Straka (1959:280 f.); for English: Straka (1959:276), Delattre (1962) and Ladefoged et al. (1972); for Czech: Chlumský (1928:ix), Straka (1959:281 f.), Lehiste (1970:30 ff.), and Hála, B. (1923): K popisu pražské výslovnosti [Contribution to the description of the Prague

pronunciation], quoted by Straka (1959:282); for Serbo-Croatian: Lehiste (1970:30 ff.); and for Finnish: Sovijärvi, (1938:24) and Wiik (1965:59).

2. When differences in timbre go with quantity, short vowels tend to be more centralized than their long counterparts (e.g. Durand 1946:152, Jakobson et al. 1963:36, Fant 1960:210). However, the way centralization affects different vowels in an inventory is language- specific. For Finnish, Sovijärvi (1938:24) points out minor differences between long vs. short /i y e o/; Wiik's (1965:56 ff.) acoustic data suggest that /i/ and /e/ are affected more than the remaining vowels. For a Czech speaker, Lehiste (1970:30 f.) shows spectral differences to be considerably greater for long and short /i/ and /u/ than for the remaining vowel quantity pairs (long and short /e a o/). Swedish has quite considerable differences in most pairs (Elert 1964:12 ff.). The most notable exceptions are long and short /ɛ/ and /ø/ including the contextually restricted open allophones [æ] and [œ̞].

3. Perceptual cues to quantity perception are further discussed in Sec. 4 of this paper.

4. The question of predictability vs. autonomy of length has been raised repeatedly in the phonetic literature; cf., for example, Jespersen's (1926) distinction between "äusserlich" and "innerlich" determined quantity. However, the causal connection between quality and length has been approached from varying viewpoints. For example, Panconcelli-Calzia, discussing the formation of vowels, suggests that "man in eine Stellung, die eine gewisse Zeit gehalten werden muss, mit Schwung und Kraft geht; in eine Stellung, die nur kurze Zeit dauern soll, geht man dagegen ohne grössere Anstrengung aus Rucksicht auf die kurz darauf zu leistende Arbeit, die Nötig ist, um diese Stellung zu verlassen" (Die experimentelle Phonetik in ihrer Anwendung auf die Sprachwissenschaft 1924:96 ff., quoted by Straka 1959:278). A similar interpretation was offered by Mazlová: "Au cours de l'émission d'une voyelle longue, les organes ont plus de temps pour occuper exactement la position typique caractérisant ce phonème" (Výslovnost na Zábřežku, fonetická studie z moravské dialektologie [The pronunciation at Zábřeh, a phonetic study in Moravian dialectology] 1949:83-84.

5. Quantity functions independently of lexical stress in Finnish and Czech, whereas Swedish quantity is linked to lexical stress. In Swedish, the durational aspect of quantity displays a complementary pattern in lexically stressed VC sequences; a long vowel is followed by a short consonant, and a short vowel is followed by a long consonant (Elert 1964). In Finnish, as well as in other Finno-Ugric languages such as Lappish and Hungarian, all four possible combinations occur in VC sequences; Lappish even has the potential of six combinations, since there is a grammatically conditioned three-way length opposition for the consonants. Long vowel segments in Finnish are sometimes considered as geminates (cf. Sovijärvi 1956, Lehiste 1965); this analysis makes the long vowels structurally equivalent to the large inventory of diphthongs (cf. Karlsson 1976:12 ff.). In Czech, quantity alteration at the lexical level is restricted to the vowels. Jakobson (1962:626) states that "the lengthening of Czech consonants plays an emphatic

role and serves, furthermore, as a secondary component of the stress".

6. Czech and Finnish words carry lexical stress on the first syllable. Swedish words generally carry lexical stress on the first syllable of the stem; the Swedish test words used in the present experiment carry lexical stress on the first syllable.

7. Part of the Swedish speech material and data were described in Engstrand (1983). For the sake of the cross-language comparison, they will be repeated in relevant detail here. The transcription of length used for the present Swedish material represents one of several possibilities. For a discussion of alternative notations, see Elert (1970:54-60).

8. Perceptual data available are inconclusive on this point. Thus, data by Hadding-Koch and Abramson (1964) suggest that Swedish listeners may use either spectral or durational cues to the quantity affiliation of a given vowel; given a clear spectral distinction, phonologically long (or tense) and short (or lax) vowels, e.g. [ʉ:] and [ɵ], were identified correctly even in the absence of the durational distinction required by qualitatively more similar cognates, e.g. [ø:] and [œ]. Similar studies, yielding partly incompatible results, have been carried out on quantity perception in English and German (Bennet 1968, Delattre and Hohenberg 1968, Heike 1970, Sendlmeier 1981). I am not aware of any analogous experiments involving Finnish or Czech. However, Czech long and short /a/ and /e/, which are relatively similar in quality (Lehiste 1970:30 f.), have been shown to bear a clear-cut relationship between duration and judgments of quantity by listeners (Janota and Jančák 1970, Janota and Ondráčková 1975).

9. Besides phonological length, particularly emphasized by Malmberg (1944), Elert (1964) and Bannert (1979a,b), suprasegmental and grammatical factors that condition segment duration include word and phrase length (e.g. Lindblom et al. 1981, Lehtonen 1970, 1974, Iivonen 1974, Chlumský 1928), word and phrase boundaries (e.g. Cooper 1976), various fundamental frequency requirements (e.g. Lyberg 1981, Bruce 1981, Bannert 1982), and stress (e.g. Öhman et al. 1979). For example, Öhman et al. (1979) attempted to derive the complementary distribution of length in Swedish from assumptions about syllabification and gestural coproduction.

ACKNOWLEDGMENTS

This work was supported in part by the Swedish Council for Research in the Humanities and the Social Sciences (HSFR) under grant F 734/85. I am grateful to Peter Ladefoged, Mona Lindau and Ian Maddieson for generous support during the course of this work. I am also indebted to Abby Cohn, Karen Emmorey, Bruce Hayes, Marie Huffman, Michel Jackson and Pat Keating for generous help during my daily lab work. Una Cunningham-Andersson has corrected my English language errors. I thank them all warmly.

REFERENCES

Bannert, R. 1979a. Ordstruktur och prosodi [Word structure and prosody]. In Hyltenstam, K. (Ed.): Svenska i invandrarperspektiv [Swedish from an immigrant perspective]. Lund: Liber Läromedel, 132-173.

Bannert, R. 1979b. The effect of sentence accent on quantity. In Proceedings of the 9th International Congress of Phonetic Sciences, Copenhagen 1979, Vol. II, 253-259. Institute of Phonetics, University of Copenhagen.

Bannert, R. 1982. An F0-dependent model for duration? Reports from Uppsala University, Department of Linguistics (RUUL) 8, 59-80.

Bennet, D. C. 1968. Spectral form and duration as cues in the recognition of English and German vowels. Language and Speech 11, 65-85.

Bruce, G. 1977. Swedish word accent in sentence perspective. Travaux de l'institut de linguistique de Lund XII. Lund: Gleerup.

Bruce, G. 1981. Tonal and temporal interplay. Working Papers 21, 49-60. Dept. of Linguistics, Lund University,

Chlumský, J. 1928. Česká kvantita, melodie a přízvuk (La quantité, la mélodie et l'accent d'intensité en tchèque). C.A.V.U., Prague, 367 pp.

Cooper, W. E. 1976. Syntactic control of timing in speech production: a study of complement clauses. Journal of Phonetics 4, 151-171.

Delattre, P. 1962. Some factors of vowel duration and their cross-linguistic validity. Journal of the Acoustical Society of America 34, 1141-1142.

Delattre, P. and M. Hohenberg. 1968. Duration as a cue to the tense/lax distinction in German unstressed vowels. International Review of Applied Linguistics in Language Teaching (IRAL) 6, 367-390.

Durand, M. 1946. Voyelles longues et voyelles brèves: essai sur la nature de la quantité vocalique. Collection linguistique, Soc. Ling. Paris, XLIX, Klincksieck, Paris.

Elert, C.-C. 1964. Phonologic studies of quantity in Swedish. Uppsala: Almqvist & Wiksell.

Elert, C.-C. 1970. Ljud och ord i svenskan. [Sounds and words in Swedish.] Stockholm: Almqvist & Wiksell.

Engstrand, O. 1983. Articulatory coordination in selected VCV utterances: A means-end view. Reports from Uppsala University, Department of Linguistics (RUUL) 10.

Engstrand, O. and L. Nordstrand. 1984. Acoustic features correlating with tenseness, laxness and stress in Swedish: Preliminary observations. In C.-C. Elert, I. Johansson and E. Strangert (Eds.): Nordic Prosody III. Papers from a Symposium. Acta Univ. Um., Umeå Studies in the Humanities 59, pp. 51-66.

Fant, G. 1970. Acoustic Theory of Speech Production (second edition). The Hague: Mouton.

Gay, T. 1981. Mechanisms in the control of speech rate. Phonetica 38, 148-158.

Hadding-Koch, K. and A. S. Abramson. 1964. Duration versus spectrum in Swedish vowels: Some perceptual experiments. Studia Linguistica 18, 2, 94-107.

Heike, G. 1970. Lautdauer als Merkmal der wahrgenommenen Quantität, Qualität und Betonung im Deutschen. Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, Sept. 1967. Prague: Acad. Publ. House of the Czechoslovak Academy of Sciences, 433-437.

Iivonen, A. 1974. On the dependence of sound duration on word length in isolated words. Virittäjä 78, 150-151. (English summary of paper written in Finnish.)

Jakobson, R. 1962. On the elimination of long consonants in Czech. In R. Jakobson, Selected Writings I: Phonological Studies. 's-Gravenhage: Mouton, pp. 626-627.

Jakobson, R., G. Fant and M. Halle. 1969. Preliminaries to Speech Analysis. The Distinctive Features and their Correlates. Cambridge: The M.I.T. Press. (Ninth Printing.)

Jakobson, R. and M. Halle. 1969. Tenseness and laxness. Suppl. to R. Jakobson, G. Fant and M. Halle: Preliminaries to Speech Analysis. The Distinctive Features and their Correlates. Cambridge: The M.I.T. Press, pp. 57-61. (Ninth Printing.)

Janota, P. and P. Jančák. 1970. An investigation of Czech vowel quantity by means of listening tests. Acta Univ. Carol., Philologica 1, Phonetica Pragensia II, 31-68.

Janota, P. and J. Ondráčková. 1975. Some experiments on the perception of prosodic features in Czech. In G. Fant and M.A.A. Tatham (Eds.): Auditory Analysis and Perception of Speech. London: Academic Press, pp. 485-496.

Jespersen, O. 1926. Lehrbuch der Phonetik (fourth edition). Berlin.

Jones, D. 1934. English Phonetics (fourth edition). Leipzig.

Karlsson, F. 1976. Finskans struktur. [The structure of Finnish.] Lund: Liber Läromedel.

Ladefoged, P., J. DeClerk, M. Lindau and G. Papçun. 1972. An auditory-motor theory of speech production. UCLA Working Papers in Phonetics 22, 48-75.

Lehiste, I. 1965. The function of quantity in Finnish and Estonian. Language 41, 3, 447-456.

Lehiste, I. 1970. Suprasegmentals. Cambridge, Mass.: The M.I.T. Press.

Lehtonen, J. 1970. Aspects of quantity in Standard Finnish. Studia Philologica Jyväskyläensia VI. Jyväskylä: K. J. Gummérus.

Lehtonen, J. 1974. Word length and sound durations. Virittäjä 78, 160. (English summary of paper written in Finnish.)

Lindblom, B. 1963. Spectrographic study of vowel reduction. Journal of the Acoustical Society of America 35, 1773-81.

Lindblom, B. 1964. Articulatory activity in vowels. Speech Transmission Laboratory Quarterly Progress and Status Report 2, Royal Institute of Technology, Stockholm, 1-5.

Lindblom, B., B. Lyberg and K. Holmgren. 1981. Durational Patterns of Swedish Phonology: Do They Reflect Short-Term Memory Processes? Indiana University Linguistics Club. Bloomington, Indiana.

Lyberg, B. 1981. Temporal properties of spoken Swedish. Monographs from the Institute of Linguistics, University of Stockholm (MILUS), 6.

Maddieson, I. 1984. Patterns of Sounds. Cambridge: Cambridge University Press.

Malmberg, B. 1944. Die Quantität als phonetisch-phonologischer Begriff. Eine allgemeinsprachliche Studie. Lund: Gleerup.

Niemi, J. 1984. Word level stress and prominence in Finnish and English. Publications in the Humanities, 1, University of Joensuu.

Öhman, S., S. Zetterlund, L. Nordstrand and O. Engstrand. 1979. Predicting segment durations in terms of a gesture theory of speech production. In Proceedings of the 9th International Congress of Phonetic Sciences, Copenhagen 1979, Vol. II, 305-311. Institute of Phonetics, University of Copenhagen.

Ondráčková, J. 1980. Stress and vowel quality in Czech. Acta Univ. Carol., Philologica 2, Phonetica Pragensia VI, 123-5.

Rigault, A. 1972. Accent et démarcation en tchèque. Acta Univ. Carol., Philologica 1, Phonetica Pragensia III, 207-19.

Sendlmeier, W. F. 1981. Der Einfluss von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen. Phonetica 38, 291-308.

Sovijärvi, A. 1938. Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache. Physiologisch-physikalische Lautanalysen. Helsinki.

Sovijärvi, A. 1956. Über die phonetischen Hauptzüge der finnischen und der ungarischen Hochsprache. Wiesbaden: Ural-Altaische Bibliothek II.

Straka, G. 1959. Durée et timbre vocaliques. Zeitschrift für Phonetik und allgemeine Sprachwissenschaft 12:276-300.

Wiik, K. 1965. Finnish and English vowels. Turku: Annales Universitatis Turkuensis. Series B, Tom 94.

Appendix 1: Swedish

Table A1-I. Mean durations and standard deviations (msec) of i- and l-segments in Swedish unstressed (-str) and stressed (+str) /vi:la/ and /vila/ spoken in a carrier phrase. Six subjects.

Subject:		OE		CN		GI		FF		LA		TF	
	n=	10+-2		14+-1		17+-3		12+-3		16+-1		5	
Segment:		V	C	V	C	V	C	V	C	V	C	V	C
/i:l/ +str	x	193	84	165	57	152	56	175	39	181	80	154	66
	s	11	8	18	4	17	14	12	5	9	5	7	8
/il/ +str	x	118	153	106	110	116	97	105	106	113	142	83	104
	s	8	12	6	9	13	13	8	12	8	8	6	4
/i:l/ -str	x	92	72	76	55	123	59	131	52	108	57	79	63
	s	5	6	13	8	18	7	25	10	10	11	6	7
/il/ -str	x	94	73	71	56	95	82	86	63	82	70	68	79
	s	5	4	12	13	14	17	7	9	9	13	5	5

Table A1-II. Vowel/consonant duration ratios between i- and l-segments in the Swedish test words. Six subjects.

Subject:	OE	CN	GI	FF	LA	TF
/vi:la/ (+str)	2.30	2.89	2.71	4.49	2.26	2.33
/vila/ (+str)	0.77	0.96	1.20	0.99	0.80	0.80
/vi:la/ (-str)	1.28	1.38	2.08	2.52	1.89	1.25
/vila/ (-str)	1.29	1.27	1.17	1.37	1.17	0.86

Table A1-III. Change in V/C ratio as a function of stress in the Swedish test words. Six subjects.

Subject:	OE	CN	GI	FF	LA	TF
/vi:la/	+1.02	+1.51	+0.63	+1.96	+0.37	+1.08
/vila/	-0.52	-0.31	+0.04	-0.38	-0.37	-0.06

Table A2-I. Mean durations and standard deviations (msec) of i-and l-segments in Czech stressed and unstressed /vi:la/ and /vila/ spoken in a carrier phrase under two different sentence stress conditions. Four subjects.

Subject:		KS		HKP		KV		MA	
n=		7		6+-1		7+-2		6+-1	
Segment:		V	C	V	C	V	C	V	C
/vi:la/ +Str	x	218	80	152	50	183	68	146	66
	s	23	8	8	3	7	3	7	4
/vila/ +Str	x	107	75	82	45	85	64	88	52
	s	10	12	7	5	5	4	7	3
/vi:la/ -Str	x	167	70	128	40	129	52	109	41
	s	24	5	4	2	13	3	6	2
/vila/ -Str	x	92	64	66	42	80	51	70	43
	s	9	13	4	6	8	2	6	3

Table A2-II. Vowel/consonant duration ratios between i-and l-segments in the Czech test words. Four subjects.

Subject:	HKP		KS		KV		MA	
	+Str	-Str	+Str	-Str	+Str	-Str	+Str	-Str
/vi:la/	3.04	3.20	2.73	2.39	2.69	2.48	2.21	2.66
/vila/	1.82	1.57	1.43	1.44	1.33	1.57	1.69	1.63

Table A2-III. Change in V/C ratio as a function of stress in the Czech test words. Four subjects.

Subject:	HKP	KS	KV	MA
/vi:la/	-0.16	+0.34	+0.21	-0.45
/vila/	+0.25	-0.01	-0.24	+0.06

Appendix 3: Finnish

Table A3-1. Mean durations and standard deviations (msec) of a- and l-segments in the Finnish words Malinen, maalinen, mallinen and maallinen spoken in a carrier phrase under two different sentence stress conditions, n=7. Subj. RJ.

		V		C	
		x	s	x	s
Malinen	Stressed	71	6	44	8
	Unstressed	60	14	42	4
maalinen	Stressed	184	10	47	9
	Unstressed	150	9	45	9
mallinen	Stressed	88	6	113	13
	Unstressed	81	18	113	16
maallinen	Stressed	175	15	81	9
	Unstressed	140	11	61	8

Table A3-II. Mean durations and standard deviations (msec) of a- and l-segments in the Finnish words Malinen, maalinen, mallinen and maallinen spoken in a carrier phrase under two different sentence stress conditions, n=9+-2. Subj. MLJ.

		V		C	
		x	s	x	s
Malinen	Stressed	76	5	56	5
	Unstressed	81	5	52	7
maalinen	Stressed	154	12	61	5
	Unstressed	157	13	60	9
mallinen	Stressed	90	5	115	9
	Unstressed	100	10	104	12
maallinen	Stressed	155	16	105	5
	Unstressed	144	10	93	11

Table A3-III. Mean durations and standard deviations (msec) of a- and l-segments in the Finnish words Malinen, maalinen, mallinen and maallinen spoken in a carrier phrase under two different sentence stress conditions, n=5. Subj. LP.

		V		C	
		x	s	x	s
Malinen	Stressed	88	7	48	2
	Unstressed	78	-	45	5
maalinen	Stressed	177	8	49	5
	Unstressed	143	7	51	3
mallinen	Stressed	109	8	123	14
	Unstressed	100	3	63	2
maallinen	Stressed	173	6	84	3
	Unstressed	127	20	64	4

Table A3-IV. Mean durations and standard deviations (msec) of a- and l-segments in the Finnish words Malinen, maalinen, mallinen and maallinen spoken in a carrier phrase under two different sentence stress conditions. For the stressed version n=6; for the unstressed version n=10+-1. Subj. SK.

		V		C	
		x	s	x	s
malinen	Stressed	55	5	55	6
	Unstressed	39	3	47	8
maalinen	Stressed	149	8	47	2
	Unstressed	118	16	47	8
mallinen	Stressed	71	3	118	8
	Unstressed	55	8	102	16
maallinen	Stressed	149	16	133	16
	Unstressed	110	16	71	8

Appendix 3: Finnish

Table A3-V. Vowel/consonant duration ratios between a- and l-segments in the Finnish test words. Four subjects.

Subject:		RJ	MLJ	LP	SK
		V/C	V/C	V/C	V/C
Malinen	(+Str)	1.61	1.36	1.83	1.00
	(-Str)	1.43	1.56	1.73	0.83
maalinen	(+Str)	3.91	2.52	3.61	3.17
	(-Str)	3.33	2.62	2.80	2.51
mallinen	(+Str)	0.78	0.78	0.89	0.60
	(-Str)	0.72	0.96	1.23	0.54
maallinen	(+Str)	2.16	1.48	2.06	1.12
	(-Str)	2.30	1.55	1.98	1.55

Table A3-VI. Change in V/C ratio as a function of stress in the Finnish test words. Four subjects.

Subject:	RJ	MLJ	LP	SK
Malinen	+0.18	-0.20	-0.10	+0.17
maalinen	+0.58	-0.10	+0.81	+0.66
mallinen	+0.06	-0.18	-0.34	+0.06
maallinen	-0.14	-0.07	+0.08	-0.43

Patterns of Coarticulation in English

Marie K. Huffman

I INTRODUCTION

Recent studies of coarticulation (e.g. Krakow and Manuel 1983; Magen 1984, 1985) have identified several ways in which the coarticulatory patterns of languages may differ. These include differences in the freedom of vowels to vary, in how far coarticulatory effects extend into a vowel, and (differences) in the relative strength of carryover versus anticipatory coarticulation (sometimes called "directionality"). Such results indicate that coarticulatory processes of individual languages are more complicated than earlier work has assumed. Given the prospect of substantial linguistic control over coarticulation, it is clear that the finer details of coarticulation warrant further investigation.

This work is part of a larger project concerned with the effect consonantal feature specifications have on patterns of vowel–to–vowel coarticulation. Another way to look at this question is to ask how independent consonant and vowel production are. Ohman (1966) proposed that vowel–to–vowel movement, as in a VCV sequence, is basically diphthongal, with the consonant gesture superimposed. However, in order to evaluate Ohman's proposal, we need to know much more about the factors that influence the shape of the vowel–to–vowel or vowel–consonant and consonant–vowel gestures. Ohman showed that when consonants have secondary articulations that involve vowel features (e.g. as in Russian), vowel–to–vowel coarticulation is blocked. If a plain consonant is simply overlayed on a vowel–to–vowel movement, how minor a perturbation to the vowel gesture is the consonant? Consonants without secondary articulations might still vary considerably in the effect they have on vowel–to–vowel movements.

In many studies of vowel–to–vowel coarticulation, /b/ is chosen as the intervening consonant to avoid any consonantal perturbation of the vowel gesture. However, several studies have included examination of different consonantal contexts. For English, Ohman (1966) varied place of articulation in stops, and touched on coarticulation across /s/, noting that /s/ appeared to show little if any coarticulatory effects. Other studies (e.g. Bell–Berti and Harris 1976, Gay 1977) have also considered coarticulation across different English stop consonants. In general, there is still some disagreement about the extent and magnitude of "typical" vowel–to–vowel coarticulatory effects across stops in English. There has been less work done comparing coarticulation across consonants with different manners of articulation. However, previous research on English /l/ (e.g. Bladon and Al–Bamerni 1976, Keating 1985) suggests that its coarticulatory properties might affect vowel–to–vowel coarticulation. Therefore, we chose to compare coarticulation across /l/ with coarticulation across /d/.

In addition to the question of the consonant's perturbation of the vowel–to–vowel gesture, another basic issue in understanding vowel–to–vowel coarticulation is the role played by stress. Since stress affects vowel quality and duration, it presumably influences vowel–to–vowel movements. Fowler's (1981a,b) theory of coarticulation extends Ohman's notion of a diphthongal gesture with a slightly more abstract model intended to account for coexistent and complementary patterns of acoustic shortening and coarticulatory effects. She proposes that stressed vowels affect neighboring segments due to articulatory overlap. Fowler notes that stressed vowels are shortened when they are followed by one or more segments. She relates this acoustic shortening to an overlap in production with these segments. This overlap results in coarticulatory effects of a stressed vowel on neighboring consonants and stressless vowels. Thus Fowler assumes that stressed vowels have precedence at some level of speech planning. In other words, there is an asymmetry in effects. The more overlap there is, the more stressed vowels affect unstressed vowels, but not vice versa. However, the theory tells us nothing about what to expect in terms of coarticulation between weakly stressed vowels and unstressed vowels. Therefore, we decided to examine vowel–to–vowel coarticulation while varying stress environment.

A third major consideration in explaining patterns of vowel–to–vowel coarticulation is directionality of coarticulatory effects. In acoustic studies similar to the present investigation (e.g. Ohman 1966, Bell–Berti and Harris 1976, Magen 1984), carryover effects were found to be stronger than anticipatory effects in English. However, a number of researchers (Gay 1974, 1977, Parush et al. 1983, Henke 1966) suggest that carryover and anticipatory coarticulation are quite different phenomena, and probably cannot be compared, per se. Gay claims, for example, that carryover effects are mechanical inertial effects, which do not extend beyond the immediate phonetic context. On the other hand, he says, anticipatory effects are timing effects; different aspects of a segment's articulation begin at different times, under linguistic control at some level. This picture is muddled by the fact that anticipatory effects can presumably override the inertial carryover influences that would otherwise occur. In cases where no anticipatory effects are present, these "same" carryover mechanical effects may appear to last longer. We will look for differences in carryover and anticipatory coarticulation, and possible conditioning factors for these differences, in the data reported below.

II METHODS

Two male Californian speakers of American English were recorded saying items of two types. One type of token had the form VCV, where C = {d,l} and V = {i,a,u,ə}. Stress was placed on either the first or second vowel; schwas were only used in unstressed positions. The other type of token had the form bVCəCVb, where both C's were the same, either /d/ or /l/, and V = {i,a,u}. As pronounced by the subjects, there was one primary and one secondary stress per token, falling of either the first or third vowel. Schwa was always unstressed.

Both kinds of items were said in a carrier phrase which was constructed to be as symmetrical as possible immediately adjacent to the experimental items. The complete block was repeated twice, embedded between two sentences intended to aid token identification. The complete block of sentences for each word type was as below; main stresses were located as marked:

Say ´_____ for the fírst time.
Say ´_____ yet agáin.
Say ´_____ yet agáin.
Say ´_____ for the lást time.

The VCV items were pseudo–randomized and recorded as a set, followed by the complete set of b...b items, also pseudo–randomized.

The tokens were digitized and formant frequencies were determined using LPC analysis. Formants one and two were measured at the schwa midpoint in the b...b items, and at the VC and CV boundaries in both kinds of items. Wideband spectrograms were made of all tokens. These were used to measure segment durations, and to aid in segmentation of the output of LPC analysis. Analyses of variance were used to evaluate the contribution of stress, vowel identity and consonant identity to variance in vowel formant frequencies.

III RESULTS–VCV's

We can look at patterns of coarticulation with plots like Figure 1, which shows mean onset formant values for V2 following different V1's speaker two, where the consonant is /d/. Each data point represents the average formant onset values of /i, a, u/ following a particular V1. The vowel quality of V1 is as labelled. Since schwas were always unstressed, they were not included in these plots, so that stressed and unstressed conditions could be compared. Dots indicate average formant values for all 3 V2's when main stress falls on V1, and stars denote average formant values when main stress falls on V2 (i.e. V1 is only weakly stressed). The more spread there is in average formant values in this kind of plot, the more coarticulatory effect V1 is assumed to have on V2. Thus, we are looking at a quantity similar to the "coarticulatory range" used by Fowler (1981).

In Figure 1, there is spread between data points in the F1 and F2 dimensions in both the context of stressed V1 (●'s) and the context of weakly stressed V1 (★'s). This indicates that identity of V1 does influence V2 onset values, in both stress contexts. Plots of this kind give us a preliminary idea about patterns of coarticulation, which we can then test for statistical significance within the complete data set. Figures 1–4 illustrate carryover and anticipatory coarticulation across /d/ for both speakers and both stress contexts. The first thing to note is that mean V2 offset values (Figures, 1 and 2) and mean V1 offset values (Figures 3 and 4) show spread in the F1 and F2 dimensions, reflecting the identity of the transconsonantal vowel. Thus, we have both carryover and anticipatory vowel–to–vowel effects. Analyses of variance indicated that identity of V1 had a highly significant effect ($p<.0001$) on both F1 and F2 of V2. Similarly, V2 has a

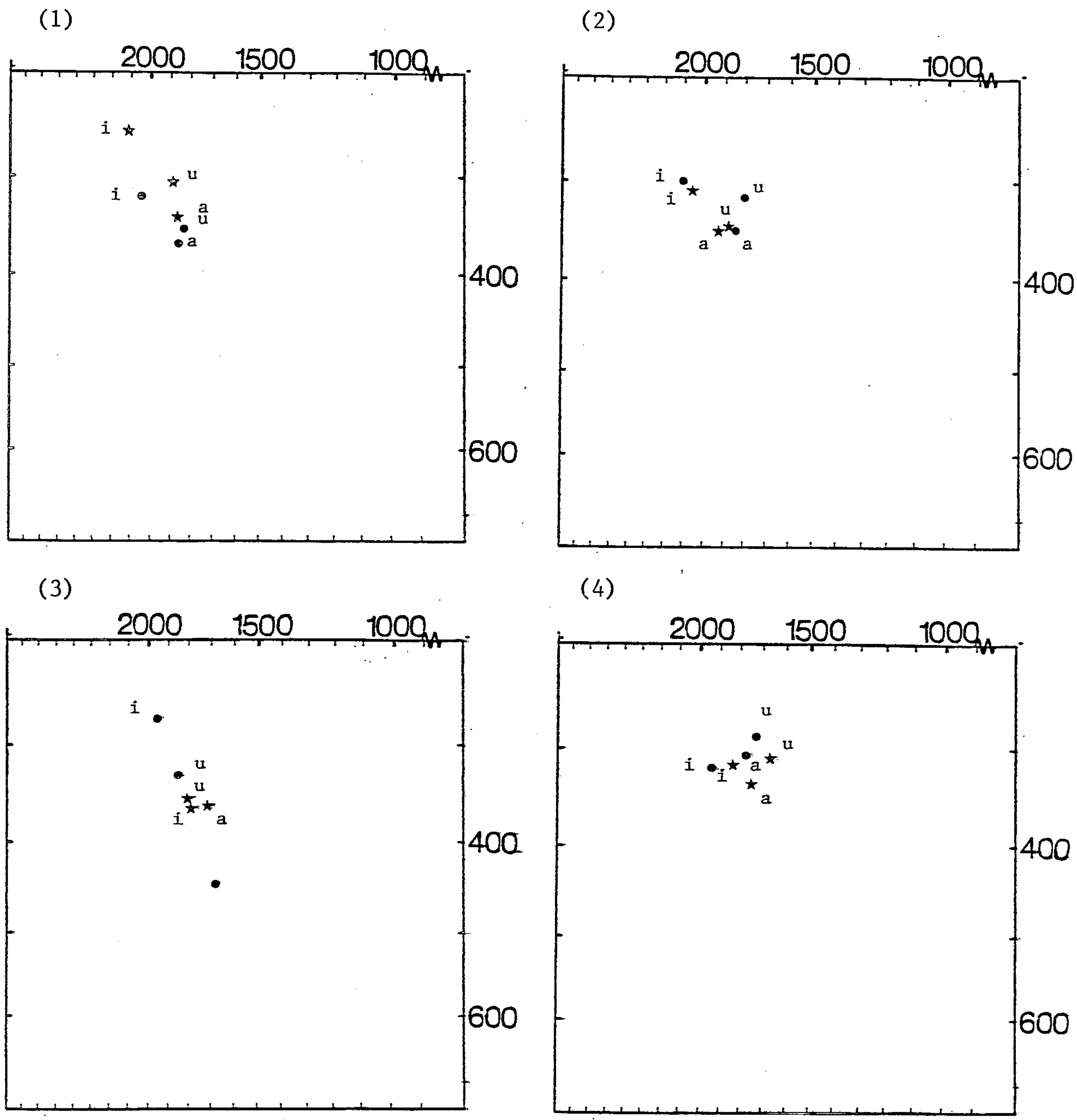


Figure 1–4. Plots of coarticulatory effects in VdV items. Figures 1 and 2 illustrate carryover and anticipatory effects, respectively, for speaker one. Figures 3 and 4 are the comparable plots for speaker two. For legend, see Figures 5 and 6.

highly significant effect ($p<.0001$) on V1 formant frequencies, except for F1 of speaker two, where the effect is less robust ($p<.011$).

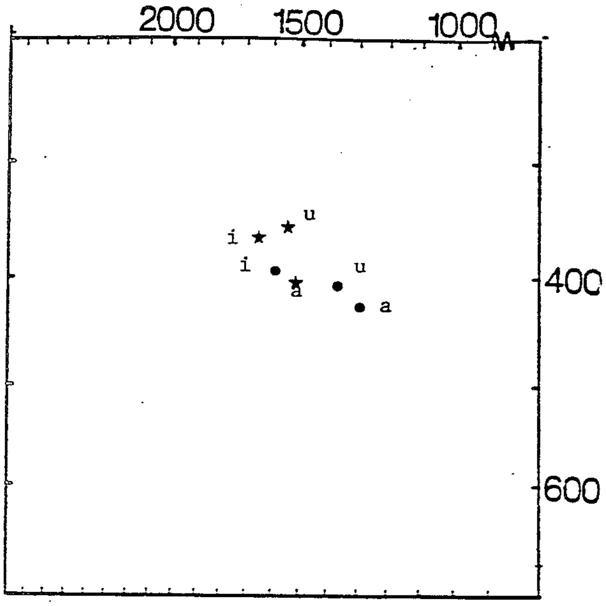
Figures 5 and 6 illustrate vowel–to–vowel effects when the intervening consonant is /l/. As with the /d/ tokens, the spread in the F1 and F2 dimensions seen in these figures indicates that there is also vowel–to–vowel coarticulation across /l/. Analyses of variance indicated that for both speakers, V1 identity had a very significant effect ($p<.001$) on the onset value of F2 for V2, and V2 had an equally significant effect on the offset value of F2 for V1. F1 values generally did not vary consistently with changes in the transconsonantal vowel. The only significant effect for F1 occurred in anticipatory coarticulation for speaker two, where F1 of V1 varied significantly with changes in V2.

Is amount of vowel–to–vowel coarticulation affected by stress differences? If stressed vowels have the dominance in coarticulatory effects which Fowler suggests, we should see considerable differences in amount of coarticulation in the stressed and weakly stressed conditions. Specifically, we expect a stressed vowel to have a stronger coarticulatory influence on a transconsonantal vowel than a weakly stressed vowel does. This will be reflected in the amount of spread between data points for comparable stressed and weakly stressed conditions. However, comparing the amount of spread in the two stress contexts (dots versus stars) for each of Figures 1–4 reveals that in the /d/ tokens, stress of a vowel generally has little effect on the amount of influence it will have on a transconsonantal vowel. There is only one case where stress context makes a big difference in amount of spread, namely anticipatory coarticulation for speaker one (Figure 3), where there is much more F1 spread in V1 offset formant values when V2 is stressed.

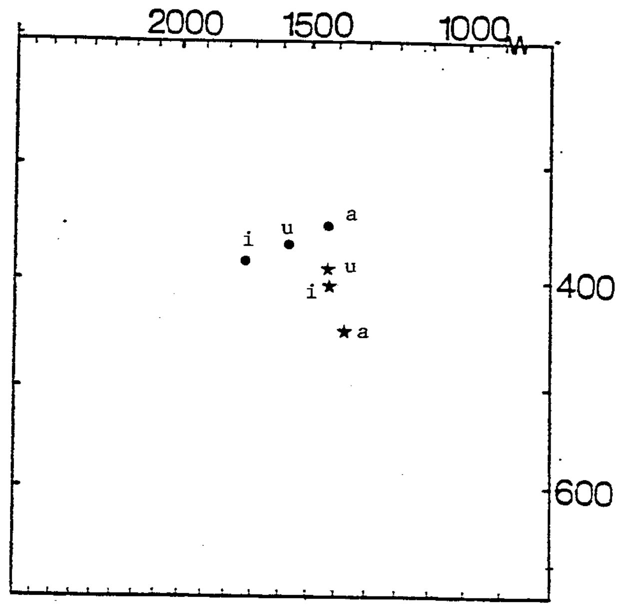
Comparing stress contexts in each of Figures 5 and 6, we see that in the /l/ tokens as well, stress does not have a strong effect on amount of spread in F1 and F2. However, when there is a difference between stress contexts, the data patterns as Fowler would have predicted. That is, stressed vowels have the stronger effect on transconsonantal vowels. For speaker one (Figure 5), a stress effect is seen in carryover coarticulation. There is slightly more spread in F2 values of V2 when V1 has primary stress, than when it is weakly stressed. For speaker two (Figure 6), a somewhat stronger effect is seen in anticipatory coarticulation. There is less F2 variation (less spread) in V1's that precede a weakly stressed V2 as opposed to a stressed V2.

While stress context generally only weakly influences the strength of coarticulatory effects, it does appear to affect the overall vowel–to–vowel movement as measured at V–C and C–V boundaries. This effect can be seen by comparing the location in the vowel space of the data points, rather than by looking for differences in spread between points. For the /d/ tokens, stress context had a significant effect (to at least the level $p<.01$) on mean formant values as indicated by asterisks (*) in Table 1. For speaker one, V2 onsets were higher and fronter when main stress was on V2. For speaker two, V1 offset and V2 onset values were generally fronter when stress fell on V2.

(5)



(6)



Figures 5–6. Plots of coarticulatory effects in VIV items. Figure 5 shows carryover effects for speaker one; Figure 6 shows anticipatory effects for speaker two.

V● = average of vowels across the consonant from stressed V.
V★ = average of vowels across the consonant from weakly stressed V.

Table 1. Significant stress effects for /d/ tokens.

Speaker		V1 offset	V2 onset
1	F1	---	*
	F2	---	*
2	F1	---	---
	F2	*	*

Overall effects of stress on formant frequencies of the /l/ tokens are almost identical to those reported in Table 1 for the /d/ tokens. Significant effects for the /l/ tokens are as indicated in Table 2. For speaker one, the /l/ tokens are like the /d/ tokens. V2 onsets are generally higher and fronter after weakly stressed V1. For speaker two, V1 offsets are fronter and higher, and V2 onsets are fronter, when V2 is stressed.

Table 2. Significant stress effects for /l/ tokens.

Speaker		V1 offset	V2 onset
1	F1	---	*
	F2	---	*
2	F1	*	---
	F2	*	*

While examining consonant and stress differences, we have treated carryover and anticipatory coarticulation separately. If we hold consonant context and stress of the affecting vowel constant, are there differences in amount of carryover and anticipatory coarticulation? In general, differences are more visible in F2 than F1, but even F2 differences are often slight. For the /d/ tokens, for both speakers, when there is a difference in F2 values, there is more carryover coarticulation than anticipatory coarticulation. For speaker one, there is a notable difference in F1 effects as well, but it patterns the other way. As can be seen by comparing Figures 1 and 3, for this speaker, when the affecting vowel is stressed (marked by dots), there is more spread in F1 in the anticipatory condition than in the carryover condition. Comparing Figures 2 and 4 we see that for speaker two, there are no large differences in amount of carryover and anticipatory coarticulation in F1 across /d/.

In the /l/ tokens, there are substantial differences in amount of carryover and anticipatory coarticulation in F2 when the affecting vowel is only weakly stressed. Results for our two speakers pattern differently under these conditions. For speaker one, there is more anticipatory than carryover coarticulation from a weakly stressed vowel onto a stressed vowel. For speaker two, the situation is just the reverse. Finally, for the /l/ tokens, as in the /d/ tokens, differences in amount of coarticulatory influence in F1 are slight.

In summary, for the VCV items, we found both carryover and anticipatory vowel–to–vowel coarticulatory effects. Stress did not have much of an influence on amount of coarticulation in either direction, though stress tended to result in overall shifts of mean formant values in the vowel space. Finally, there were no consistent patterns of directionality in coarticulatory effects.

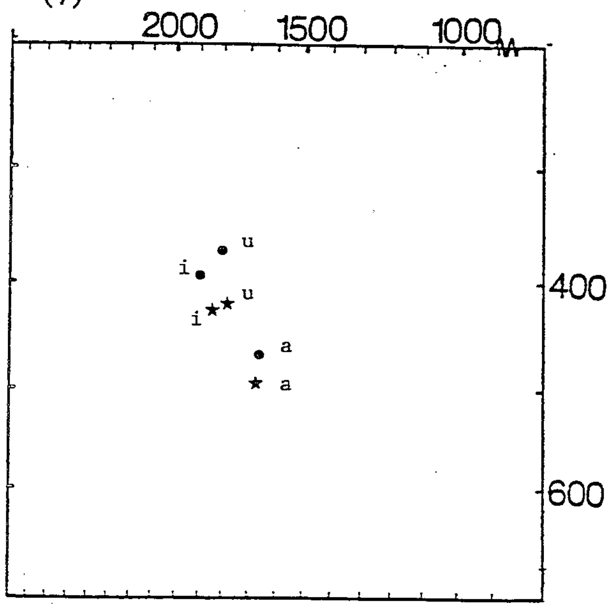
IV RESULTS–b...b's

First we will consider coarticulatory effects onto schwa. In these tokens, V2 is always schwa, so we will simply refer to it as "schwa"; the vowels flanking schwa will be designated "V1" and "V3". Since formant frequencies were measured at the schwa midpoint, we are looking at a point further away from the affecting vowel than we were in the VCV tokens. Nevertheless, we do see vowel–to–vowel effects. Figures 7–10 illustrate average schwas in different vowel and stress environments for the two consonantal contexts. In both the /d/ tokens and the /l/ tokens, the schwa midpoint reflects the identity of V1 and V3 in one or both formant dimensions. The influence of V1 and V3 identity on schwa formant values was very significant ($p<.001$), for both consonants and both speakers. The interesting exception is anticipatory coarticulation across /l/ for speaker one. In the F2 dimension, there is no significant influence of V3 on a preceding schwa in either stress context. Figure 10 illustrates this levelling of F2 distinctions in schwas preceding stressed vowels. Notice that while F2 distinctions are lost in this context, F1 differences remain.

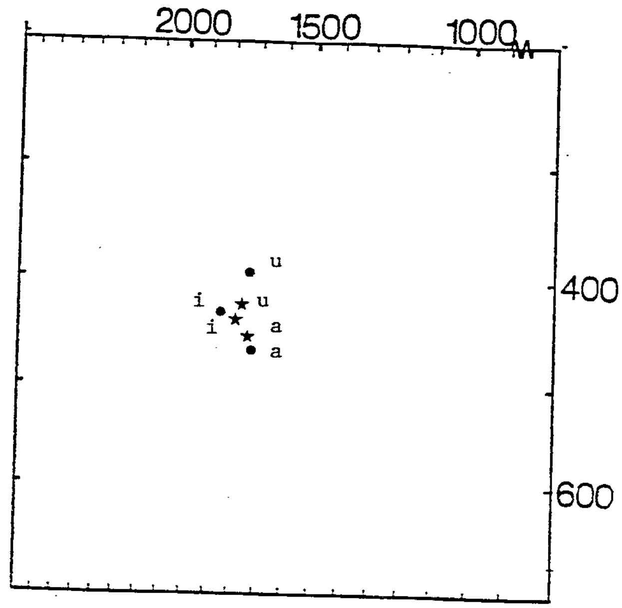
For the b...b tokens, we looked for long–range coarticulatory effects by measuring formants at the V1–C and C–V3 boundaries. Significant coarticulatory effects were found in only one case: carryover coarticulation across /l/, for speaker one. Figure 11 shows average V3 onset formant values after V1 = {i,a,u}, where C = /l/. Results are pooled across stress context. Analysis of variance found that for speaker one, F2 of V3 varied highly significantly ($p<.0001$) with changes in V1. In particular, front and back V1's affected V3 differently. There was no significant difference between /u/ and /a/ contexts as Figure 11 suggests. There were no significant differences in F1 of V3 in different V1 contexts, and stress context had no effect on V3 formant values.

How does the stress on a vowel affect its coarticulatory influence over a transconsonantal schwa? One reason that the b...b items were included in this study was to allow us to control more completely for the effect on coarticulatory patterns of changes in stress of the affecting vowel; the affected vowel does not change in stress: it is always unstressed schwa. Taking the /d/ tokens first, in carryover coarticulation (Figure 7) we see that for speaker one there is very little difference between the two stress contexts. When V1 is weakly stressed there is slightly less spread in the F1 and F2 directions than when V1 is stressed. However, there was no significant interaction between stress and V1 in F1 or F2 values in carryover coarticulation for either speaker. The only significant stress effect was that F1 of speaker one's schwas was higher overall following weakly stressed vowels than following stressed vowels. In Figure 8, the anticipatory case, we also see little

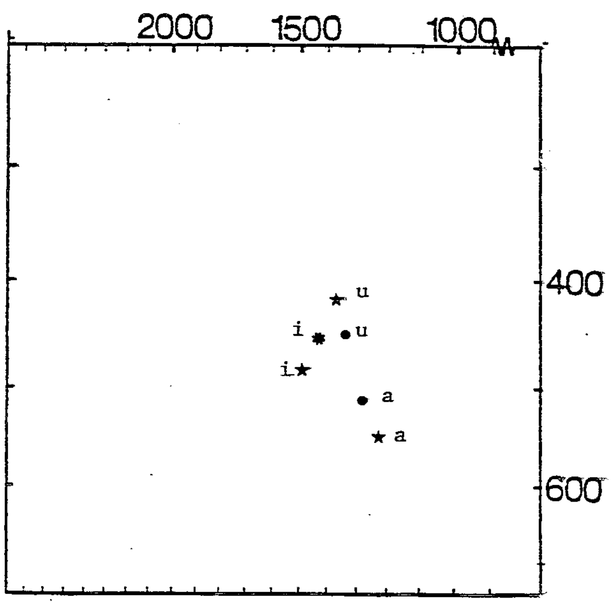
(7)



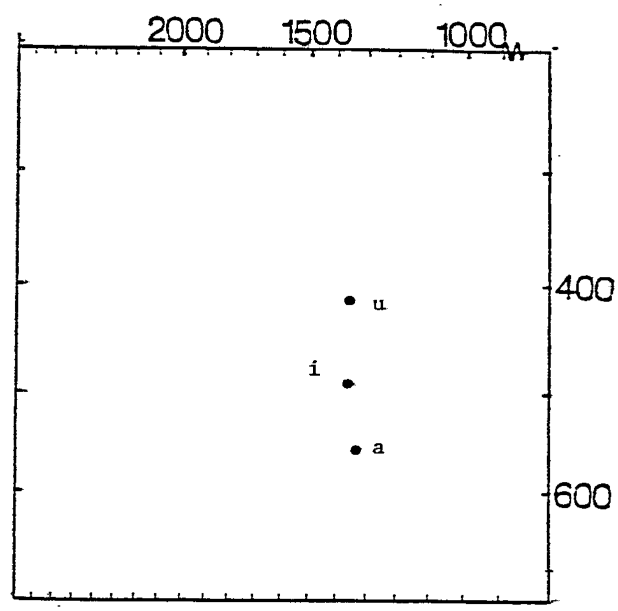
(8)



(9)



(10)



Figures 7–10. Plots of coarticulatory effects in b...b items. Figures 7 and 8 illustrate carryover (speaker one) and anticipatory (speaker two) coarticulatory effects onto schwa across /d/. Figures 9 and 10 illustrate carryover and anticipatory effects onto schwa across /l/ for speaker one. Asterisk indicates that a few data points are missing.

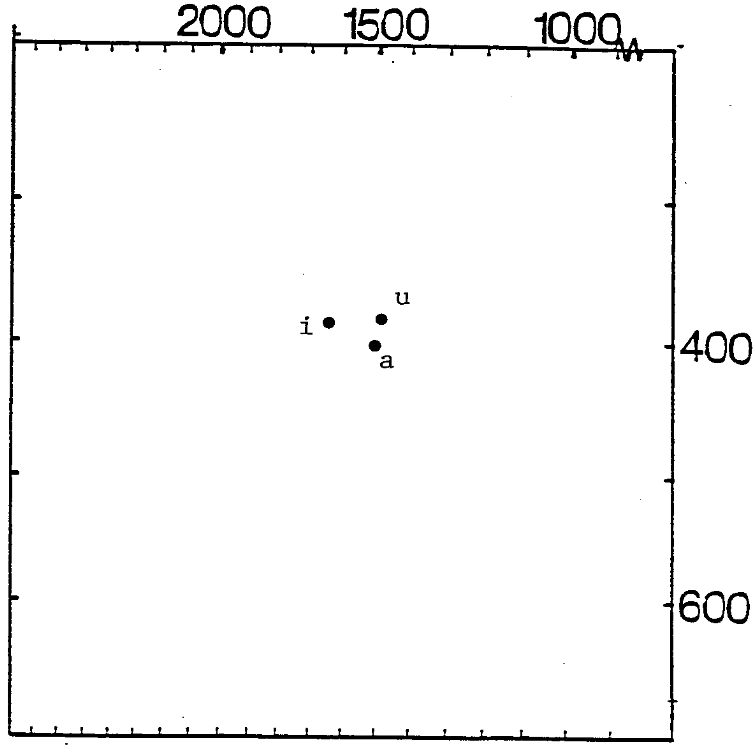


Figure 11. Long–range coarticulatory effect of V1 on V3 in the context of /l/ for speaker one. Values are averaged across stress contexts.

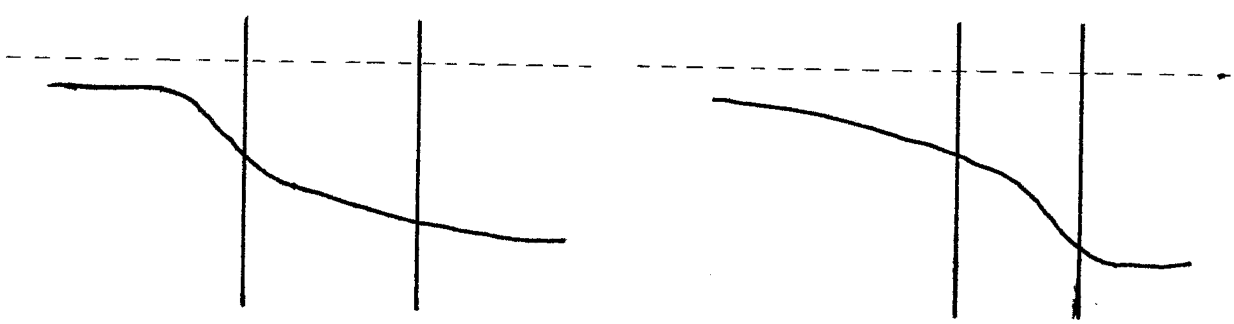


Figure 12. Sketch of how consonant duration differences can mask differences in vowel–to–vowel movements, represented here by F2. Vertical lines represent consonant closure and release, repectively. The dashed line is an arbitrary reference line.

difference in amount of spread in the stress conditions. For anticipatory coarticulation across /d/, there was no significant stress effect in either formant dimension, for either speaker.

Stress seems to affect coarticulation more in the /l/ tokens. As can be seen in Figure 9, in carryover coarticulation for speaker one, schwas after weakly stressed vowels are slightly more peripheral in the F2 dimension. That is, weakly stressed V1's appear to have a stronger carryover effect on F2 of a following schwa than stressed V1's have. This is reflected in the significant ($p<.01$) interaction between V1 and stress for these tokens. For speaker two, there is no real difference in amount of spread in the two stress contexts, but there is a general lowering of F2 after weakly stressed vowels. Anticipatory coarticulation across /l/ shows minimal effects of stress context. Neither speaker shows stress effects on the amount of spread in F1 or F2 values. For speaker one, there is also no stress effect on mean formant values in either the F1 or the F2 dimensions. For speaker two, the story is the complement of the carryover case. That is, schwas before stressed V3 (thus following a weakly stressed V1) show a uniformly lower F2.

V DISCUSSION

The results reported above show a considerable amount of variation between speakers and between the two kinds of experimental items (VCV's and b...b's). However, there are several general patterns which should be noted. First, we observed both carryover and anticipatory coarticulation across /d/ and /l/, in both kinds of tokens, for both speakers. For one speaker, we found significant long–range effects of V1 on V3 in the b.l.b items. Second, stress effects on amount of coarticulation are minimal, though they are more common in /l/ tokens. Again, this is true for both kinds of tokens, for both speakers. In addition, overall stress influences on mean formant values are fairly consistent in the VCV tokens; they are less consistent in the b..b items. For the VCV tokens, formant values are generally fronter, and sometimes, higher, in the vowel space in the immediate environment of a stressed vowel. There are some speaker differences here. Finally, patterns in coarticulatory "directionality" are more detectable for the b...b items, especially in the /l/ consonantal context. For F2 in these items, carryover coarticulation is always stronger than anticipatory coarticulation, for both speakers, though sometimes the differences are slight. For F1, there is a stress effect on relative strength of coarticulatory influences in the /l/ items, particularly for speaker one. When the affecting vowel is stressed, anticipatory coarticulation is stronger than carryover coarticulation; when the affecting vowel is weakly stressed, carryover coarticulation is stronger than anticipatory coarticulation.

What can these trends tell us about the issues in coarticulation raised in the introduction? First we will consider the question of the consonant as a perturbation to the vowel gesture, in terms of the strength and extent of consonantal effects. Although physiological data is needed to treat this issue conclusively, here we will infer the vowel gesture from formant movement. One way to identify consonant effects on vowel formants is to look for differences in the two consonantal contexts. When we see differences between /d/ and /l/ tokens, we are presumably seeing consonantal effects, since the items are the same in all other senses. One gross difference between /d/ and /l/ items is the part of the vowel space occupied by neighboring vowels. For both VCV and b...b items, vowels in the /l/ context are lower and backer in the vowel space than they are in the /d/ context. This can be seen for the VCV's by comparing Figures 1–4 with Figures 5 and 6. For the b...b's, compare Figures 7 and 8 with 9 and 10.

The extent (in time) of consonantal influences on vowel formants can be investigated by comparing results of measurements made at different distances from the consonantal boundary. In our data, we can do this by looking for differences between our two kinds of tokens: the VCV's, where vowel formants were measured at the consonantal boundary, and the b...b's, where vowel formants were measured at some distance into the schwa. When interpreting the results of such a comparison, we must keep in mind that there is also a difference in vowels between the two kinds of items. That is, in the b...b tokens, the affected vowel is always schwa, which might be expected to show coarticulatory effects more readily than other vowels. It was already noted above that the consonant's effect on general placement of neighboring vowels or vowel transition edges in the vowel space extends as far as the schwa midpoint.

One way in which the VCV and b...b items differ is in overall stress influences on average formant values. In the b...b items, such effects are weak, and show no real pattern. In the VCV items, there are slight differences between speakers, but each speaker is fairly consistent. For speaker one, V2 onset is fronter and higher when V2 is stressed, for both /d/ and /l/ tokens. For speaker two, V–C and C–V boundary values are always fronter when V2 is stressed, for both /d/ and /l/ tokens. For this same speaker, in /l/ items, V1 offset (i.e. at the V–C boundary) is higher in this context as well. The fact that this general stress effect is seen clearly in the VCV items but not in the b...b items suggests that it is only a local effect of the consonant on neighboring vowels. More precisely, it is a stress effect on consonant quality, which affects vowel quality at consonantal boundaries.

We have neither the space nor the articulatory data which would allow us a thorough discussion of stress effects on consonants. However, we might note that such an effect is not completely surprising. In the /d/ tokens, the effect is at least partly due to the stress conditioned presence or absence of flapping. When V2 is stressed, the preceding /d/ is always a stop, while when V1 is stressed, the following /d/ is sometimes a flap. It is possible that the tongue body is more forward in the mouth for [d] than for [ɾ], perhaps to aid complete occlusion for the stop. This could produce the higher F2 observed for tokens with stress on V2. As for /l/, we already know there is allophonic variation between clear [l] and velarized, backer [ɫ] in English, but we usually do not think of the rule as stress dependent. However, velarization can be determined by position of the /l/ in the syllable, with [l] occuring syllable initially, and [ɫ] occurring elsewhere. Perhaps stress on V2 supports syllabification of the nonsense word as V.lV, while initial stress may favor interpretation of the /l/ as ambisyllabic, thus not strictly syllable initial, and therefore dark (backed). This would result in (fronter) clear l's with stressed V2 and (backer) velarized l with stressed V1, which is basically the pattern observed.

A second way in which results differ for the VCV and b...b items is in the consistency of patterns in directionality. The b...b tokens show several consistent patterns in relative amount of carryover and anticipatory coarticulation, while the VCV items show almost none. That is, at consonant boundaries (as in VCV items) consonantal influence is strong enough to overshadow differences in coarticulatory effects of neighboring vowels. However, the picture is much clearer when vowel properties are measured at a short distance from the consonant (as in the b...b items). Interestingly, at this point (i.e. the schwa midpoint), there are still some differences between consonantal contexts. In particular, the /l/ tokens show a straightforward effect of stress environment on amount of coarticulation (as mentioned earlier), while the /d/ tokens do not. In other words, there is still some influence of consonantal environment. In summary, we have seen both a very local consonantal effect, and slightly more extensive (in time) effects on vowel formant values. We will discuss consonant effects further below.

What can we say about how stress affects the vowel–to–vowel gesture? As was mentioned earlier, Fowler (1981a) suggests that stress will affect the nature of the vowel–to–vowel movement; specifically, stressed vowels will dominate unstressed ones. Magen's (1984) study presents data which seem to support Fowler's claims. Magen found stress effects on the extent (in time) of vowel–to–vowel coarticulation in English VCV's. In particular, she found that anticipatory effects appear later in a stressed vowel than in an unstressed vowel. Likewise, carryover effects decrease sooner in stressed vowels than in unstressed vowels. On the other hand, as was noted earlier, we found little consistent effect of stress on amount of vowel–to–vowel coarticulation, whether it was measured near consonant boundaries or at the schwa midpoint.

Is there some other effect in our data which can explain this lack of a stress influence? One possibility is that consonant duration differences are affecting how vowel–to vowel gestures are realized acoustically. If consonants are superimposed on a vowel–to–vowel gesture, then consonant duration can presumably vary independently of that gesture. This also means that consonant duration could affect what part of the vowel–to–vowel gesture is realized acoustically. This in turn would influence measured coarticulation. In our case, to get similar vowel–to–vowel gestures in different stress environments, we need a scenario like that sketched in Figure 12. 12a and b are possible schematics of vowel–to–vowel movements (represented by F2) in a VCV sequence when main stress falls on V1 and V2, respectively. With a longer C after stressed V1 and a shorter consonant after weakly stressed V1, we get similar F2 values at the consonant boundaries. However, in actual fact, consonant durations pattern in exactly the opposite way.

Table 3 gives consonant durations for the VCV tokens in the two stress contexts, for each speaker. Both /d/ and /l/ are shorter after a stressed V1. In this stress context, /d/'s are flapped about one third of the time, for both speakers. The durations in Table 3, taken with the coarticulatory patterns shown in Figures

1–6, indicate that consonant duration does not have a simple effect on observed coarticulation. For one thing, there are instances where considerable consonant durations exist without differences in amount of coarticulation. Examples include carryover and anticipatory coarticulation across /d/ for speaker two (Figures 2 and 4), where /d/ durations are different in the two stress environments ($p<.0075$), but amount of coarticulation is very similar. In addition, it is not true that a longer consonant results in a weaker vowel–to–vowel coarticulatory effect; there are cases where we see stronger coarticulatory effects across a longer consonant. This is true for speaker one (Figure 3), who shows more anticipatory coarticulation across /d/ when V2 is stressed.

Table 3. VCV consonant durations in msec.

Speaker		V1	V2
1	/d/	46.4	82.9
	/l/	58.4	91.1
2	/d/	47.3	96.9
	/l/	70.2	134.5

Table 4 shows duration patterns for the b...b items. Notice that stress–conditioned consonant duration differences are most notable in the consonant following schwa. For both speakers, the first /d/ was usually flapped. For speaker one, the first /d/ in the b...b tokens was flapped in all but two tokens, regardless of the location of primary stress. For speaker two, /d/ after stressed V1 was always flapped, but /d/ after weakly stressed V1 was flapped only half of the time, with no real pattern except that five out of six /d/'s after /i/ were not flapped. The second /d/ in the b...b tokens was never flapped, for either speaker. Nonetheless, despite stress related duration differences in the consonant following schwa, amount of coarticulation across it did not vary appreciably in different stress environments.

Table 4. b...b segment durations in msec.

Speaker		´V1	´V2
1		´b V d ə `d V b	`b V d ə ´d V b
	C Dur.	31.3 76.9	34.0 86.7
	V Dur.	91 67 147	72 71.3 195
		´b V l ə `l V b	`b V l ə ´l V b
	C Dur.	48.0 77.0	52.0 84.0
	V Dur.	97 66 139	77 63 148
2		´b V d ə `d V b	`b V d ə ´d V b
	C Dur.	26.7 79.6	32.9 108.8
	V Dur.	121 94 194.4	94.5 96.5 257.8
		´b V l ə `l V b	`b V l ə ´l V b
	C Dur.	52.4 84.9	57.6 126.1
	V Dur.	97.6 77.8 182.2	81.1 78.7 236

What we can conclude from all of this is that if stress influences the nature of the vowel–to–vowel gesture (as indicated by Magen's results), the effect is not great enough to be detectable amid the noise introduced into the data by differences in consonant duration and quality (see below) which also occur with changes in stress.

We have not yet discussed what our data have to say about directionality in coarticulation. First we will summarize the directionality patterns found in the data. These are based on comparisons between carryover and anticipatory effects where the affecting vowels carry the same stress (either main or weak stress). It was mentioned earlier that the VCV items showed much less consistency in directionality differences than b...b items. The one pattern which was clear in the VCV items, and which held for both speakers, was that for coarticulation in F2 across /d/, carryover coarticulation was stronger than anticipatory coarticulation. For the b...b items, coarticulation in F2 showed the same behavior: carryover coarticulation was stronger than anticipatory coarticulation, for both speakers, in both consonantal contexts. For coarticulation in F1, there is a stress effect on relative amounts of carryover and anticipatory coarticulation. Finally, it should be noted that the long–range coarticulatory influence of V1 on V3 in speaker one's b.l.b items, was a carryover effect, with no comparable anticipatory effect.

So in these acoustic data, we see a general dominance of carryover effects. However, consonantal context does influence the relative strength of coarticulatory effects in the two directions. We need an explanation for the consonant differences, but as was established earlier, consonant duration does not do any work for us in explaining coarticulatory patterns. There are at least two other ways that consonants could affect vowel–to–vowel gestures. One way is by having a specification for a vowel feature. As was discussed earlier, velarization has been found to affect coarticulation across /l/ (Keating 1985). Keating suggests that a velarized /l/ carries specification for the vowel feature [back], and that its [+back] feature can enhance or block vowel–to–vowel coarticulation in different contexts.

While we did not observe clear cases of blocking, we did observe instances of /l/ affecting vowel formant values in a way related to backness, for speaker two. In carryover coarticulation, while identity of V1 was reflected in schwa values, there was a general lowering of schwa's F2 (i.e. backness) after weakly stressed V1, compared to stressed V1. Inspection of wide–band spectrograms reveals that this lowering is present during the /l/. In addition, the amount of formant movement between /l/ and a following schwa changes very little between stress contexts. This suggests that to a large extent, the value of F2 for schwa is determined by a combination of V1's influence on the intervening /l/, and a general lowering of F2 of /l/ in this environment.

If F2 of /l/ is taken as an indication of velarization (e.g. Bladon and Al–Bamerni, 1976), then we might say that /l/ following unstressed vowels is more velarized than /l/ following stressed vowels, for speaker two. One problem with this analysis is that for this speaker, F2 of /l/ still strongly reflects the identity of V1, whereas Bladon and Al–Bamerni (for British English) and others (e.g. Keating 1985, for American English) have found velarized /l/ to be resistant to coarticulatory influences. Thus, for speaker two, we seem to be able to get a lowered F2 of /l/ by a phonetic detail rule which applies in the F2 (front–back) dimension, as in the b...b items for this speaker. The rule partially determines the quality of non–velarized /l/'s, specifying that they are backed somewhat, but without the coarticulatory resistance usually associated with velarized allophones.

A second way a consonant can influence vowel–to–vowel gestures is by having some low–level articulatory requirement which constrains a vowel–to–vowel gesture without causing substantive changes in the "shape" or overall time course of the gesture. Put another way, what appears to be at issue is the freedom of the tongue body to vary its position during production of a consonant. Recasens (1984) found for Catalan that the more tongue–dorsum contact there was during a consonant, the less vowel–to–vowel coarticulation was observed in physiological and acoustic measurements. In English, the lateral is somewhat less constrained in tongue body position than the alveolar stop. The change in amount of coarticulation Recasens observed was continuous, unlike the discrete on–off nature of coarticulation as affected by the binary feature [back] in Keating's results. While we did find an example of backing without total blocking (the low level backing rule just described), we also observed examples of even freer, more "continuous" variability in backness of /l/; namely, in the b.l.b tokens of speaker one.

For speaker one, the relative freedom in production of /l/ combines with stress–conditioned V1 duration differences to produce the carryover coarticulatory patterns seen in Figure 9. As can be seen in the spectrograms in Figure 13, during stressed V1, F2 moves considerably before closure for the /l/. On the other hand, during the shorter, weakly stressed V1, F2 does not move as far before closure for the /l/, and there is still vowel–to–vowel transition visible during the /l/. The duration of /l/ is basically the same in the two conditions, so the end result is that by the midpoint of the schwa, the influence of a stressed V1 is weaker than that of a weakly stressed V1.[1] In addition, F2 of /l/ ends up varying considerably with changes in V1 identity. Compare this situation to the spectrograms in Figure 14. When /d/ is the intervening consonant, we do not see these differences in formant movement, even in this vowel context where a transition into /d/ is likely to be visible.

This same weak influence of stressed V1 on following schwa helps explain the stress–related differences in directionality observed for speaker one. We could say that, as for F2, carryover coarticulation is stronger than anticipatory coarticulation, at least when the affecting vowel is weakly stressed. However, when the affecting vowel is stressed, carryover coarticulation is weak enough that the influence of a stressed V3 on schwa is considerably stronger. The variability in front–back tongue body position during /l/ for speaker one is crucial to an explanation of his long–range carryover effects. As can be seen by comparing the spectrograms in Figure 15, the influence of V1 is evident in F2 of both /l/'s and the intervening schwa. For ˈbiləˌlub, F2 of schwa is 1428 Hz; for ˈbaləˌlub, F2 of schwa is 1314 Hz. It is clear that no segment between V1 and V3 is blocking carryover coarticulatory effects. Rather, the /l/'s help propagate the effect of V1.

This result is interesting, because it suggests that Gay's and others' characterization of carryover coarticulatory effects as inertial is appropriate. In this case, the inertial forces are quite strong; so strong, it seems, that, given no conflicting articulatory requirements of following segments, carryover effects can last a long time. Note that these results go directly counter to the predictions made by Henke's (1966) look–ahead model of speech production. A better understanding of carryover effects like those observed in these /l/ tokens may clarify whether or not the /l/'s are to be considered totally unspecified for vowel features. If they are, then we would have counterevidence to Henke's proposal, since in his model, unspecified segments should receive the next non–contradictory specifed value for a feature. However, evidence may be found demonstrating how these /l/'s are in fact assigned values for vowel features via some interpretation of the qualities of the preceding vowel. In this case, lack of look–ahead effects would be consonant with Henke's theory.

The above interpretation of long–range carryover effects has consequences for the explanation of the collapse of F2 differences in Figure 10 (anticipatory coarticulation for speaker one). Figure 10 would appear to be a case of coarticulatory blocking by /l/. However, we have just seen that the second /l/ in the b.l.b's is not blocking long–range carryover coarticulation effects. If we argue that lack of carryover blocking is due to a lack of specification for the relevant vowel feature (i.e. backness), then this /l/ cannot be blocking anticipatory coarticulatory effects by being velarized. Consequently, lack of anticipatory effects

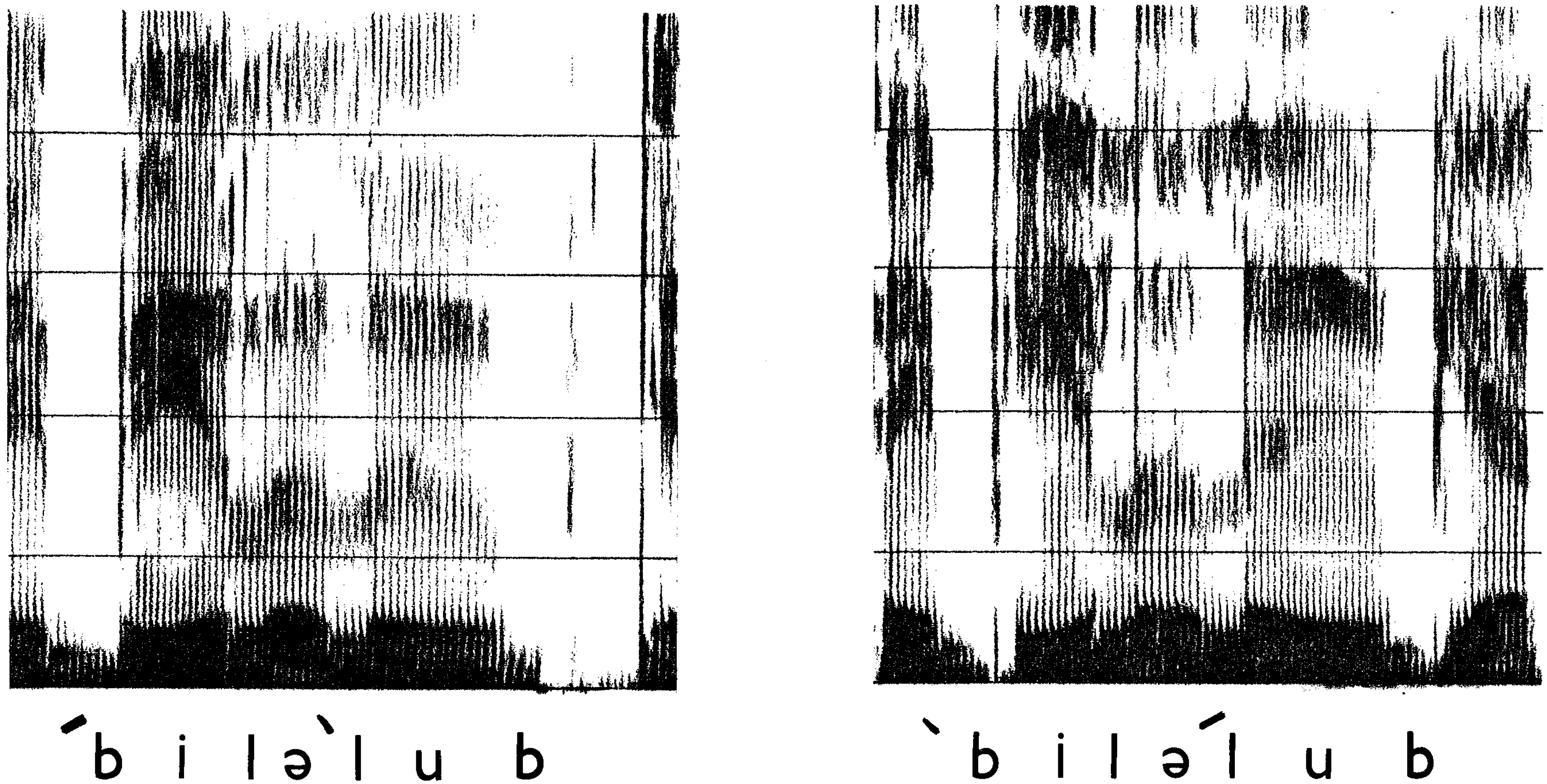


Figure 13. Spectrograms illustrating formant movement during stressed and weakly stressed V1 = /i/, in the context of /l/, for speaker one.

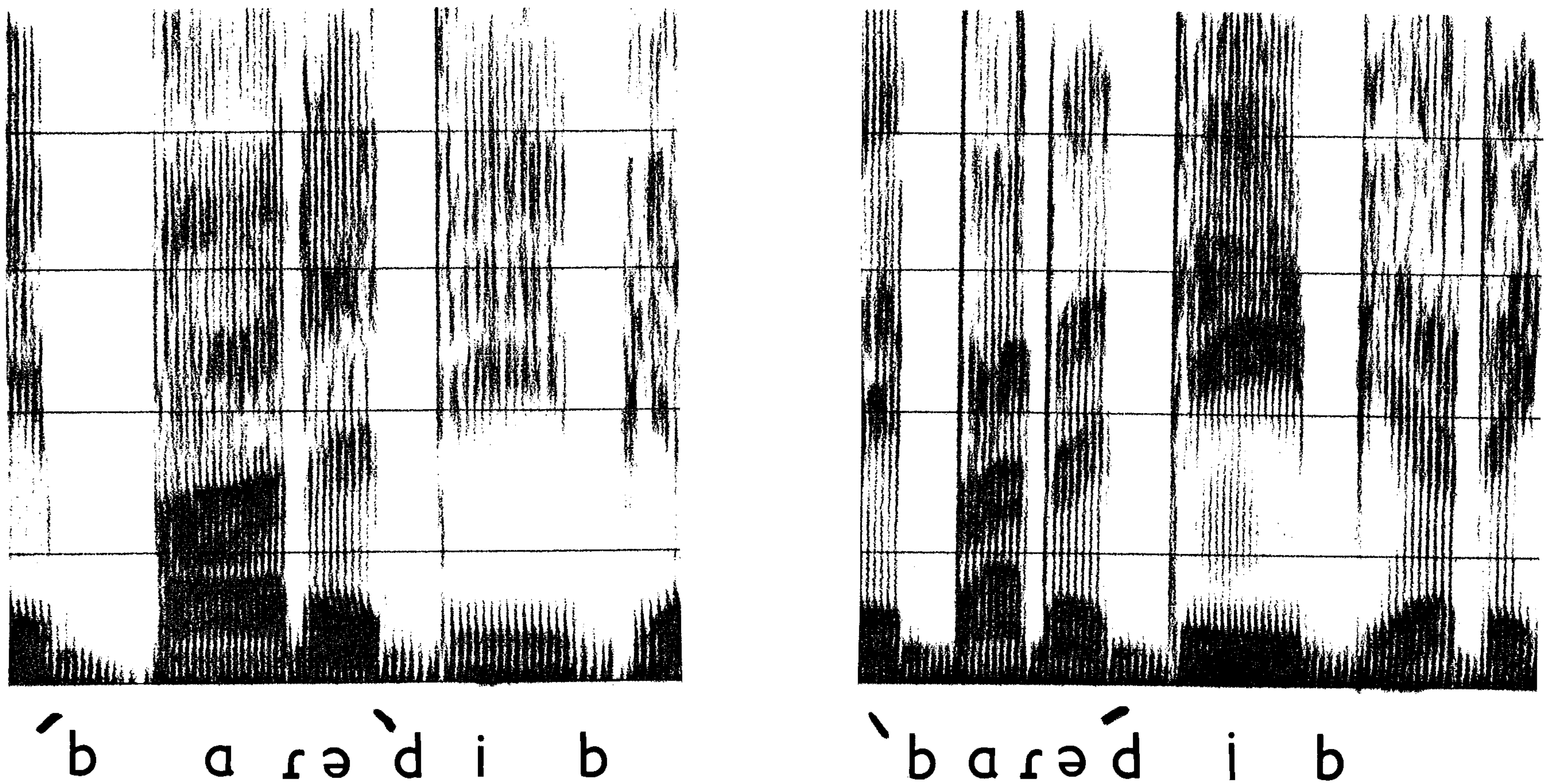


Figure 14. Spectrograms illustrating formant movement during stressed and weakly stressed V1 = /a/ in the context of /d/, for speaker one.

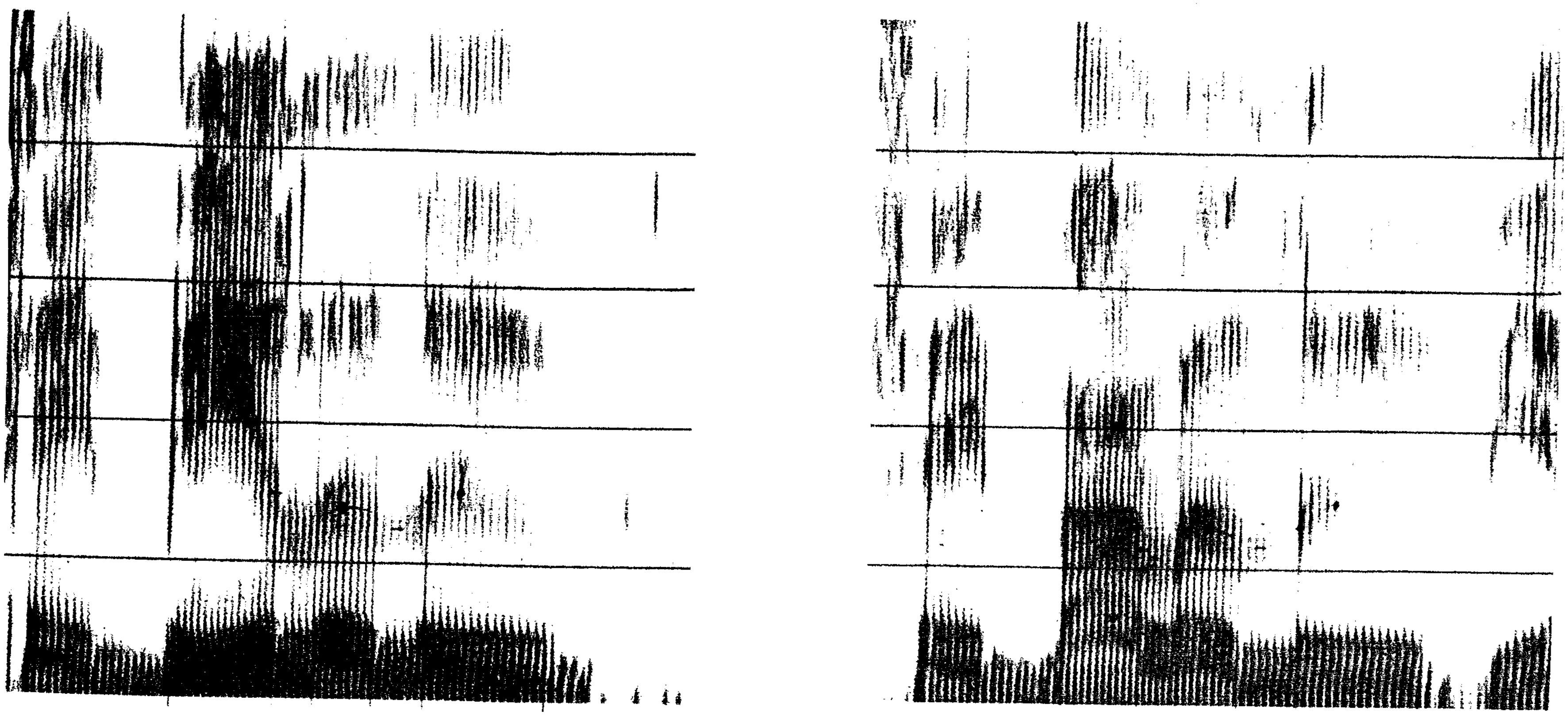


Figure 15. Spectrograms illustrating the effect of V1 on following vowels and /l/'s, for speaker one.

must be due to some other factor, like the sheer strength of the carryover effects.

In conclusion, degree of constraint of tongue body movement in the consonants /d/ and /l/ appears to be one of the strongest factors affecting patterns of vowel–to–vowel coarticulation in English. In addition, most stress effects on coarticulation are moderated by these consonant differences.

NOTES

[1] P. Keating has aptly pointed out that this situation is counterevidence to Fowler's "overlap" scenario in that the stressed vowel is overlapping very little with the schwa.

ACKNOWLEDGEMENTS

This work was supported by NSF Grant No. BNS–8418580 to Patricia Keating. Portions of this work were presented at the 111th meeting of the Acoustical Society of America, Cleveland, Ohio 1986.

REFERENCES

Bell–Berti, F.; Harris, K. S.: "Some Aspects of Coarticulation." *Haskins Laboratories Status Report, Speech Res.* SR–45/46 (1976).

Bladon, R.; Al–Bamerni, A.: "Coarticulatory resistance in English /l/." *J. Phonet.* 4: 135–150.

Fowler, C.A.: "Production and perception of coarticulation among stressed and unstressed vowels." *J. Speech. Hear. Res.* 46:127–139 (1981a).

Fowler, C.A.: "A Relationship between Coarticulation and Compensatory Shortening." *Phonetica* 38: 35–50 (1981b).

Gay, T.: "Articulatory movements in VCV sequences." *J. Acoust. Soc. Am.* 62: 183–193 (1977).

Henke, W.L.: *Dynamic Articulatory Model of Speech Production Using Computer Simulation.* MIT dissertation (1966).

Keating, P.: "CV Phonology, Experimental Phonetics, and Coarticulation." Presented at Colloque Phonologie Pluri–Lineaire, Lyon (1985).

Krakow, R.A.; Manuel, S.: "Vowel–to–Vowel coarticulation in Swahili." *J. Acoust. Soc. Am.* 74, Suppl. 1: S91 (1983).

Magen, H.: "Vowel–to–vowel Coarticulation in English and Japanese." *J. Acoust. Soc. Am.* 75, Suppl. 1: T11 (1984).

Magen, H.: "Prosodic effects in Vowel–to–Vowel Coarticulation." *J. Acoust. Soc. Am.* 77, Suppl. 1: SS2 (1985).

Ohman, S.: "Coarticulation in VCV utterances." *J. Acoust. Soc. Am.* 39: 151–168 (1966).

Parush, A.; Ostry, D.J.; Munhall, K.G.: "A Kinematic Study of lingual coarticulation in VCV sequences." *J. Acoust. Soc. Am.* 74: 1115–1123 (1983).

Recasens, D.: "Vowel–to–vowel coarticulation in Catalan VCV sequences." *J. Acoust. Soc. Am.*: 1624–1635 (1984).

Wuxi Tone Sandhi: From Last to First Syllable Dominance

Marjorie K.M. Chan & Hongmo Ren

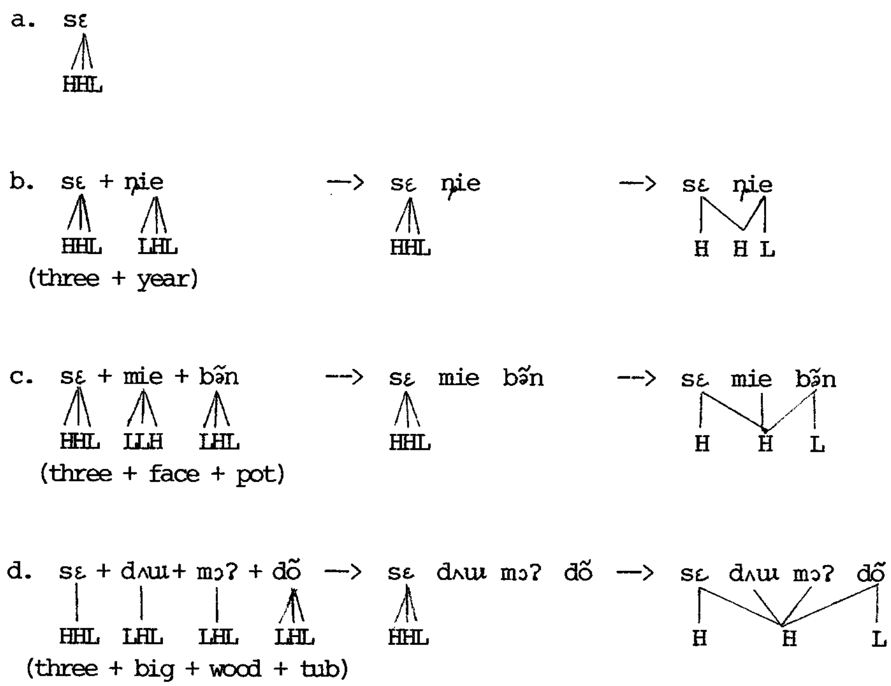
0. Introduction.

Wuxi 無錫 is a Northern Wu dialect of Chinese spoken in Wuxi district, the administrative center of which is Wuxi City, located about thirty miles northwest of Suzhou, and another fifty miles further from Shanghai. The Wuxi dialect exhibits some interesting behaviour with respect to its tone sandhi patterning that has not been reported in the literature.[1] The dialect has two strategies for tone sandhi patterning -- pattern extension and pattern substitution -- both of which involve the first syllable as "dominant", as determining the tone patterning in the sandhi span (Yue-Hashimoto 1980). Pattern extension simply involves taking the tone melody that was on the first syllable and extending that pattern onto the entire tone sandhi span. Pattern substitution requires one additional step before the tone-spreading; namely, the replacement of the tone melody on the first syllable with a new melody before its spread onto the tone sandhi domain. While pattern extension has already been encountered in the literature on Tangxi (Kennedy 1953) and Shanghai (Sherard 1980, Zee and Maddieson 1980), pattern substitution as described above has not, to our knowledge, been reported in the literature. In this paper, we argue that the pattern substitution in Wuxi originally involved last syllable dominance, and further propose that it is the shift from last to first syllable dominance that accounts for the unusual tone sandhi behaviour in modern Wuxi.

To better understand the difference between pattern extension and pattern substitution in the dialect, the two strategies are illustrated in (1) and (2) respectively on the following page. The examples are presented autosegmentally, without displaying the syllable and CV tiers. A three-toneme sequence for the tone patterns, or tone melodies, in Wuxi is assumed here, proposed originally in Ren (1982). The language-specific rule he formulated for Wuxi is as follows: "Associate the first toneme with the first syllable and the last toneme with the last syllable, and associate the middle toneme with all syllables including the first and last syllable." The tri-toneme sequence functions as a timing unit, which becomes more salient in longer tone sandhi spans. This analysis, therefore, does not take the Obligatory Contour Principle (OCP) into consideration. Given the toneme-mapping rule in Wuxi, "LLH" and "LHH", for example, yield different pitch contours. The difference lies in the location of the pitch rise: in a tone sandhi span with the LHH melody, the rise occurs immediately, on the first syllable; in a span with LLH, on the other hand, the rise is delayed until the final syllable.

Illustrating pattern extension, (1) gives the tone pattern on the first syllable as "HHL", which serves as the tone pattern for the entire tone sandhi domain; the tone pattern is thus extended over the whole sandhi span. The tones on the non-initial syllables of the sandhi span are deleted, and the pitch contour, or tone pattern, of the first syllable is then mapped onto the entire span. The example includes a broad phonetic representation in which the HHL tone pattern is converted to tone numbers ('1' to '5' representing ascending pitch height) for a more traditional rendition of the tone sandhi forms. For typographical ease, the tone numbers and 'h' for aspiration are not superscripted.

(1) Example of Pattern Extension.



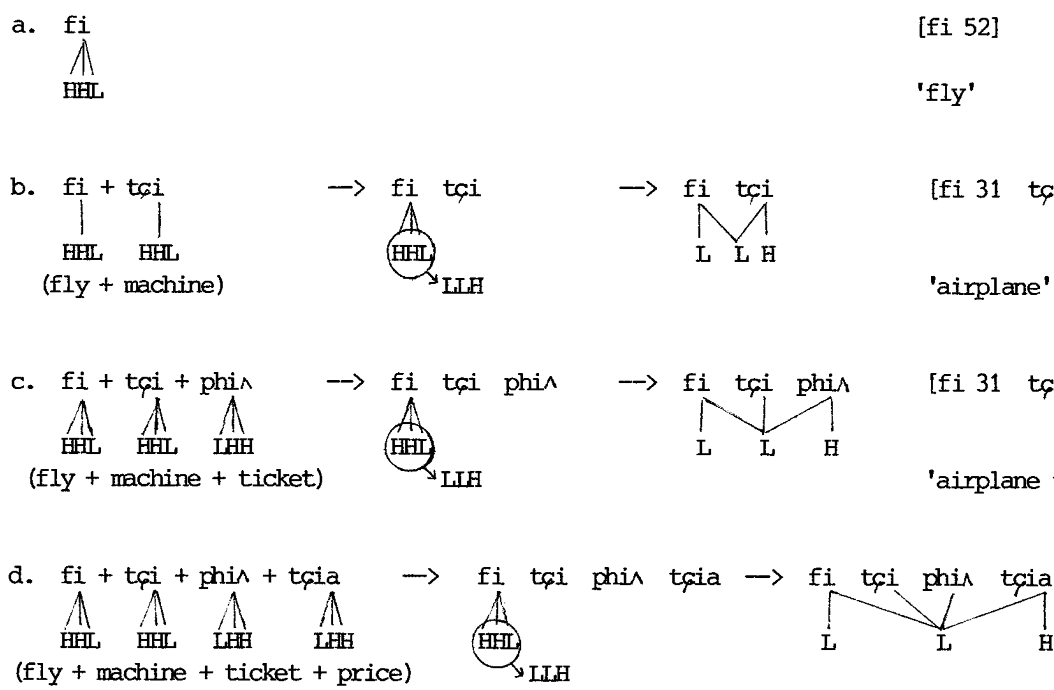
a. [sɛ 52] 'three'

b. [sɛ 44 ɲie 21] 'three years'

c. [sɛ 44 mie 44 bə̃n 21] 'three pots of'

d. [sɛ 44 dʌɯ 44 mɔʔ 4 dɔ̃ 21] 'three big wooden tubs'

(2) Example of Pattern Substitution.



a. [fi 52] 'fly'

b. [fi 31 tɕi 14] 'airplane'

c. [fi 31 tɕi 11 phiʌ 14] 'airplane ticket'

d. [fi 31 tɕi 11 phiʌ 11 tɕia 14]
'airplane ticket price'

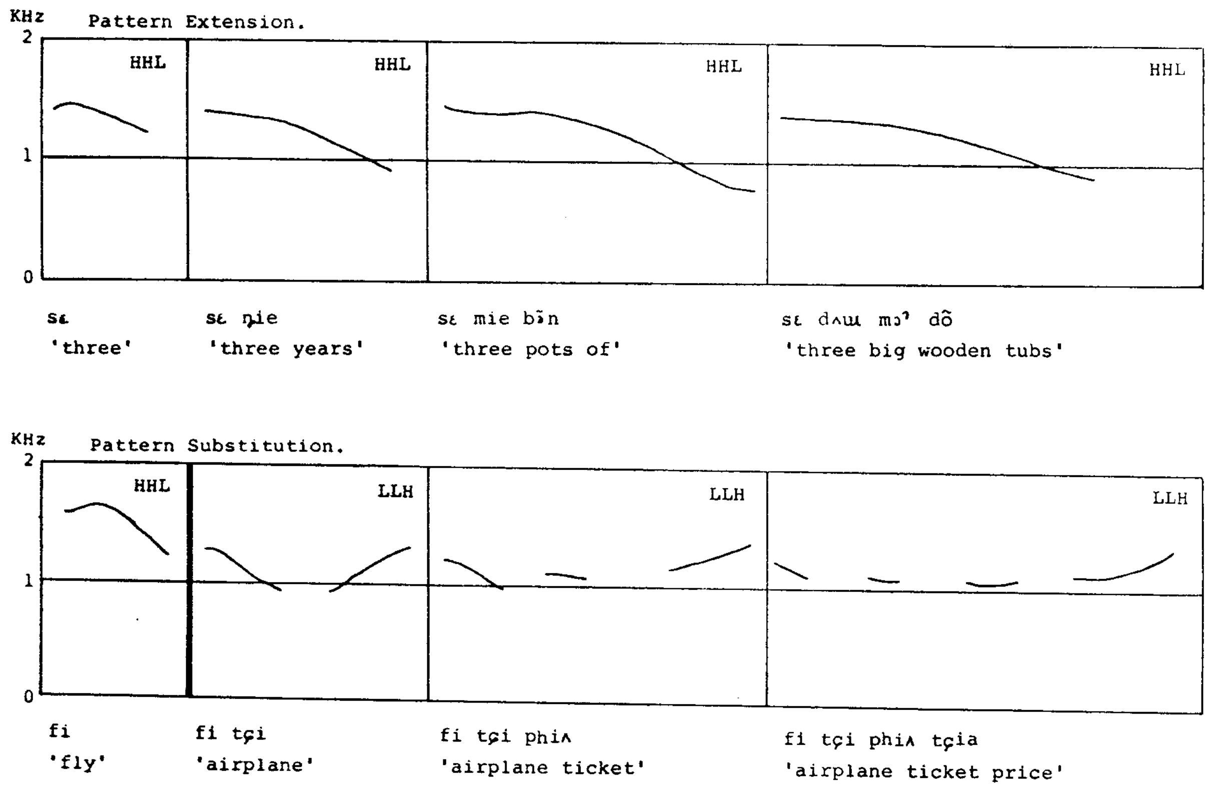


Figure 1. Two Strategies in Wuxi: Pattern Extension and Pattern Substitution. (Tenth harmonic on narrow-band spectrograms)

The tone patterns in (1) can also be observed in Figure 1, where the tenth harmonic on narrow band spectrograms is traced, revealing identical pitch contours for all four examples.[2] All the figures presented here are tracings of the tenth harmonic on narrow-band spectrograms to obtain the pitch contours.

In the case of pattern substitution, as shown in (2), the tone pattern of the tone sandhi span is not (i.e., not typically) a copy of the first syllable's citation tone. The isolation syllable in (2a), for example, has the HHL tone pattern, but the tone pattern of the sandhi forms in (2b) through (2d) is LLH. The original tone pattern on the first syllable is thus replaced by another tone pattern. The pattern substitution in (2) can also be seen in Figure 1.

The two strategies affect different types of morphological and syntactic structures, as summarized in (3). Pattern extension involves a minority of cases, and is limited to reduplicated verbs, verbs followed by a directional complement, reduplicated nouns used in baby talk, resultative verbs, as well as expressions containing a number and a classifier. Pattern substitution, on the other hand, affects the remaining types of constructions, including reduplicated nouns, compounds, and phrases. Therefore, of the two types of tone sandhi patterning in Wuxi, the main one is pattern substitution, which is also the one of primary concern in this paper. The proposal here is that the tone sandhi patterning in Wuxi was originally last syllable dominant, affecting the majority of the constructions in the dialect. Only in a minor set of morphological structures was the tone sandhi patterning first syllable dominant.

(3) Summary of Tone Sandhi in Wuxi.

Strategy	Syll. Dominance	Constructions
Pattern Extension	first	reduplicated verbs, reduplicated nouns in baby talk verbs with resultative or directional complements expressions with 'number + classifier'
Pattern Substitution	first	regular reduplicated nouns, compounds, phrases

The underlying assumption is that there is a split in the dialect of first and last syllable dominance. Such a split is not unique to Wuxi. The first well-known case is reported by Kennedy (1953) for the Tangxi (or Tangsic) dialect. Yue-Hashimoto (1980) also cites Shaoxing (Wang 1959) as exhibiting both first and last syllable dominance within a single dialect. It is noteworthy that Wuxi, Tangxi, and Shaoxing all belong to the northern Wu dialect group. Regarding Tangxi and Shaoxing, Yue-Hashimoto suggests that they represent a transitional type of dialect. On the basis of the Wuxi data, we would argue, instead, that Tangxi and Shaoxing are in fact conservative, having preserved the two different types of tone sandhi patterning. Wuxi, on the other hand, has lost that contrast of first and last syllable dominance affecting different morphological and syntactic constructions, with tone sandhi patterning having shifted to that involving primarily first syllable dominance. It is, nonetheless, the effects of last syllable dominance that accounts for the pattern substitution found in modern Wuxi. Only vestiges of last syllable dominance remain in the dialect today.

Prior to developing the case for the shift in dominance in Wuxi, some preliminaries are needed, since many of the basic facts concerning Wuxi tones and tone sandhi are not familiar to scholars. Thus, for the sake of clarity, background on the Wuxi tonal system and its tone patterns on monosyllables and sandhi forms will be presented in the next section.

1. Tones and Tone Patterns in Wuxi.

Following traditional practice, the citation tones in Wuxi are presented in (4) below, arranged, firstly, according to the four historical tone categories, Ping ('Even'), Shang ('Ascending'), Qu ('Departing'), and Ru ('Entering'), and, secondly, according to the register split into the Yin (Upper) and Yang (Lower). The combination yields eight tones, which will be referred to as "T1", "T2", and so forth. An example with each of the tones is given in (4), and a display of their pitch contours is provided in Figure 2. (For the initials and finals in Wuxi, see the appendix.)

(4) Wuxi Citation Tones.

	Tone		Tone Number	Example	
a.	T1	Yinping	/52/	tõ 52	'east'
b.	T2	Yangping	/213/	dõ 213	'together'
c.	T3	Yinshang	/313/	tõ 313	'to understand'
d.	T4	Yangshang	/131/	dõ 131	'to move'
e.	T5	Yinqu	/34/	tõ 34	'freeze'
f.	T6	Yangqu	/213/	dõ 213	'hole'
g.	T7	Yinru	/53/	tɔʔ 53	'sincere'
h.	T8	Yangru	/13/	dɔʔ 13	'only'

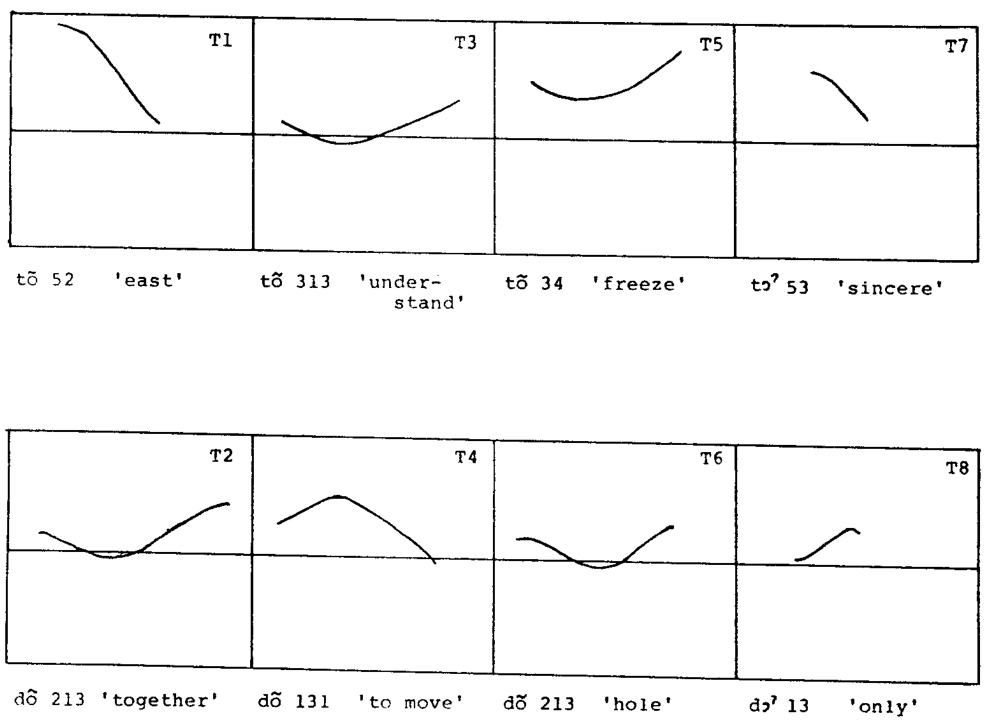


Figure 2. The 8 citation tones in Wuxi.

A few comments are needed concerning (4). First of all, Wuxi has voiced initials on Yang (lower) register syllables, accounting for the initial or overall lower pitch on these syllables. Concerning individual tones, T1 has an alternant form which is high level, phonetically [44]. In slow speech, T5 often has a slight dip, and is then phonetically [324]. T7 and T8 are short, with glottal stop closure. T7 is furthermore often truncated, and is simply high ([5]). Syllable spans with initial T7 followed by T7 or T8 are high, with or without a final pitch drop. T3 is [213] sometimes, but can be distinguished from T4 and T6 in that it occurs on syllables with a voiceless onset. In citation context, T2 and T6 are regularly non-contrastive. The two tones were still distinct in Chao's 1928 account of Wuxi, where T2 was /13/, and T6 /213/.[3] In our data, however, T2 monosyllables (citation forms) are regularly non-distinct from their T6 counterparts. Despite tonal merger on citation forms, the two tones in modern Wuxi continue to behave differently in tone sandhi context, a phenomenon that is by no means unique to Wuxi.[4] Distinctions between T2 and T6 in Wuxi surface in tone sandhi spans, as shown in (5). (5a) illustrates pattern extension involving reduplicated nouns occurring in baby talk, whereas (5b) involves regular compounds undergoing pattern substitution. As in the earlier examples, we are including tone numbers in the sandhi forms as a guide to the pitch shapes, which will be omitted once we have established the inventory of tone patterns in the dialect.

(5) a. Pattern Extension with Reduplicated Nouns in Baby Talk.

T2: dɛ 12 dɛ 41 'play (e.g., piano)'
(dɛ 213 'play, pluck')

T6: dɛ 32 dɛ 24 'egg'
(dɛ 213 'egg')

b. Pattern Substitution with Compounds/Phrases.

T2: lɛi 24 vã 21 '(two or more) storey building'
(lɛi 213 'storey')

T6: lɛi 12 vã 41 'leaky house'
(lɛi 213 'leak')

The examples demonstrate, above all, the need to treat T2 and T6 as having distinct underlying tone patterns that are neutralized in citation forms.

The underlying tone patterns for the eight tones in Wuxi are given in (6). A tri-toneme melody is assigned to each citation tone. Phonological distinctions can be accounted for using a contrast of only two pitch heights: H and L, to be interpreted as (relatively) high and non-high respectively. Recall that T7 and T8 are short, terminating in a glottal stop.

(6) Wuxi Tones and Tone Patterns.

a.	T1	/52/	HHL
b.	T2	/213/	LHL
c.	T3	/313/	LLH
d.	T4	/131/	HHL
e.	T5	/34/	LHH
f.	T6	/213/	LLH
g.	T7	/53/	HHL
h.	T8	/13/	LHL

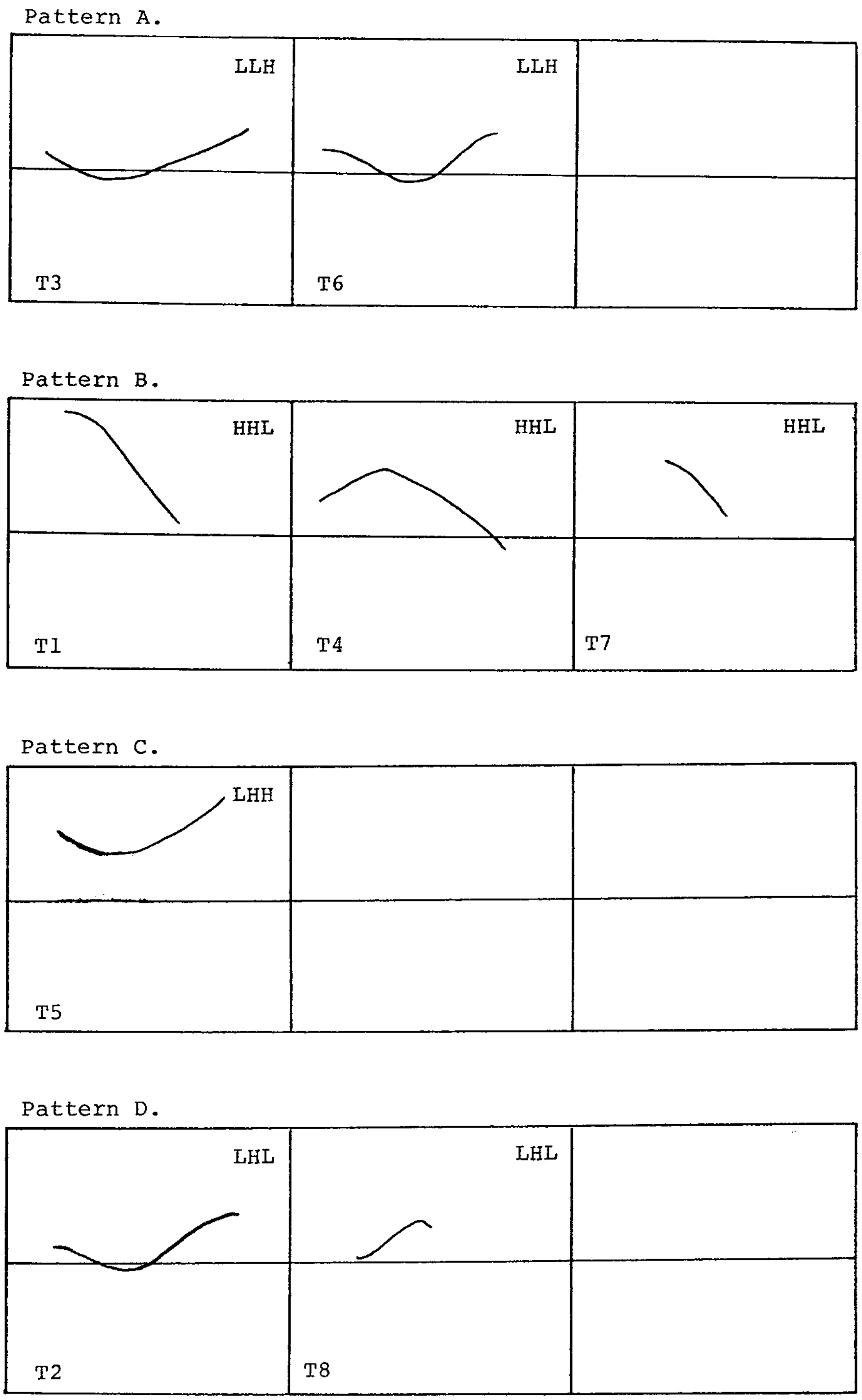


Figure 3. The four tone patterns on the eight citation tones in Wuxi.

Four distinct tone patterns emerge from (6):

(7) The Four Tone Patterns.

Pattern A:	LLH	T3	(313)
		T6	(213)
Pattern B:	HHL	T1	(52)
		T4	(131)
		T7	(53)
Pattern C:	LHH	T5	(34)
Pattern D:	LHL	T2	(213)
		T8	(13)

For comparative purposes, the pitch contours of the eight tones in Figure 2 are re-arranged on the preceding page in Figure 3, according to the four patterns in (7). Observe that the treatment of T4 (/131/) in (7) as having Pattern B, an HHL tone melody, needs some explanation. T4 actually has an initial low pitch due to the pitch lowering effect of voiced initials on T4 syllables. The initial rise is completely predictable and can thus be treated as phonetic detail. It is also important to note that the HHL pattern with voiced onset is distinct from LHL with or without initial voicing. A comparison of the two tone patterns is presented in Figure 4, using the examples in (8) and (9) on the next page. The HHL patterns shows a sharp rise followed by a gradual decline in pitch, while the LHL pattern is almost like a mirror image, exhibiting a delayed rise and then a sharp fall.

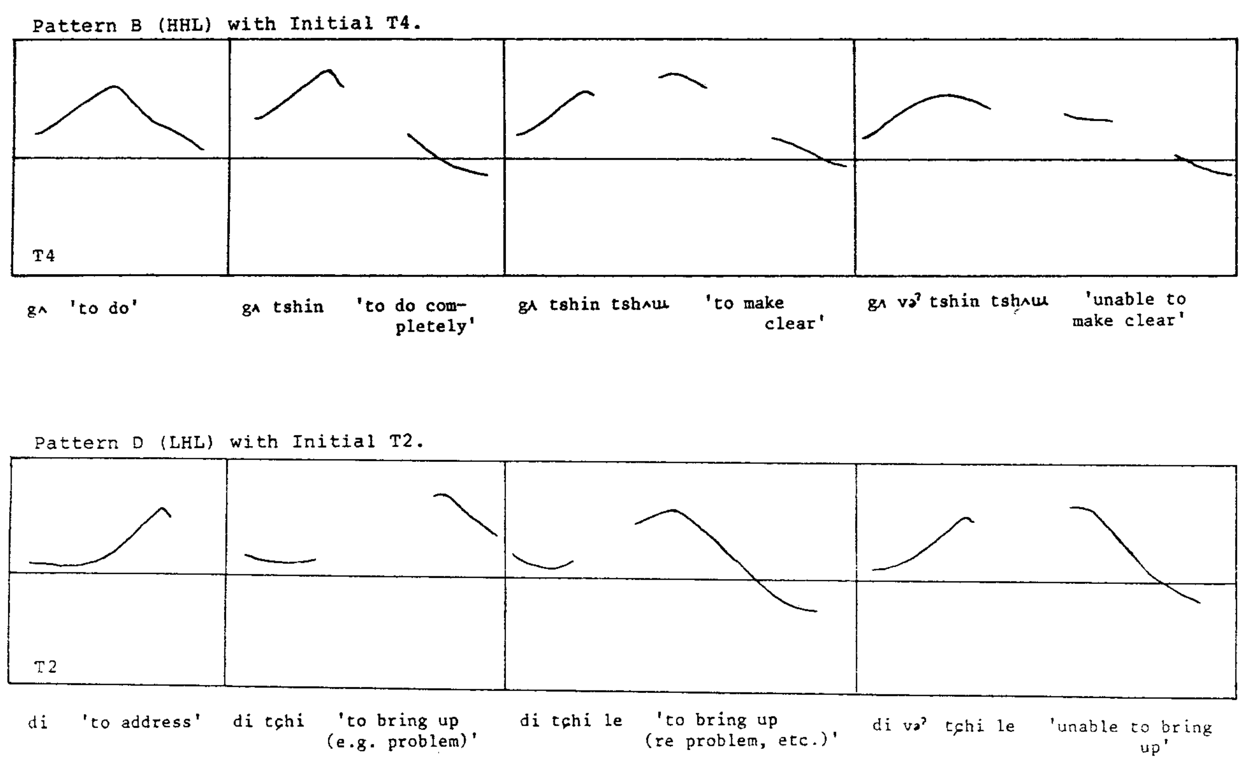


Figure 4. Comparison of Patterns B and D with Voiced Onset.

(8) Pattern B (HHL) with Initial T4. (Pattern Extension)

a. gʌ [gʌ 131]
HHL 'to do'

b. gʌ tshin [gʌ 24 tshin 21]
H H L 'to do completely'

c. gʌ tshin tshʌɯ [gʌ 24 tshin 44 tshʌɯ 21]
H H L 'to make clear'

d. gʌ vəʔ tshin tshʌɯ [gʌ 24 vəʔ 4 tshin 44 tshʌɯ 21]
H H L 'unable to make clear'

(9) Pattern D (LHL) with Initial T2.

a. di [di 213]
LHL 'to address'

b. di tɕhi [di 12 tɕhi 41]
L H L 'to bring up (e.g., problem)'

c. di tɕhi le [di 12 tɕhi 44 le 41]
L H L 'to bring up (re problem, etc.)'

d. di vəʔ tɕhi le [di 12 vəʔ 4 tɕhi 44 le 41]
L H L 'unable to bring up'

The other tones with a potentially disputable tone pattern are T2 and T8, assigned Pattern D, with an LHL melody. The assignment of LHL is based primarily on the pitch shape of polysyllabic forms that had undergone pattern extension, where the initial syllable originally bore T2 or T8. A set of examples with the LHL tone melody for initial T2 sandhi spans was given in (9). A set for initial T8 sandhi spans is presented in (10) below. Pitch contours of (9) and (10) are given in Figure 5, showing an LHL melody on the polysyllabic forms.

(10) Pattern D (LHL) with Initial T8.

a. ɦɔʔ — LHL — [ɦɔʔ 13] — 'to learn'

b. ɦɔʔ zə̃n — L H L — [ɦɔʔ 12 zə̃n 41] — 'to achieve learning (s.t)'

c. ɦɔʔ zə̃n kõ — L H L — [ɦɔʔ 12 zə̃n 44 kõ 41] — 'to be successful in learning (s.t.)'

d. ɦɔʔ vəʔ zə̃n kõ — L H L — [ɦɔʔ 12 vəʔ 4 zə̃n kõ 41] — 'to be unsuccessful in learning (s.t.)

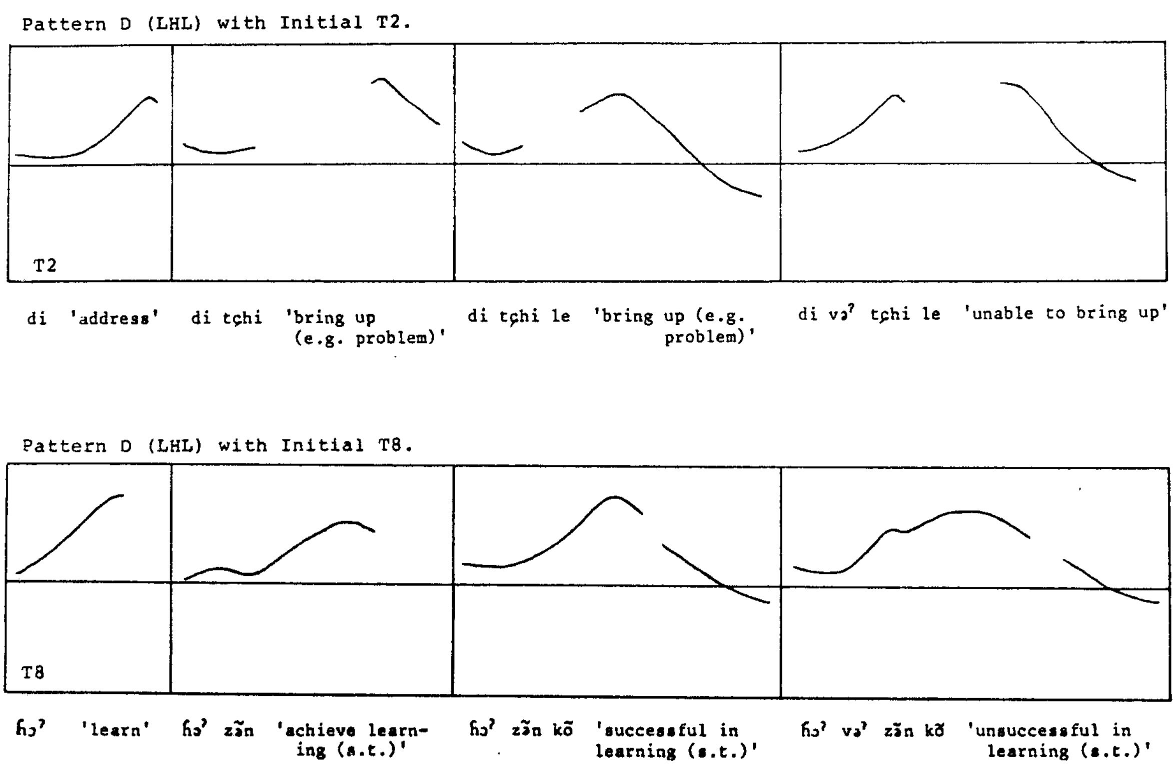


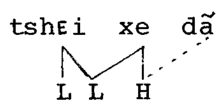
Figure 5. Pattern D (LHL) on Sandhi Spans with Initial T2 and T8 Syllables.

Concerning T2, recall also that (5) showed the need to contrast T2 with T6 in sandhi forms even though the two tones are merged in citation form. The assignment of LHL for T2 thus serve additionally to distinguish it from T6, which is given an LLH tone pattern. In the examples with pattern extension in (5a), one can also observe the LHL tone pattern on the word 'play (e.g., piano)', with reduplicated T2 syllables, contrasting with the LLH tone pattern on the word 'egg', with reduplicated T6 syllables.

The four tone melodies -- LLH (A), HHL (B), LHH (C), and LHL (D) -- account for the monosyllables and tone sandhi spans in Wuxi. Figure 6 presents the four tone patterns mapped onto words or phrases of one to four syllables in length. A gradual pitch drop at the end of utterances with Pattern C (LHH) is due to intonation factors, which are more discernible in Ren's speech than in Wang's. The initial mid-low instead of very low pitch at the onset of the LLH tone can be treated as due to the pitch-raising effect of voiceless consonants.

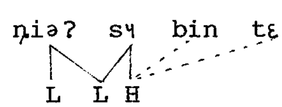
In addition to the regular mapping of the four tone patterns, there is also a slight variation in the mapping for Pattern A (LLH) occurring on tri- and quadrisyllabic forms. For easy reference, this irregular mapping of Pattern A will be identified as Pattern A'. It differs from Pattern A in that the rise occurs on the second syllable and not delayed to the last syllable of the tone sandhi span. Thus, the third toneme must be linked to the second syllable before the regular association rule in Wuxi applies. This modified mapping of Pattern A occurs in limited environments. Examples are given in (11), with the pitch contours included in Figure 6. Informally, one can regard (11a) as exhibiting an "L-LH-H" sequence, and (11b) an "L-LH-H-H" sequence. In the examples, the spreading of H onto the third and fourth syllables (shown by the use of dotted lines), occurs in order not to violate the original version of the Wellformedness Condition, which requires all vowels to be associated with a tone.

(11) Pattern A': A Variation in the Mapping of Pattern A (LLH).

a. 

[tshɛi 21 xe 13 dã 33]

'begonia'

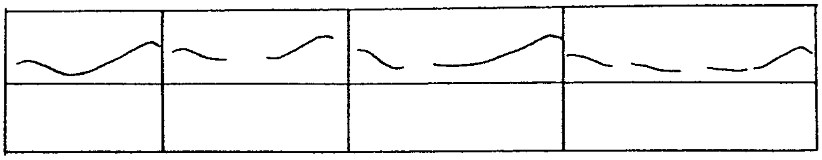
b. 

[ȵiəʔ 21 sɥ 12 bin 33 tɛ 33]

'thermos bottle lining'

The above provides the basic background that we need to discuss the shift in Wuxi of regular compounds and phrases from last syllable dominance to first syllable dominance. At the same time, the inclusion of the pitch contours traced from narrow-band spectrograms provide support for the four basic tone patterns posited here for modern Wuxi.

Pattern A - LLH.



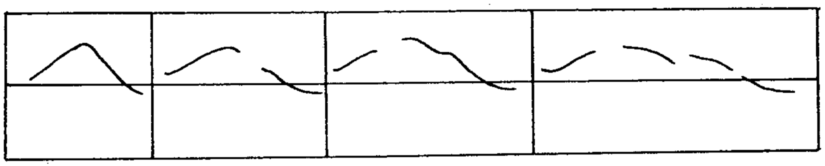
siã 'think'

sie sã 'teacher'

tshin ɕiõ di 'blood brother'

tɕhĩə̃n phiʌ phiʌ li 'floating lightly in air'

Pattern B - HHL.



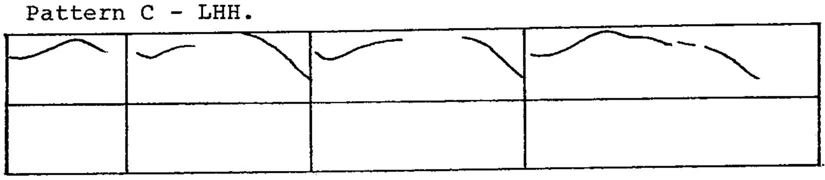
lʌ 'old'

zɛ̃ bʌ 'long robe'

no kʌ iə̃n 'male sophrano'

liɛ̃ sEi sEi li 'cooled by breeze'

Pattern C - LHH.



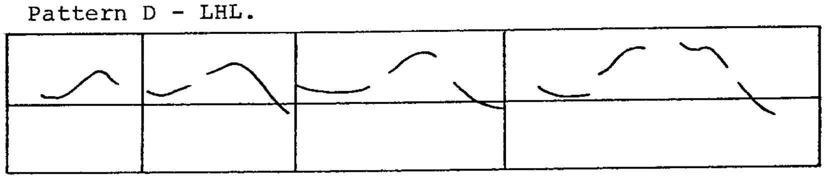
si 'four'

tsʌɯ tsõ 'ancester'

siʌ ɦu tsɿ 'small moustache'

sɥ liə̃n liə̃n li 'fresh as if recently watered'

Pattern D - LHL.



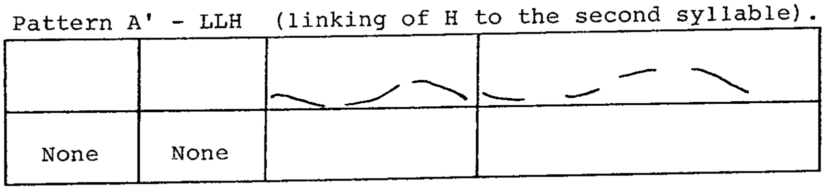
ȵiəʔ 'hot'

na dã 'milk candy'

lʌ di fã 'old place'

mɛ thə̃n thə̃n li 'slowly, step by step'

Pattern A' - LLH (linking of H to the second syllable).



tshEi xe dã 'begonia'

ȵiəʔ sɥ bin tɛ 'thermos bottle lining'

Figure 6. The four tone patterns in Wuxi on mono- to quadrisyllabic forms.

2. From Last to First Syllable Dominance.

We have proposed that Wuxi historically had two types of tone sandhi patterning based on syllable dominance: pattern extension applied in tone sandhi spans where the first syllable was dominant, and pattern substitution applied in cases of spans with last syllable dominance. At this point it is worth conducting a brief study of the Tangxi (Northern Wu) dialect recorded by Kennedy (1953), which has preserved the first and last syllable dominance distinction. Tangxi, thus, serves as a kind of prototype.

A word should first be said concerning the correlation between syllable dominance and stress in Tangxi, shown in (12), which is not so salient in Wuxi. Included in the chart for reference are the morphological and syntactic constructions affected by the two different tone sandhi strategies. The term "pattern substitution" is used very loosely in (12) for comparative purposes.

(12) Summary of (Disyllabic) Tone Sandhi in Tangxi.

Strategy	Syll. Dominance	Stress	Constructions
Pattern Extension	first	initial	compounds, NP's
Pattern Substitution	last	final	V-O expressions

Turning now to first syllable dominance involving pattern extension in Tangxi, the rule for generating the correct tone sandhi pattern is identical to that in modern Wuxi: "The original tone of the first syllable is spread over the two combined syllables." (Kennedy 1953:369). An informal, autosegmental treatment is given in (13), using disyllabic examples from Kennedy. Only L and H tonemes are needed for the analysis. As shown below, the tone melody of the second syllable is deleted, while that of the first syllable is spread onto both syllables. In (13b), the LH melody permits two options in its linkages, with a sequence of either "L-H" or "L-LH". For reference purposes, the tone numbers assigned by Kennedy are included in the examples.

(13) Pattern Extension in Tangxi.

```
a.  tzong 33 + nyin 24   -->  tzong  nyin -->  tzong 33  nyin 33
      |         /\              |                |_________/
      L         L H             L                L
    (middle  +  man)                           'middleman'

b.  tseo 24  + sea 51    -->  tseo   sea  -->  tseo 22   sea 44
     /\         /\             /\                |         |
     L H        H L            L H               L         H
   (smelly   + water)                OR
                                          -->  tseo 22   sea 24
                                                 |_________|
                                                 L         H
                                    'smelly water' (disinfectant)
```

What we are calling "pattern substitution" in (12) is not a mirror image of pattern extension in Tangxi. The second syllable actually retains its citation tone, while the tone on the first syllable is neutralized to simply low; that is, phonetically lower than the /33/ tone, and is, hence, treated here as [22]. One could assume a default low tone in Tangxi that is inserted in preterminal position of last-syllable dominant tone sandhi spans. In any case, due to tone neutralization in preterminal position, to 'buy wine', for example, becomes homophonous with to 'sell wine'.

(14) Pattern Substitution in Tangxi.

```
a.  ma 24 + tseo 51    -->  ma  tseo 51   -->  ma 22   tseo 51
    /\       /\                  /\             :        /\
    L H      H L                 H L           (L)       H L
   (sell  +  wine)                             'sell wine'

b.  ma 51 + tseo 51    -->  ma  tseo 51   -->  ma 22   tseo 51
    /\       /\                  /\             :        /\
    H L      H L                 H L           (L)       H L
   (buy   +  wine)                             'buy wine'
```

In Wuxi, the neutralization of the tones on the initial syllable did not take place. Instead, the initial tone is substituted by another tone, and is thus more similar to the paradigmatic tone replacement found in Min tone sandhi, with last syllable dominance (e.g., Xiamen (Bodman 1955), Fuzhou (Chan 1985)). The pattern substitution in Wuxi, first presented in Ren (1982) is summarized in Table 1 below:

Table 1. Pattern Substitution in Wuxi.

(A = LLH, B = HHL, C = LHH, D = LHL)

First Syllable	Original Tone Pattern on the First Syllable	Replacement Tone Pattern on Sandhi Spans of 2 to 4 Syllables
T1	B	A
T2	D	B
T3	A	C
T4	B	D
T5	C	B
T6	A	D
T7	B	A
T8	D	D

Table 1 reads as follows: tone sandhi spans of two to four syllables in length are organized according to the historical tone category of the first syllable, namely, T1, T2, T3, and so forth, as presented in the first column. The original tone pattern on the first syllable is then specified in the second column, while the third and final column shows the tone pattern which replaces it, and serves as the tone pattern for the entire tone sandhi span. Thus, a T1 syllable in isolation, for example, has tone pattern B, which is replaced by that of A when the T1 syllable occurs as the first syllable in a tone sandhi span. It is this new tone pattern A that is spread onto the entire tone sandhi span, wiping out the tone patterns that were originally on the remaining syllables of the span. Table 1 thus provides a clear picture of how pattern substitution in Wuxi operates, and why the tone sandhi in modern Wuxi is characterized as exhibiting first syllable dominance.

Two questions might be raised at this point. First of all, if the tone sandhi patterning of regular compounds and phrases in Wuxi had always been first syllable dominant, what led the dialect to undergo the unusual operation of substituting another tone on the first syllable? And secondly, what are the bases for our claim that these compounds and phrases in Wuxi were originally last syllable dominant? These two questions are clearly intertwined. To address them, we need first to say a few words about the length of the tone sandhi spans in Tangxi and Shaoxing, because the observations concerning these two Northern Wu dialects have a direct bearing on the Wuxi case.

With regard to Tangxi, Kennedy (1953) includes examples of tri- and quadrisyllabic tone sandhi spans for pattern extension, which show that the longer strings behave similarly to the disyllabic cases. However, longer sequences of pattern substitution are ostensibly absent in his article. It seems that "last syllable dominance", applying to pattern substitution in Tangxi, refers more precisely to disyllabic spans. In Shaoxing, also reported to have a dual strategy for tone sandhi patterning, last syllable dominance is found in disyllabic phrases, and in only a subset of trisyllabic numeral combinations.

The above observations suggest that the occurrence of last syllable dominance may be very restricted in both Tangxi and Shaoxing, so that if they occur at all, they tend to be limited to disyllabic forms. In Wuxi, it is precisely in the second syllable position in polysyllabic tone sandhi domains that we found vestiges of our last syllable dominance. Whether this had ever extended beyond two syllable spans or not can only be speculative at this time.

Evidence of last syllable dominance at some stage in the tonal history of Wuxi can be obtained by conducting a very systematic study of the tone sandhi patterning in the dialect. Such a study reveals that there are a number of exceptions to the pattern substitution given in Table 1. Many of these exceptions are not random, however, but reflect, instead, some tonal conditioning. There are splits in the tone sandhi patterning that are dependent upon the original tone on the second syllable of tone sandhi spans. To be even more precise, the particular historical tone category to which the second syllable belonged -- Ping, Shang, Qu, or Ru -- also partially conditions the substitution of a tone pattern. Register distinction, on the other hand, plays no role in that process. The picture that emerges is presented in Table 2, which is a revision of the previous table. In Table 2, the original tone pattern on the first syllable is given in parentheses on the first column. Moreover, since the Yin-Yang register distinction is not relevant, one could simply treat both T1 and T2 as Ping tones, T3 and T4 as Shang, T5 and T6 as Qu, and T7 and T8 as Ru.

Table 2. Pattern Substitution in Wuxi.

(Ping = T1,T2. Shang = T3,T4. Qu = T5,T6. Ru = T7,T8)

<table>
<tr><th>First Syllable</th><th>Two Syllable Spans</th><th>Three and Four Syllable Spans</th></tr>
<tr><td rowspan="2">T1 (B)</td><td rowspan="2">A</td><td>A (second syllable is Ping or Ru)</td></tr>
<tr><td>A' (second syllable is Shang or Qu) *</td></tr>
<tr><td>T2 (D)</td><td colspan="2">B</td></tr>
<tr><td>T3 (A)</td><td colspan="2">C</td></tr>
<tr><td rowspan="2">T4 (B)</td><td>A (2nd syl. Shang or Qu)</td><td>A' (second syl. is Shang or Qu)</td></tr>
<tr><td colspan="2">D (second syllable is Ping or Ru)</td></tr>
<tr><td>T5 (C)</td><td colspan="2">B</td></tr>
<tr><td>T6 (A)</td><td colspan="2">D</td></tr>
<tr><td rowspan="3">T7 (B)</td><td>A (2nd syl. Shang or Qu</td><td>A' (second syl. is Shang or Qu) ***</td></tr>
<tr><td colspan="2">B (second syllable is Ru)</td></tr>
<tr><td colspan="2">D (second syllable is Ping) **</td></tr>
<tr><td rowspan="2">T8 (D)</td><td>A (2nd syl. Shang or Qu</td><td>A' (second syl. is Shang or Qu) ***</td></tr>
<tr><td colspan="2">D (second syllable is Ping or Ru)</td></tr>
</table>

* There are some variant forms with pattern A.
** A number of tri- and quadrisyllabic forms have shifted (optionally or obligatorily) to pattern A.
*** A number of tri- and quadrisyllabic forms have shifted (optionally or obligatorily) to pattern D.

Examples of the regular pattern substitution are given in (15) through (22) arranged in accordance with Table 2.

(15) T1 as the Initial Tone.

a. Pattern A.

T1 + T1:	sie sã	'teacher'
T1 + T8 + T1:	tsõ ɦɔʔ sɛ̃	'high school student'
T1 + T2 + T5 + T1:	sin mʌ sie sɛ	'new sweater'

b. Pattern A'.

T1 + T4 + T6:	sin mu lʌɯ	'new road'
T1 + T6 + T6 + T1:	kõ lʌɯ yən sɥ	'highway transportation'

(16) T2 as the Initial Tone - Pattern B.

T2 + T2:	no ŋ̥ə̃n	'a man'
T2 + T1 + T1:	no kʌ iə̃n	'male sophrano'
T2 + T6 + T8 + T3:	jã ŋ̥i zə phin	'foreign art works'

(17) T3 as the Initial Tone - Pattern C.

T3 + T3:	sɛi piʌ	'watch'
T3 + T2 + T3:	siʌ ɦu tsɿ	'small moustache'
T3 + T6 + T1 + T2:	sɥ li kõ zə̃n	'hydro-construction'

(18) T4 as the Initial Tone.

a. Pattern A.

T4 + T3:	lɛ̃ sɥ	'cold water'
T4 + T5:	vu tɕ̹hi	'weapons'

b. Pattern A'.

T4 + T5 + T1:	mu phi tsin	'flatterer'
T4 + T3 + T6 + T7:	lɛ̃ sɥ liʌ faʔ	'cold water therapy'

(18) c. Pattern D.

T4 + T2:	na dã̃n	'milk candy'
T4 + T8 + T2:	lʌ zəʔ ȵən	'honest person'
T4 + T1 + T4 + T5:	lʌ xu ŋɛ tɕiə̃n	'prebyopic glasses'

(19) T5 as the Initial Tone - Pattern B.

T5 + T5:	tshã phie	'(phonograph) records'
T5 + T2 + T5:	si mʌ sie	'fine wool'
T5 + T5 + T1 + T3:	fu pi sɥ tsã	'vice-secretary general'

(20) T6 as the Initial Tone - Pattern D.

T6 + T1:	ŋa tɕiʌ	'diplomatic'
T6 + T4 + T1:	dʌɯ ŋɛ tsin	'big eyes'
T6 + T7 + T8 + T1:	dʑiʌɯ tɕiaʔ daʔ tshʌɯ	'used bike'

(21) T7 as the Initial Tone.

a. Pattern A.

T7 + T3:	xəʔ pɛ	'blackboard'
T7 + T5:	khaʔ tɕhi	'polite'

b. Pattern A'.

T7 + T3 + T6:	tshiʔ tsi mɛ	'seven sisters'
T7 + T5 + T7:	kuəʔ tsi faʔ	'international law'
T7 + T6 + T1 + T2:	pɔʔ da si iã	'northern (part of) Atlantic Ocean'

c. Pattern B.

T7 + T8:	tɕhiəʔ liʔ	'to be a strain'
T7 + T7 + T8:	kuəʔ tɕiəʔ ɦəʔ	'bone tuberculosis'
T7 + T7 + T7 + T4:	xəʔ tshəʔ tshəʔ li	'dark'

(21) d. Pattern D.

T7 + T1:	kuəʔ tɕia	'country'
T7 + T2 + T6:	pɔʔ mə̃n ŋa	'outside the North Gate (of the city)'
T7 + T1 + T1 + T1:	kuəʔ tɕia tɕi kuɛ	'government agency'

(22) T8 as the Initial Tone.

a. Pattern A.

T8 + T3:	mɔʔ pɛ	'wooden board'
T8 + T5:	liʔ tɕhi	'strength'

b. Pattern A'.

T8 + T3 + T1:	ɦuəʔ xʌɯ sɛ	'active volcano'
T8 + T3 + T2 + T3:	ȵiəʔ sɥ bin tɛ	'thermos bottle lining'

c. Pattern D.

T8 + T2 + T2:	ɦuəʔ ȵie wɑ̃	'living King of Hell'
T8 + T2 + T5 + T1:	baʔ mʌ sie sɛ	'white sweater'

Having presented the pattern substitutions in Table 2, and provided some examples in (15) through (22), let us turn now to our proposal of an historical shift in Wuxi from last to first syllable dominance. The arguments and evidence in support of the proposal would also answer the two concerns expressed earlier, (1) why Wuxi would have undergone tone pattern substitution if the tone sandhi in the compounds and phrases in the dialect had always been first syllable dominant; and (2) what are the bases for our claim that the tone sandhi in these constructions originally involved last syllable dominance.

In proposing an historical shift from last to first syllable dominance, the clearest evidence comes from tone sandhi spans with initial T4 syllables. In such cases, tone pattern B, which T4 possesses in isolation, is replaced by one of three tone patterns -- A, A', or D -- depending on the original tone on the second syllable and on how many syllables there are in the span. Pattern B is replaced by pattern A or A' if the second syllable of disyllabic spans was Shang or Qu: pattern A occurs on disyllabic spans, and A' on longer ones. If the second syllable of the tone sandhi spans was Ping or Ru, then the pattern replacing B is D, regardless of syllable length. In order for the second syllable to condition the change in tone pattern, it must have retained its tone at the time of tone pattern substitution. The second syllable must, therefore, have served as the dominant syllable, since it determined the precise tone change on the first syllable. Only after pattern substitution is accomplished does the second syllable lose its citation tone. The new tone pattern on the first syllable is then entednded onto the entire tone sandhi span.[6]

Tone sandhi spans with initial T1 or initial T8 also display a split in tone patterning that is conditioned by the tone on the second syllable. For tri- and quadrisyllabic forms in particular, one observes a tendency towards the merging of the tone patterns. For example, T1 initial spans in which the second syllable originally bore Shang or Qu should display tone pattern A'. However, in some cases, these spans have A as an optional tone pattern. It is significant that the reverse does not hold: spans with tone pattern A do not have A' as an alternate pattern. The above facts suggest that, eventually, sandhi spans with initial T1 may have tone pattern A as the only substitute pattern, with the original tone on the second syllable irrelevant. Thus, it appears that the dialect is moving towards simplifying its tone pattern substitutions. Such a levelling process is affecting the tone sandhi spans with initial T7 and T8 syllables. Again, it is the trisyllabic and quadrisyllabic forms that are leading the way in this levelling process.

Related to the above comments is the observation that the regular pattern substitutions in Wuxi are not without exceptions, however, as noted in Table 2. Particularly with respect to trisyllabic and quadrisyllabic tone sandhi spans, there is a tendency toward eliminating the splits in the tone pattern substitutions conditioned by the original historical tone on the second syllable. Although not without exceptions, the disyllabic forms are more resistant to this levelling process. Consequently, it is the disyllabic sandhi forms that best preserve vestiges of tone substitution, or paradigmatic tone replacement, a process that is typical of sandhi patterns involving last syllable dominance.[5] Nonetheless, the trend toward context-free pattern substitution is obscuring the tonal conditioning, yielding, for modern Wuxi, a contradictory situation in which tone sandhi in the modern dialect entails pattern substitution on the first syllable despite an apparent case of first syllable dominance.

To recapitulate, we have seen evidence of tone pattern changes conditioned by the tone on the second syllable, suggesting that the compounds and phrases in Wuxi must have been last syllable dominant historically. It would be reasonable to expect that the second syllable of the sandhi spans had retained its citation tone at that time, as is typically the case for tone sandhi domains with last syllable dominance. The paradigmatic replacement of one tone with another in preterminal position of sandhi spans with last syllable dominance is a commonplace occurrence. This is obscured, however, in modern Wuxi by at least three factors: (1) the loss of the citation tone on the second syllable, effecting a shift in syllable dominance, (2) the subsequent spread of the tone pattern from the first syllable onto the entire sandhi span of two to four syllables in length, a tone-spreading process that is identical in manner to that found in tone sandhi spans with first syllable dominance in the dialect, and (3) the trend toward context-free pattern substitution that depends entirely on the tone category of the first syllable, due to the merging of tone patterns that had earlier reflected tonal conditioning of the second syllable.

In summary, it is precisely the vestiges of last syllable dominance that provides us with evidence for how Wuxi acquired its interesting pattern substitution. At the same time, Wuxi provides support for regarding the split dominance system in Tangxi as a more conservative system rather than a transitional one. Furthermore, because of the lack of data and analysis on Wuxi tone sandhi, we have included a fairly lengthy introduction on the topic. Instrumental data was also provided to support our treatment of Wuxi as having four basic tone patterns, derivable by either of the two strategies, pattern extension and pattern substitution.

Acknowledgements

The first author's research was made possible by funding from the Social Sciences and Humanities Research Council of Canada. We wish to thank Mona Lindau for suggesting the term, "pattern extension", which is preferable to our more ambiguous choice of "pattern duplication". Special thanks go to Ian Maddieson for his generous assistance in setting up the initial project in 1982. This paper, with minor changes, was presented at the Nineteenth International Conference on Sino-Tibetan Languages and Linguistics at Columbus, Ohio, 12-14 September 1986.

Notes

1. General published works on the Wuxi dialect include Chao (1928), Nanjing Normal Institute (1959), and Weng (1984, 1985). In addition, Ren (1982, 1983) contain the preliminary studies of the tone sandhi patterning in Wuxi.

2. Two native speakers from Wuxi City were used in this study: Jing Wang, a Visiting Scholar at UCLA whose speech was recorded in February 1982, and Hongmo Ren, a graduate student at UCLA. Early recording and instrumental work were done in 1982 by Hongmo Ren and Ian Maddieson. Additional recordings and acoustic data were obtained in the summer of 1986 using Ren as the informant. Ren in the speaker in all the spectrographic recordings in this paper.

3. Weng (1985) states that the Wuxi City dialect has eight tones, but does not provide an inventory of those tones. For both of our speakers, T2 and T6 have merged in citation context, emerging in sandhi domains with different tonal behaviour. (See also Weng (1984) where the author provides the tonal values for the eight tones in Bidian, a Wuxi dialect spoken in the eastern part of the district.

4. See Chan (1986), for example, where a number of cases are presented from different Chinese dialects where distinctions in historical tone categories are preserved in tone sandhi environment, although they have become lost in citation form. Such finds, thus, argue for the need to distinguish underlying forms from both citation and sandhi forms.

5. Of the pattern substitutions shown in Table 2, the most complex case involves T7 as the initial tone in the sandhi domain. Although the disyllabic forms are quite regular, revealing different tone patterns that are conditioned by the tone category of the second syllable, the same is not true in the case of tri- and quadrisyllabic forms: one sometimes finds alternate tone patterns, and at other times, the regular or the exceptional tone pattern is the obligatory pattern. As a result, the "regular" tone patterns for spans exceeding two syllables are based on the patterns for the disyllabic forms. The situation is then further complicated by the fact that forms which normally undergo tone extension undergo pattern substitution instead if the first syllable bears T7. Thus, tone sandhi spans with an initial T7 syllable behave exceptionally in at least a couple of ways. What is important for our purposes in knowing that,

despite the levelling process, there remains strong evidence that the pattern substitution is conditioned by the tone on the second syllable.

6. It may be that, at some stage, the second syllable supplied the last toneme in the tone pattern of disyllabic tone sandhi spans, which later became re-interpreted as part of the tri-toneme sequence in the pattern substitution process. Such a proposal would account, in particular, for the bidirectional tone pattern, LHL. In the paradigmatic tone replacement typical of the last syllable dominant type of tone sandhi spans, one frequently encounters level and falling tones, and, occasionally, rising tones as well. A convex tone surfacing in preterminal position, on the other hand, is rare, if it occurs at all. In Wuxi, if the shift from last to first syllable dominance involved an intermediary stage of tone neutralization on the second syllable, resulting in a default low tone on the second syllable, as one finds in the neighbouring Suzhou dialect, one would only need to posit an LH sequence on the initial syllable as the pitch contour resulting from the paradigmatic tonal substitution. Conceivably, then, the LHL sequence arose from the concatenation of LH from the first syllable, and L from the second syllable. This LHL sequence then became re-analyzed as a tone pattern that is mapped onto tone sandhi spans of three and four syllables in length.

References

Bodman, Nicholas. 1955. Spoken Amoy Hokkien. Kuala Lumpur: Government Federation of Malaya. 2 vols.

Chan, Marjorie K.M. 1985. Fuzhou phonology: a non-linear analysis of tone and stress. Ph.D. dissertation, University of Washington.

__________. 1986. "On the status of 'basic' tones." In this volume.

Chao, Yuen Ren. 1928. Xiandai Wuyude yanjiu (Study of modern Wu dialects). Beijing: Tsinghua Research Institute. Monograph 4.

Kennedy, George. 1953. "Two tone patterns in Tangsic." Language 29.3:367-373.

Nanjing Normal Institute. 1959. Wuxi ren Putonghua shouce. (Manual of Putonghua for Wuxi speakers). Nanjing, Jiangsu.

Ren, Hongmo. 1982. "A multi-level mapping tone system in Wuxi Chinese." U.C.L.A., ms.

__________. 1983. "A 'copying' type of tone sandhi in Wuxi Chinese." Paper included in: Proceedings of the Sixteenth International Conference on Sino-Tibetan Languages and Linguistics, edited by Marjorie K.M. Chan. Seattle: University of Washington.

Sherard, Michael. 1980. A synchronic phonology of modern colloquial Shanghai. Computational Analysis of Asian and African Languages, Monograph Series 5.

Wang, Futang. 1959. "Shaoxinghua jiyin." (Record of the sounds in Shaoxing speech) Yuyanxue Luncong 3:73-126.

Weng, Shouyuan. 1984. "Wuxi fangyan benzi kao." (Etymological notes on the Wuxi dialect) Fangyan 1984.3:199-212.

Weng, Shouyuan. 1985. "Wuxi fangyande guishu." (A note on the Wuxi dialect). Fangyan 1985.2:124-127.

Ye, Xiangling. 1979. "Suzhou fangyande liandu biandiao." (Tone sandhi in the Suzhou dialect) Fangyan 1979.1:30-46.

Yue-Hashimoto, Anne O. 1980. "Tone sandhi across Chinese dialects." Paper presented at the Thirteenth International Conference on Sino-Tibetan Languages and Linguistics, Charlottesville, Virginia.

Zee, Eric and Ian Maddieson. 1980. "Tones and tone sandhi in Shanghai: phonetic evidence and phonological analysis." Glossa 14.1:45-88.

Appendix. The Initials and Finals in Wuxi

1. Initials.

p	ph	b	m	f	v
t	th	d	n		l
ts	tsh			s	z
tɕ	tɕh	dʑ	ȵ	ɕ	j
k	kh	g	ŋ	x	ɦ

2. Finals.

ɿ	ʮ		u		y	m̩	ŋ̍	əl
a	iə	ua		ian	aʔ	iaʔ	uaʔ	yaʔ
ã		uã						
i	in							
o	io		uo			iɔʔ	uɔʔ	
õ	iõ							
ɛ			uɛ					
ɛ̃	iɛ̃		uɛ̃					
e	ie ([iɪ])		ue	ᴇi	əʔ	iəʔ	uəʔ	yəʔ
ʌ	iʌ							
ʌɯ	iʌɯ							
ə̃n	iə̃n		uə̃n	yə̃n				

On the Status of "Basic" Tones*

Marjorie K.M. Chan

0. Abstract.

This paper questions the traditional concept in Chinese dialectology of ben diao, "basic tones", which takes the "basic" tones to be coterminous with the citation tones, and proposes that the concept be revised based on the existence of an increase in tonal contrasts that arise in tone sandhi context in a number of Chinese dialects. The paper concludes with two examples from the Fuzhou dialect to demonstrate that tone sandhi positions may also preserve segmental contrasts that are totally or partially neutralized in citation context.

1. The Traditional Approach.

Until recently, what is considered a "basic" tone, or ben diao (root/original + tone) in Chinese dialectology has been quite straightforword and uncontroversial: it is a citation tone, a tone falling on a monosyllabic word. By contrast, a tone surfacing as a result of the juxtaposition of two or more tones is then a "derived" tone, or bian diao (change + tone).[1] This derived tone occurs in tone sandhi context and is referred to specifically as a "sandhi" tone. The most well-known example in Chinese of a so-called "basic" tone, as opposed to its sandhi counterpart(s) is that of Tone 3 in the four-tone system of Beijing Mandarin. In that dialect, Tone 3 has three phonetic shapes, as outlined in (1), where the numbers '1' to '5' represent ascending pitch height, based on Y.R. Chao's system of tone numbers.

(1) The Phonetic Shapes of Tone 3 in Mandarin.

a. [214] occurs prepausally (including in citation forms), and is a low falling-rising tone.

b. [35] occurs before another Tone 3 syllable, and is a mid-rising tone.

c. [21] occurs before any other tone, and is a low (falling) tone.

It is the phonetic shape of the citation tone in (1a) that is regarded as the basic one, and is hence the one taught first to foreign students in the study of the lexical tones in Mandarin. The other two tonal shapes of Tone 3 are considered the sandhi tones, assumed to be derived from the basic one, and are taught later.

A point concerning the "basic" form of Tone 3 needs to be made at this time. Woo, in her 1969 dissertation, proposed to treat Mandarin Tone 3 as low underlyingly, based on the sandhi form rather than the citation form, which is extra long in duration. Other generative phonologists have since adopted Woo's choice of an underlyingly low tone for Tone 3. Observe that distributionally it is the low variant (1c) which occurs in most environments: (1a) occurs prepausally, (1b) before another Tone 3, and (1c) elsewhere (namely, before Tones 1, 2, and 4, as well as before an atonic syllable). The low tone variant is, in fact, the choice of Hartman (1944), Hockett (1947), Kratochvil (1968), and other Structuralists, who treat Tone 3 as simply low, without any explicit statements as to their criteria for selecting low as the basic tone shape.

In China, the traditional analysis of the tones in the various Chinese dialects considers factors other than the distribution of the tonal variants to be of paramount importance. Rather than study the variants of each lexical tone, the focus is on the tonal system as a whole; that is, on the number of citation tones in relation to the number of tones in sandhi position in a dialect. Typically, a greater number of contrasts are found in the citation forms than in the sandhi forms. This can be illustrated using examples from Beijing Mandarin in (2); the Mandarin dialect of Luoyang in (3) from He (1984); the Northeastern Min dialect of Fuzhou in (4) from Chen and Norman (1965b); and the Southern Min dialect of Yongcun in (5) from Tay (1968), where the forms are given in broad phonetic transcriptions below. Aspiration is marked by an apostrophe ('). For ease of comparison, all tones are represented using tone numbers.) In all three examples, contrasts occurring on the citation forms are lost in (preterminal) sandhi position. Thus, in the Fuzhou example in (4), for instance, one assumes that the citation tones are basic, yielding the following analysis: tones /55/, /52/, and /12/ become high level before tone /52/; the result is the homophony of the disyllabic forms.

(2) Mandarin.

	Citation Tone		Sandhi Tone (on Preterminal Syllable)	
a.	mai^{214}	'to buy'	mai^{35} ma^{214}	'to buy a horse'
b.	mai^{35}	'to bury'	mai^{35} ma^{214}	'to bury horse'
	(ma^{214}	'horse')		

(3) Luoyang.

	Citation Tone		Sandhi Tone (on Preterminal Syllable)	
a.	iəu^{53}	'to have'	iəu^{31} ʂuei^{53}	'to have water'
b.	iəu^{31}	'oil'	iəu^{31} ʂuei^{53}	'oil and water'
	(ʂuei^{53}	'water')		

(4) Fuzhou.

	Citation Tone		Sandhi Tone (on Preterminal Syllable)	
a.	siŋ^{55}	'new'	siŋ^{55} iŋ^{52}	'bride'
b.	siŋ^{52}	'to become'	siŋ^{55} iŋ^{52}	'to become an adult'
c.	siŋ^{12}	'sacred'	siŋ^{55} iŋ^{52}	'a saint, a sage'
	(iŋ^{52}	'person, man')		

(5) Yongcun.

	Citation Tone		Sandhi Tone (on Preterminal Syllable)	
a.	kau^{24}	'monkey'	kau^{11} $p'ɔ^{24}$	'monkey's skin'
b.	kau^{11}	'thick'	kau^{11} $p'ɔ^{24}$	'thick skin'
	($p'ɔ^{24}$	'skin')		

By positing the citation tones as basic, or underlying, one can readily account for the different tones in the citation forms. The citation forms in (2) through (5) can be regarded as having preserved distinctions lost in the preterminal, sandhi forms: the reduction of tones in sandhi position is then the result of tone neutralization. It is examples such as those above that form the basis for positing the citation tones as basic, and the sandhi tones as having been derived from the citation ones. Furthermore, in dialects such Zhongshan and Standard Cantonese, where tone sandhi is either practically non-existent or of low-level phonetic relevance, the citation tones are chosen as the basic tones by default.

2. <u>Increase Tonal Contrasts in Sandhi Position</u>.

For dialects with tone sandhi, the simple rule-of-thumb in selecting the basic tones from the citation forms remains unchallenged as long as the sandhi position consistently shows a reduction in tonal contrast. As more studies on Chinese dialects with tone sandhi systems become available, however, one finds that, for quite a few dialects, the convention of treating the citation tones as the basic ones cannot be maintained. In this paper I will present a number of cases drawn from various Chinese dialects.

The first case to be examined is the Yongcun dialect of Southern Min (Tay 1968), a dialect which we have already encountered in (5) above. For comparative purposes, the tones are identified based on their historical tone categories. Roman numerals I, II, III, and IV represent Ping, Shang, Qu and Ru tones, while 'a' and 'b' designate the Yin and Yang registers respectively. Historically, the Yin-register syllables bore a voiceless initial consonant, whereas Yang-register syllables bore a voiced initial consonant. The tones in Yongcun are presented in (6) on the next page. Tone IV syllables in Yongcun end in -p, -t, -k, or -ʔ, and are hence shorter than the other, non-checked syllables. Due to their shorter duration, in (6) and elsewhere in this paper, checked syllables are represented by a single number if it is not a contour tone. (A high checked tone is '5', a mid checked tone is '3', etc.) Tones IIa and IIb are no longer contrastive, and have merged to become Tone II.

Given the merger of Tones IIa and IIb to form Tone II, [51], in the modern Yongcun dialect, the tonal value, [11], for both Tones IIIa and IIIb in citation tones would suggest another occurrence of tone merger. That is not the case, however. Although Tones IIIa and IIIb have merged in citation forms, they do not behave similarly in tone sandhi context. Tones IIIa and IIIb are contrastive in tone sandhi position, as illustrated in (7). (Tone numbers are assigned by me on the basis of the examples in (5), and the schematic representation given by Tay (1968:56)).

(6) Yongcun tone system (citation tones).

	I	II	III	IV
a	44	51	11*	32
b	24		11	24**
				5***

* Tone IIIa is phoneticaly [32] in certain tone sandhi contexts.
** in syllables checked by -p, -t, or -k
*** in syllables checked by -ʔ.

(7) Examples with Tones IIIa and IIIb in Yongcun.

	Citation Tone		Sandhi Tone (Preterminal Syllable)	
a. Tone IIIa:	si^{11}	'four'	si^{32} tsap^{24}	'forty'
c. Tone IIIb:	si^{11}	'to be'	si^{11} tsap^{24}	'it is ten'
	(tsap^{24}	'ten' - Tone IVb)		

In (7), observe that the citation forms are merged not only with respect to their tones, but also with respect to their segments. This is only possible because the obstruent initial of Yang ('b') register syllables have devoiced. The contrast between Tones IIIa and IIIb only surfaces in tone sandhi position. In the same dialect, there are also cases where two historical tone categories are neutralized in tone sandhi context. This was illustrated in (5) earlier, repeated in (5') below, with the historical tone categories included for comparative purposes. Tones Ib and IIIb are contrastive in citation forms, but neutralized in preterminal, tone sandhi context.

(5') Examples with Tones Ib and IIIb in Yongcun.

	Citation Tone		Sandhi Tone (Preterminal Syllable)	
a. Tone Ib:	kau^{24}	'monkey'	kau^{11} p'ɔ^{24}	'monkey's skin'
c. Tone IIIb:	kau^{11}	'thick'	kau^{11} p'ɔ^{24}	'thick skin'
	(p'ɔ^{24}	'skin' - Tone IVb)		

The Yongcun case is valuable in showing that tone sandhi positions are not necessarily always the sites where tone neutralizations occur. Tonal contrasts may emerge in tone sandhi context or in citation context. In the case of Tones IIIa and IIIb, the tonal contrasts only emerge in sandhi context, whereas in the case of Tones Ib and IIIb, it is in the citation context that the tonal contrast occurs. From an historical perspective, what is significant about Yongcun is that the augmentation of tones in sandhi position reflects preservation of earlier tonal distinctions that are now lost in the citation forms. It would thus be a fallacy to consider the citation tones to be the "basic" ones, and to further regard them as the earlier tones historically, as might be implied by the term, ben diao (original + tone).[2]

Despite the above comments, it is still the case that examples of tone neutralization in sandhi context predominate in the literature. Thus, an increase in tonal distinctions in that environment is rare, although by no means unique to Yongcun. Similar types of examples are found in other Chinese dialects, such as Ningde, a Northeastern Min dialect (Norman 1977); Pingyao, a Northwestern Mandarin dialect (Hou 1980); Yinchuan, a northern Mandarin dialect spoken in the Ningxia Hui Autonomous Region, immediately west of the Inner Mongolia Autonomous Region (Zhang 1984); and Wenling, a Wu dialect (Li 1978, 1979). (For additional cases in the Mandarin and Wu dialects, see Hashimoto 1982 and Pan 1986).

Examples from these dialects are given below. Further tonal changes on the [kæ] syllable in (9a) may be ignored for our purposes. In (8) through (10) below, the initial obstruents of Yang ('b') register syllables have become voiceless in the modern dialects, thereby paving the way for total neutralization of different tonal categories in citation context. The Wenling case in (11) involves the merger of Tones Ib and IIb, both in the Yang register. It is, therefore, irrelevant in this dialect whether the devoicing of obstruents has taken place or not. Nonetheless, Wenling is typical of the Wu dialects in preserving the voicing on the obstruent series.

(8) Examples with Tones IIa and IIIb in Ningde.

	Citation Tone		Sandhi Tone (Preterminal Syllable)	
a. Tone IIa:	hɔ^{41}	'good'	hɔ^{35} ma^{41}	'good horse'
b. Tone IIIb:	hɔ^{41}	'number'	hɔ^{33} ma^{41}	'number (e.g., phone)'

(ma^{41} 'horse' - Tone IIa)
(ma^{41} 'sign for number' - Tone IIa)

(9) Examples with Tones Ia and Ib in Pingyao.

	Citation Tone		Sandhi Tone (Preterminal Syllable)	
a. Tone Ia:	təŋ^{13}	'to step on'	təŋ^{31} k'æ^{35}	'to begin to step on'
b. Tone Ib:	təŋ^{13}	'to mount'	təŋ^{13} k'æ^{13}	'to begin to mount'

(k'æ^{13} 'to begin' - Tone Ia)

(10) Examples with Tones Ib and IIa in Yinchuan.

	Citation Tone		Sandhi Tone (Preterminal Syllable)	
a. Tone Ib:	faŋ^{53}	'to guard against'	faŋ^{53} tʂʅ^{14}	'prevention and cure'
b. Tone IIa:	faŋ^{53}	'to spin'	fan^{35} tʂʅ^{14}	'spinning and weaving'

(tʂʅ^{14} 'to cure' - Tone III < *Tone IIIb)
(tʂʅ^{14} 'to weave' - Tone III < *Tone IVa)

(11) Examples with Tones Ib and IIb in Wenling.

	Citation Tone		Sandhi Tone (Preterminal Syllable)	
a. Tone Ib:	bi^{31}	'skin'	bi^{13} ʔli^{31}	'in the skin'
b. Tone IIb:	bi^{31}	'quilt'	bi^{31} ʔli^{31}	'in the quilt'

(ʔli^{31} 'in' - Tone IIb)

It is worth noting that Tones Ib and IIb in Wenling are distinguished not only with respect to their sandhi tones, but also with respect to their "changed tones", or bian yin, as first reported in Li (1966), and further exemplified in Li (1978:96). Wenling has a morphologically-conditioned tone change the function of which is similar to the er-suffixation in Beijing Mandarin and the tone changes in Standard Cantonese. What is of interest here is that the changed tone has two alternant forms, [15] and [51]. Despite the same tone, [31], in citation forms, Tones Ib and IIb do not bear the same variant when they undergo morphological tone change. As shown in (12), Tone Ib surfaces with the rising alternant of the changed tone, and Tone IIb the falling one.[3]

(12) Examples with Tones Ib and IIb in Wenling
(in the context of morphological tone change).

	Citation Tone		Changed Tone (Terminal Syllable)	
a. Tone Ib:	bi^{31}	'skin'	bo^{21} bi^{15}	'thin skin'
b. Tone IIb:	bi^{31}	'quilt'	bo^{21} bi^{51}	'thin quilt'

(bo^{21} 'thin' - Tone IVb)

Thus, in Wenling, despite the neutralization of Tones Ib and IIb in citation context, both tone sandhi and morphological tone change provide evidence that one must recognize a contrast between Tones Ib and IIb in the tonal system of the dialect. The distinction, although a historical one initially, continues to be relevant to the modern dialect. Given the different conditions under which the tonal contrast emerges, the Wenling case is of particular value for demonstrating the need to examine other contexts besides that of the citation one to determine the so-called "basic" tones in a given dialect. Wenling is thus far unique in having a changed tone serving to contrast two historical tone categories that have merged in citation context.

A final case to be cited here concerns Tones IIb and IIIb in Suzhou, a Wu dialect. The case is presented in Ye (1984), involving a more restricted occurrence of tonal contrast surfacing in sandhi context. Some preliminary comments are needed at this point. The tonal system of Suzhou is presented in (13), based on the citation tones. As one can see, Tones IIb and IIIb are merged as [31] in their citation tone.

(13) Suzhou tone system (citation tones).

	I	II	III	IV
a	55	51	513	5
b	13	31	31*	3

* Tone IIIb is phonetically [313] when it is preceded by Tone IVa.

In the dialects that we have studied thus far, tone sandhi changes occur on the preterminal syllable. The situation is different for Suzhou, as is generally true for the Wu dialect group. Simplifying the picture somewhat, in the disyllabic tone sandhi patterning in Suzhou, the terminal syllable undergoes tone sandhi changes, whereas the initial syllable bears the citation tone.[4] There is, nonetheless, an exception to that general observation: a syllable following a Tone IVa syllable normally retains its citation tone rather than undergo sandhi changes. As a result, in a disyllabic form, one would expect Tones IIb and IIIb to remain non-contrastive following a Tone IVa syllable; that is, both Tones IIb and IIIb should bear the mid-falling [31] citation tone. Interestingly enough, that is not the case. Tones IIb and IIIb have different pitch contours in that environment: while Tone IIb remains mid-falling, Tone IIIb emerges with a dipping [313] tone. Thus, Tone IIIb has the form of [31] in monosyllables, and [313] following Tone IVa.[5] A set of examples contrasting Tones IIb and IIIb is given in (14).

(14) Examples with Tone IIb and IIIb in Suzhou.

	Monosyllabic Word		Disyllabic Word	
a. Tone IIb:	bu^{31}	'section'	$iə^{25}$ bu^{31}	'one section'
b. Tone IIIb:	bu^{31}	'step'	$iə^{25}$ bu^{313}	'one step'
	($iə^{25}$	'one' - Tone IVa)		

Ye (1984) considers the [313] form to be the original tone value of Tone IIIb. It is the [313] tone that has preserved the distinction between Tones IIb and IIIb, a distinction which is lost in monosyllabic forms. In addition, he notes that the [313] pitch value of Tone IIIb corrresponds well to the Tone IIIa counterpart, with a [513] tone value. Tones IIIa and IIIb share one feature in common: [313] of Tone IIIa and [513] of Tone IIIb are both dipping tones, with a falling-rising pitch contour. The two tones also yield the expected initial pitch contrast, as a result of the difference in their syllable onset. Tone IIIa syllables have a voiceless initial, and hence bear a higher initial pitch. Tone IIIb syllables have a voiced onset, and thus surface with the expected lower pitch initially. It is worth adding that parallel situations obtain for Tones IIa and IIb on the one hand (Tone IIa is <u>high</u>-falling [51], while Tone IIb is <u>mid</u>-falling [31]), and Tones IVa and IVb on the other (Tone IVa is a <u>high</u> [5] tone, and Tone IVb a <u>mid</u> [3] tone).

To summarize the Suzhou case, the [313] variant of Tone IIIb provides crucial evidence that a distinction between Tones IIb and IIIb is preserved in the dialect, albeit in an extremely restricted environment.

3. <u>Positing the Underlying Tones</u>.

The examples of increased tonal contrasts arising in tone sandhi environment demonstrate the need for a three-way contrast to distinguish the set of "basic" tones from the set of sandhi tones as well as from the set of citation tones. It should be emphasized at this point that in neither the generative nor Structuralist treatment of tones is it requisite for the citation tones to serve as the basic ones, as we have already noted concerning Mandarin Tone 3. Being able to select either a citation or sandhi form for the underlying tonal shape is crucial in generative studies, since the choice determines the formulation of rules to account for the surface forms. Those working in the generative approach have, thus far, chosen the citation tone as the underlying form unless there are compelling reasons to select an alternate form. Reasons may include the desire to obtain a simpler or more elegant solution, to gain greater predictive power, or to better account for certain timing phenomena. Regardless of the theoretical approach, the cases presented in this paper provide unequivocal evidence against assuming a one-to-one relationship between "basic tones" and the citation ones. It is necessary to allow the sandhi forms to be selected as part of the basic set of tones if the sandhi forms are needed to account for surface contrasts. However, if the term, "basic tones", is used in this broader sense, there is the possibility of introducing terminological confusion. Instead, we will hereafter adopt the already existing generative term, "underlying tones", for this purpose. The set of underlying tones then chooses from the citation and the sandhi form.[6]

In a dialect such as Yongcun, Tones IIIa and IIIb would not be distinguished if only the citation forms of these tones are chosen to form the underlying tones in the dialect. The different behaviour of these two tones cannot be accounted for if /11/ is assigned as the underlying form for both tones. An easy remedy would to be select /32/ as the underlying form of Tone IIIa in the dialect. The neutralization of Tones IIIa and IIIb then occurs in citation forms, or more generally, at word-final position. Thus, the underlying tones in Yongcun can be posited as follows, in (15b), with the citation tones given in (15a) for ease of comparison:

(15) a. The Citation Tones in Yongcun.

```
  :   I    :   II   :   III    :  IV
--:--------:--------:----------:-------
a :   44   :        :   11*    :  32
--:--------:        :----------:-------
  :        :   51   :          :  24**
b :   24   :        :   11     :-------
  :        :        :          :   5***
--.--------.--------.----------.-------
```

* Tone IIIa is phonetically [32] in certain tone sandhi contexts.
** in syllables checked by -p, -t, or -k
*** in syllables checked by -ʔ.

b. The Underlying Tones in Yongcun.

```
  :   I    :   II   :   III    :  IV
--:--------:--------:----------:-------
a :   44   :        :   32*    :  32
--:--------:        :----------:-------
  :        :   51   :          :  24**
b :   24   :        :   11     :-------
  :        :        :          :   5***
--.--------.--------.----------.-------
```

* The sandhi form of Tone IIIa is chosen as underlying.
** In syllables checked by -p, -t, or -k.
*** In syllables checked by -ʔ.

Recall that Tone IV syllables are checked, and remain contrastive with respect to Tone IIIa syllables, which are non-checked, ending in a vowel, glide, or nasal consonant. Note also that, historically, one would have expected a higher pitch register on a Yin ('a') register syllable. The surfacing of /32/ on Tone IIIa, contrasting with /11/ on Tone IIIb, therefore, suggests that the sandhi position not only preserved a tonal contrast, but also the expected difference in pitch height between Yin and Yang registers. For comparative purposes, it is instructive to cite Standard Cantonese, where Tone IIIa is /44/, while Tone IIIb is /33/. Again, we find an overall higher pitch on a Yin ('a') register tone than on a Yang ('b') register tone. A similar situation also obtains in Suzhou, as demonstrated earlier.

A solution similar to Yongcun can be motivated for Yinchuan, Ningde, Pingyao, and Suzhou, as illustrated in (16) through (19), where the citation tones versus the set of underlying tones are both presented for comparative purposes. Two dashed lines "--" in the slot of a tonal cateogry indicates that syllables bearing that historical tone category have shifted to other tone categories, as in the case of Tone IIb in Yinchuan, where Tone IIb syllables with sonorant initials have shifted to the Tone IIa category, while the remaining syllables from the Tone IIb category, with historically voiced obstruent initials, have shifted to the Tone III category. In the meantime, the Tone III category in Yinchuan represents the merger of two earlier tone categories, Tone IIIa and IIIb. That is, there is one single tonal reflex for these two historical categories. What were historically Tone IV syllables are no longer checked in Yinchuan, and have hence shifted to other (non-checked) tone categories.

A few comments will be made concerning the above dialects. In the Yinchuan case (16), observe that Tone IIa is treated as underlyingly /35/. In the literature on tonogenesis in Chinese, it is proposed that syllables bearing Tone II, the Shang tone, ended in a glottal stop historically. There is abundant published sources (cited in Chan 1977, 1985, for example) on the pitch-raising effect of a final glottal stop. What is worth noting here is that the sandhi form in Pingyao has preserved the rising tone.[7] Similar cases of a sandhi form preserving the rising tone can be found in other dialects, including Ningde, even though Tone IIa is treated in Ningde as having /55/ as the underlying tone. /55/ is chosen due to the need to distinguish Tone IIa (/55/) from Tone IIIa (/35/).[8] In the case of Pingyao, notice, again, that the Yin-register tone, Tone Ia, preserves the higher register, with /31/ underlyingly, contrasting with Tone Ib, with /13/ underlyingly.

(16) a. The Citation Tones in Yinchuan.

<table>
<tr><th></th><th>I</th><th>II</th><th>III</th><th>IV</th></tr>
<tr><td>a</td><td>44</td><td>53*</td><td rowspan="2">14</td><td>--</td></tr>
<tr><td>b</td><td>53</td><td>--</td><td>--</td></tr>
</table>

* Tone IIa has the sandhi tone [35] before Tone III, and before an atonic syllable.

b. The Underlying Tones in Yinchuan.

<table>
<tr><th></th><th>I</th><th>II</th><th>III</th><th>IV</th></tr>
<tr><td>a</td><td>44</td><td>35*</td><td rowspan="2">14</td><td>--</td></tr>
<tr><td>b</td><td>53</td><td>--</td><td>--</td></tr>
</table>

* The sandhi tone of Tone IIa is chosen as underlying.

(17) a. The Citation Tones in Ningde.

	I	II	III	IV
a	33	41*	35	3
b	11	--	41**	5

* Tone IIa has two sandhi forms: [35] before [41], [55] elsewhere.
** Tone IIIb has [33] as a sandhi tone occurring before any tone.

(17) b. The Underlying Tones in Ningde.

	I	II	III	IV
a	33	55*	35	3
b	11	--	41	5

* One of the sandhi tones of Tone IIa is chosen as underlying.

(18) a. The Citation Tones in Pingyao.

	I	II	III	IV
a	13*	53	35	23
b	13			54

* Tone Ia has [31] as a sandhi tone.

b. The Underlying Tones in Pingyao.

	I	II	III	IV
a	31*	53	35	23
b	13			54

* The sandhi tone of Tone Ia is chosen as underlying.

(19) a. The Citation Tones in Suzhou.

	I	II	III	IV
a	55	51	513	5
b	13	31	31*	3

* Tone IIIb has [313] as a tonal variant following Tone IVa.

b. The Underlying Tones in Suzhou.

	I	II	III	IV
a	55	51	513	5
b	13	31	313*	3

* The tonal variant of Tone IIIb is chosen as underlying.

The Suzhou case presents an added point of interest. The citation form of Tone IIIb is mid-falling [31], having merged with Tone IIb. The underlying form of Tone IIIb is /313/, based on the tonal value occurring after a Tone IVa syllable. Note that the underlying form cannot legitimately be referred to as a "sandhi tone", since in every other instance the syllable following a Tone IVa syllable retains its citation tone. Ye (1984), in fact, regards that environment as a very stable one, and uses tone stability following Tone IVa to advance the proposal that the [313] variant is the original pitch value of Tone IIIb. Thus, the [313] variant might be more appropriately regarded as an alternant citation tone than as a sandhi tone. As far as tones in tone sandhi environment is concerned -- that is, tones in non-initial position -- there is a great deal of neutralization (Ye 1979). For Tone IIIb and other tones, the regular sandhi tone for syllables in a disyllabic tone sandhi span is simply a low-falling [21] tone. Thus, we can establish for Suzhou a further confirmation of a need for a three-way distinction among (1) underlying tones, (2) citation tones, and (3) sandhi tones. For Suzhou Tone IIIb, the underlying form, based on the [313] tonal variant, differs from both its citation form, [31], and its sandhi form, [21].

This section will end with two more cases, the Northern Mandarin dialect of Baoding (Yang 1960), and the Wu dialect of Wenling (Li 1979). For both dialects, complications arise leading to some indeterminacy in positing the underlying tones. Turning first to the Baoding dialect, the citation tones are presented in (20a). Some details concerning Tone IIb are omitted to simplify our discussion.

(20) a. The Citation Tones in Baoding.

	I	II	III	IV
a	45	213	51	--
b	22	213*	51**	--

* Tone IIa has [51] as its sandhi tone, while Tone IIb has [11] as its sandhi tone.

** Tone IIIa has [55] as its sandhi tone, while Tone IIIb has [11] as its sandhi tone.

It can be seen in (20a) that Tones IIa and IIb are merged in citation context, but remain contrastive in tone sandhi environment. Similarly, Tones IIIa and IIIb are no longer distinctive in citation forms, but continue to contrast in tone sandhi position. Observe also that the sandhi forms of Tones IIb and IIIb have merged. As a result, Tone IIb has merged with Tone IIa in citation context but with Tone IIIb in tone sandhi context, and is, thus, aptly dubbed a "ghost tone" by Hashimoto (1982). Given the situation in Baoding, there is no unique set of underlying tones that can be posited. Two possibilities are given in (20b) and (20c). The ultimate choice must depend on a more comprehensive study of the dialect. What is important for our purposes is the surfacing of distinctions in tone sandhi environment despite their loss in citation context.

(20) b. The Underlying Tones in Baoding. (Option 1)

	I	II	III	IV
a	45	213	55	--
b	22	11	51	--

c. The Underlying Tones in Baoding. (Option 2)

	I	II	III	IV
a	45	51	55	--
b	22	213	11	--

Turning now to Wenling, the citation tones are given below from Li (1979).

(21) a. The Citation Tones in Wenling.

	I	II	III	IV
a	33	42	55	5
b	31*	31**	13	1

* Tone Ib has [13], [24], and [35] as sandhi forms in various environments. (In terminal position, Tone Ib sometimes also undergoes tone change to [51].)

** Tone IIb has [35] as a sandhi tone before Tone Ib.

In (21a), observe that new tonal values arise in (preterminal) tone sandhi position for both Tones Ib and IIb. As a result, there are several, competing sets of tones that can serve as the underlying tones in the dialect. Three of these are presented below.

(21) b. The Underlying Tones in Wenling. (Option 1)

	I	II	III	IV
a	33	42	55	5
b	24*	31	13	1

* A sandhi form of Tone Ib is chosen as underlying.

(21) c. The Underlying Tones in Wenling. (Option 2)

	I	II	III	IV
a	33	42	55	5
b	35*	31	13	1

* A sandhi form of Tone Ib is chosen as underlying.

d. The Underlying Tones in Wenling. (Option 3)

	I	II	III	IV
a	33	42	55	5
b	31	35*	13	1

* The sandhi form of Tone IIb is chosen as underlying.

One might mention that Ting (1983) posits our Option 2 (21c) as the set of "basic" tones for Li's Wenling data. However, he provides no explicit arguments favouring this choice over other options that seem equally viable. Particularly for a synchronic analysis, it would appear that only a more thorough study of Wenling that takes into consideration the overall structure of the tone system in the dialect would provide arguments for selecting one choice over the others. That is a direction of research that will not be pursued here. As in the Baoding case, our main concern is the need to distinguish Tones Ib and IIb underlyingly, which can be accomplished by any of the above options. To complicate matters, there is another variety of the Wenling dialect, recorded by Susan Hess (personal communication), which has the same pitch values for the citation tones, but her dialect material differs crucially from Li's in that it lacks the two rising sandhi tones, [24] and [35]. The Wenling tone system recorded by Hess is given in (21e).

(21) e. The Citation Tones in Wenling - Variety 2.

	I	II	III	IV
a	33	42	55	5
b	31*	31**	13	2

* Tone Ib has two sandhi forms: [44] before Tone Ia, and [13] elsewhere.
** Tone IIb has [55] as a sandhi form before Tone Ib.

Instead of the sandhi forms for Tone Ib given in (21a), based on Li's data, the sandhi forms for that tone are [44] and [13], as indicated above. Similarly, instead of the [35] tone surfacing in sandhi position for Tone IIb, Hess' data show a [55] sandhi tone. Thus, none of the options presented earlier in (21b) through (21d) are relevant for Hess' Wenling data. Whether one opts for a solution requiring four phonological tone levels, or a more abstract analysis that by-passes the tone numbers and utilizes, instead, the auto-segmental framework (as done for Fuzhou (Chan 1985) and Ningde, in footnote 8), needs to be explored further. For the present, a tentative proposal for the underlying tones for the second variety of Wenling is presented in (21f).

(21) f. The Underlying Tones in Wenling - Variety 2.

	I	II	III	IV
a	33	42	55	5
b	44*	31	13	2

* One of the sandhi tones of Tone Ib is chosen as underlying.

4. Distinctions Found in Tone Sandhi Environment: Cases from Segmental Phonology.

Tone sandhi position is conventionally regarded as the location where neutralizations take place, because such phenomena typically occur in that position. However, there is no a priori reason to expect that only neutralizations can occur in that context. In the above sections, we have demonstrated that tone sandhi position may actually preserve distinctions in tonal categories that have since been lost in citation forms. Parallel cases involving segments may be cited as further evidence that tone sandhi position may serve at times to preserve certain historical distinctions.

In this section, two cases involving the Fuzhou dialect will be presented from Chan (1985). The first case pertains to the final glottal stop, which came from two earlier codas, *-ʔ and *-k.[9] The case will be studied with respect both to its behaviour in tone sandhi and its effect on the following consonant. The second case involves an /a/ final (syllable minus the initial consonant) occurring with all tones and in all contexts, contrasting with an /e/ final which has two alternants: [a], which occurs only in citation forms bearing Tone IIIa or IIIb, and [e], surfacing in the remaining contexts. Historically, words with the invariant, or non-alternating, /a/ final belong to the Jia rhyme group, while words with the /e/ final, consisting of alternants [e] and [a], belong to the Xie rhyme group.

Before we discuss the first case, some preliminaries are in order concerning the tones in Fuzhou. The citation tones are presented in (22) from Chen and Norman (1965a). All the Fuzhou examples that follow are from Chen and Norman (1965b), and Norman (1971), presented here in very broad phonetic transcription. (The mid vowels are actually more open than transcribed, and [z] is lenis.)

(22) The Citation Tones in Fuzhou.

	I	II	III	IV
a	55	22	12	24
b	52	--	342	5

Regarding the first case, involving the final glottal stop, all Tone IVa and IVb syllables are closed by the glottal stop, which, as noted above, is descended from two earlier consonants, *-ʔ and *-k. In (preterminal) tone sandhi position, the glottal stop behaves differently depending on whether it came from earlier *-ʔ or *-k. Irrespective of speech rate, the glottal stop from earlier *-ʔ is lost when it occurs in sandhi forms. In the case of the glottal stop from earlier *-k, however, it is retained in slow, careful speech. Only in faster speech is it deleted. Thus, there is a differential behaviour between the two sources of the modern Fuzhou glottal stop, as shown in the near minimal sets in (23) and (24) below, where the slower speech rate is demonstrated. The disyllabic compound in each set is composed of the first syllable, A, and the second syllable, B, together with tonal and segmental changes. (The B forms are non-checked and do not contain a glottal stop.)

(23) Examples of Fuzhou words with glottal stop coda from *-ʔ and *-k: Set 1.

Coda on A	A	+	B	AB Compound
a. *-ʔ	pa^{ʔ5}	+	tøyng^{52}	pa^{22} røyng^{52}
	'white'		'copper'	'white brass'
b. *-k	ha^{ʔ5}	+	tøyng^{52} 'same'	ha^{ʔ52} tøyng^{52}
	'join'		'same'	'a contract'

(24) Examples of Fuzhou words with glottal stop coda from *-ʔ and *-k: Set 2.

Coda on A	A	+	B	AB Compound
a. *-ʔ	pa^{ʔ5}	+	ts'ai^{12}	pa^{22} zai^{52}
	'white'		'vegetable'	'cabbage'
b. *-k	la^{ʔ5}	+	ts'ai^{12}	la^{ʔ2} ts'ai^{12}
	'hot, peppery'		'vegetable'	'preserved vegetable'

In the compounds in (23a) and (24a), the glottal stop is lost in tone sandhi context, whereas it is retained in the same environment in (23b) and (24b), where the glottal stop came from earlier *-k. Observe, also, in (23) and (24) that the presence or absence of the glottal stop in tone sandhi environment also determines whether or not the initial consonant in the terminal syllable undergoes consonant lenition. Within the same tone sandhi span, consonant lenition does not occur when the syllable is preceded by the glottal stop (<*-k), as demonstrated in (23b) and (24b). Consonant lenition does occur in the case of (23a) and (24a) where -ʔ (< *-ʔ) in the preceding syllable is deleted.[10]

The second case from Fuzhou involves a well-known observation about the dialect, namely that it has vowel alternations which are tonally conditioned. The discussion is restricted here to the case of the non-alternating /a/ final, which is descended from the Jia rhyme group, contrasting with the alternating /e/ final, which is descended from the Xie rhyme group. The phonetic variants of these two finals are presented in (25) below.

(25) The /a/ and /e/ finals in Fuzhou.

a. The /a/ final is [a] in all contexts. (From the Jia rhyme group).

b. The /e/ final is [a] only in citation syllables bearing Tone IIIa or IIIb, and is [e] everywhere else (viz., in citation syllables bearing Tone Ia, Ib, or IIa, and in all tone sandhi positions). (From the Xie rhyme group).

Examples with the non-alternating /a/ final versus the alternating /e/ final are given in (26) and (27). (26) involves Tones Ia, Ib, and IIa syllables, while (27) involves Tone IIIa and IIIb syllables. Since these finals are not checked by a final glottal stop, they do not occur in Tones IVa and IVb.

(26) Tones Ia, Ib, and IIa Syllables.

a. With the /a/ Final.

In (Preterminal) Sandhi Position			In Citation Forms	
sa^{52}	mau^{224}	'desert'	sa^{55}	'sand'
(sand + desert)				
pa^{22}	pa^{52}	'rake'	pa^{52}	'rake'
(rake + rake)				
ma^{22}	tso^{52}	'manger'	ma^{22}	'horse'
(horse + trough)				

b. With the /e/ Final.

In (Preterminal) Sandhi Position			In Citation Forms	
se^{52}	βoy^{224}	'northwest'	se^{55}	'west'
(west + north)				
pe^{22}	pe^{52}	'a sign'	pe^{52}	'sign'
(sign + sign)				
me^{35}	zyo^{22}	'buyer'	me^{22}	'to buy'
(buy + '-er')				

(27) Tone IIIa and IIIb Syllables.

a. With the /a/ Final.

In (Preterminal) Sandhi Position			In Citation Forms	
tsa^{55}	riang^{55}	'to hear	tsa^{12}	'suddenly'[12]
(suddenly + hear)		suddenly'		
ha^{55}	rieng^{55}	'summer'	ha^{342}	'summer'
(summer + season)				

b. With the /e/ Final.

In (Preterminal) Sandhi Position			In Citation Forms	
tse^{52}	puo^{342}	'linen cloth'	tsa^{12}	'linen'
(linen + cloth)				
he^{55}	uong^{52}	'the spawn of	ha^{342}	'crab'
(crab + yellow)		crabs'		

As one can observe, in tone sandhi context the two finals, /a/ and /e/, remain contrastive throughout (as presented in (26) and (27)). In citation forms, they are contrastive if they bear Tone Ia, Ib, or IIa, as shown in (26). When the citation syllables occur in Tone IIIa or IIIb, however, as exemplified in (27), the two finals are neutralized, with both finals yielding [a]. Thus, in this case involving finals /a/ and /e/, neutralization of segments affects only a subset of citation forms, and none of the forms in tone sandhi context.

To summarize, both cases cited above from the segmental phonology of Fuzhou provide evidence that tone sandhi position may yield phonological contrasts that are neutralized, or partially neutralized, in citation context. From the perspective of historical Chinese phonology, these findings are significant in that tone sandhi positions can be relevant for seeking vestiges of earlier distinctions. Cases such as what we find in Fuzhou are rare among the Chinese dialects, but they should not be unique to Fuzhou and some of its neighbours. One would expect parallel cases in other dialects in which segmental contrasts are preserved in tone sandhi environment though lost in citation context.

5. Conclusion.

Despite the traditional view that the citation tones are the "basic" or "original" tones, various works (e.g., Hashimoto 1982, Ting (1982,1983), Chan 1985, and Pan 1986) present cases demonstrating that the tones in tone sandhi position may at times be the historically earlier forms in a given dialect, and would hence be the forms useful for the reconstruction of its proto-tones. The simplistic notion of tone sandhi environments as merely sites for phonological neutralization cannot be upheld in such analyses, since neutralized tones would be virtually useless in reconstruction attempts. The cases of final glottal stop and the /a/ and /e/ finals in Fuzhou further argue against the naive view of sandhi positions being strictly locations for phonological neutralizations.

In this paper, the traditional position is challenged from the perspective of the synchronic analyses of the Chinese dialects. Several dialects have been selected that show an increase in tonal contrasts in sandhi position. The cases presented here are by no means exhaustive. Hashimoto (1982), for example, mentions similar phenomena in the literature on a number of Mandarin dialects. Increases in tonal contrast in sandhi context provide the crucial evidence for requiring at least some basic, or underlying, tones to be selected from the sandhi ones. As a result, in such dialects as Ningde, Pingyao, and other cases cited here, the so-called "basic" tones can no longer be identified on a one-to-one basis with the citation tones.

In discussing diachronic and synchronic concerns above, it should be emphasized that the tonal variant chosen as the underlying form in a synchronic analysis may not correspond to the variant that is analyzed as being the oldest one historically for a given tone in a dialect. As noted for Ningde, Tone IIa is treated as underlyingly high, /55/, even though the tone was historically rising, so that the rising tonal variant, [35], is actually the one that has best preserved the pitch contour of the tone. However, further developments in the tonal system of Ningde -- namely the emergence of Tone IIIa as a rising [35] tone, due perhaps to a kind of pull-chain phenomenon -- prevent Tone IIa from being treated as /35/ underlyingly in the modern dialect (unless a more abstract solution is sought, as in footnote 8).

Similarly, although Ting (1982) considers the [35] variant of Mandarin Tone 3 (historically Tone IIa) to be the most conservative for reconstruction purposes, in a synchronic analysis, Tone 3 cannot be analyzed as /35/ underlyingly, since Tone 2 (historically Tone Ib) subsequently became a rising [35] tone that must be treated as /35/ in the modern dialect. For our purposes, the proposed three-way distinction among "basic" (underlying) tones, citation tones, and sandhi tones is intended for synchronic analyses only. The case of Tone IIIb in Suzhou also deserves to be mentioned here, since the underlying form, /313/, is chosen from neither the citation tone nor the sandhi tone. The example is valuable for demonstrating the need to study all the environments before positing the set of underlying tones in a dialect.

Notes

* Research for this paper is funded in part by the Social Sciences and Humanities Research Council of Canada. Thanks also go to Susan Hess for bringing to my attention Ting Pang-hsin's articles. This paper was originally presented at the Chinese Linguistics Workshop at UCLA in May, 1986.

1. There is another type of derived tone which arises through morphologically-conditioned tone changes, referred to in the literature as bian yin (change + sound), or "changed tone". Again, the citation tone serves as the base. Examples are cited below from the Yue dialects of Taishan (Yiu 1946) and Zhongshan (Chan 1980):

(i) Taishan.

ngwoi^{33}	'I'	(1p. sg.)
ngwoi^{22}	'we'	(1p. pl.)

(ii) Zhongshan.

hy^{22}	'go'	
hy^{35}	'gone'	(perfective)

The tone change to /22/ in Taishan to mark the first person plural pronoun is non-productive, whereas the tone change to a new tone, /35/, in Zhongshan to mark the perfective aspect is. In Standard Cantonese, the /35/ changed tone not only functions to mark perfectivity, but it also serves to mark a noun as denoting familiarity, diminution, and so forth. For further details on changed tones in Standard Cantonese, see, for example, Hashimoto (1972: 93ff). A somewhat different, and productive, morphological process involving tones can also be cited for Taishan, where the process is strictly agglutinative: a high pitch, [5], is simply added to the citation tone to form a contour tone (e.g., [33] + [5] --> [335], [21] + [5] --> [215], etc.) [5] is a tone morpheme that has a variety of functions, including those of changing a verb into a noun, and marking a verb as being in the durative aspect (similar to Mandarin zhe).

Tone change as a morphological process is also reported in other Chinese dialects, such as Wenling (Li 1978), belonging to the Wu dialect group, and in the Huangqiaozhen dialect of the Xiang dialect group (Tang 1960). In Wenling, for example, tone change could serve a nominalization function, producing words with specialized meanings, as in (iiii) below, where the base forms are adjectivals, bearing the citation tone, while the derived forms are nominals, bearing the changed tone, with its variants, [15] and [51]. (Note that example (iii.b) also shows some segmental changes as well.

(iii)	Citation Tone		Changed Tone (Terminal Syllable)	
a. Tone Ib:	$\hat{h}u\tilde{ɔ}^{31}$	'yellow'	$\hat{h}u\tilde{ɔ}^{15}$	'egg yolk'
b. Tone IVb:	ba^{21}	'white'	$b\tilde{a}^{51}$	'egg white'

Concerning the Chinese language in general, there are numerous articles on the topic of derivation by tone change. In the history of the Chinese language, such tone changes have often been analyzed in connection with the theory of tonogenesis. Haudricourt (1954), for example, suggests a hypothetical -s suffix as a derivational marker in Archaic Chinese on what became Ancient Chinese (Anc. Ch.) Qu tone forms -- that is, Tone IIIa and IIIb here -- and from which the falling tone originated. Examples are presented in (iv), where the falling tone is marked by an acute (‵). By positing the derivational suffix -s on the Archaic Chinese forms, Haudricourt is able to account for various cognate pairs in the language.

(iv)	Base Form		Derivational Form	
a.	âk	'bad'	âks (> Anc. Ch. âk‵)	'to hate'
b.	xâu	'good'	xâus (> Anc. Ch. xâu‵)	'to like'
c.	dâk	'to measure'	dâks (> Anc. Ch. dâk‵)	'a measure'
d.	ṣi	'to send'	ṣis (> Anc. Ch. ṣi‵)	'a messenger'

The extent to which derivation by tone change may still be a productive process in the modern dialects is a topic which remains to be explored, but one which is outside the scope of this paper. (For the Min dialects, for example, see Ting 1982?.) The main objective of the digression here is to outline the nature of bian yin "changed tone", in contradistinction to bian diao "tone sandhi", the main focus of this paper. The topic of "changed tones" will be dealt with again in the discussion of the Wenling dialect.

2. Hashimoto (1982), Ting (1992,1983), Chan (1985), and Pan (1986), for example, have presented arguments from various Chinese dialects for regarding the sandhi tones as sometimes preserving earlier tonal values. Thus, from a diachronic perspective, to regard the citation form as being the "original" tone would be incorrect in such cases.

3. See footnote 1 above for a discussion of bian yin.

4. The tone sandhi situation in Suzhou is somewhat more complex, in that there are cases where the initial syllable might also undergo tone sandhi. Those observations are not relevant for our purposes, however, and may be ignored. For a detailed study of the tone sandhi changes in Suzhou, see, for example, Ye (1979). It is important to realize, however, that Ye (1984) contains revisions on the differential behaviour of Tones IIb and IIIb following Tone IVa which is not reported in the earlier article, where Tones IIb and IIIb are treated identically in all environments.

5. There is a minor set of exceptions in which Tone IIIb behaves similarly to Tone IIb. In addition, there are many Tone IIIb syllables for which Ye (1984) is unable to determine whether or not they have an alternant [313] tone, because they do not combine readily with a Tone IVa syllable to form disyllabic words. In such cases, Ye treats these Tone IIIb syllables as bearing /31/, so that they are not distinct from Tone IIb syllables.

Observe, also, that in the discussion, we have been ignoring other contexts. This is because Tone IIIb behaves similarly to the rest of the tones in other contexts; namely, in non-initial position, all tones are completely neutralized, with the pitch value dependent on the tone value of the first syllable. This is the general picture; further details are not relevant to our discussion.

6. One might arguably select an alternant form of a tone that occurs in the environment preceding a "neutral"-tone syllable (that is, an unstressed syllable that lacks its own tone -- an atonic syllable). In my Fuzhou data, for example, Tone IIa is phonetically [32] in citation form, but [33] before an atonic syllable. Transcriptions of the citation form of Fuzhou Tone IIa in various sources include [22], [31], and [32]. (See Chan (1985) for details.) Which alternant one selects has important ramifications. The choice of [32] in my data, for example, would suggest that the underlying tone has a slight fall. The question, then, arises as to whether the phonology should reflect this minor drop in pitch. If, on the other hand, we choose the [33] form, we could argue very simply for an underlying level tone, since that is one of the alternants, and is the one we posit as the underlying form. (The solution is slightly more abstract in Chan (1985), involving an autosegmental treatment. The details, however, need not concern us here at this point.)

While the sources for different dialects are fairly informative with respect to citation tones (and to sandhi tones in the dialects which have them), the pitch shape of a tone preceding an atonic syllable is seldom described in the literature on a dialect. It is partially due to the paucity of such information that this paper focuses only on the importance of sandhi tones in determining the set of underlying tones.

The variant of a tone preceding an atonic syllable may be of significance historically, even if it is not chosen as the underlying tone synchronically.

One case in point is Tone IIa in the Shaowu dialect, discussed by Norman (1985). Tone IIa is a high level /55/ tone in citation form and elsewhere except before an atonic syllable, where it is a falling [53] tone. What is relevant to our present discussion is that Norman cites a nineteenth century source in which Tone IIa is actually a high falling tone. In the modern Shaowu dialect, it is thus the variant form of Tone IIa occurring before atonic syllables that has preserved the earlier pitch contour.

7. Observe that Yinchuan is one of the exceptional cases where the phonetic variant of a tone preceding an atonic syllable is also noted. The tonal value, [35], preceding the atonic syllable shows a similar conservatism that we found in the Shaowu dialect (footnote 6 above). Tone IIa in Yinchuan is one case where the sandhi tone, [35], and the tonal variant preceding an atonic syllable coincide in yielding evidence of an earlier, and very likely a more widespread contrast between Tones Ib and IIa in the dialect.

8. A more abstract solution for Ningde could be proposed within the autosegmental framework, which utilizes the concept of a "floating" tone or toneme. A floating tone is not realized phonetically unless it is "linked" to a tone-bearing unit. (The linking may be direct or indirect depending on which version of the model one is using.) Tone IIa in Ningde could be analyzed as a rising /LH/ tone historically, consisting of a sequence of two tonemes, L and H. Tone IIa underwent L-delinking at some stage, thereby setting the L toneme afloat. The tone is then realized as simply high, or [55], which is the variant serving as the underlying tone in this paper. The rising contour on the other sandhi form, [35], can be obtained by the linking of the L-toneme (already present in the underlying structure) in certain tone sandhi environments. The falling pitch on the citation form of Tone IIa can be viewed as a further development of the tone that may have originated as an intonational phenomenon, with falling pitch occurring in prepausal position. (This can be compared with Tone IIa in Fuzhou, which is [32] in prepausal position, but [33] before an atonic syllable. No information is available on the precise phonetic shape of Ningde Tone IIa before an atonic syllable.)

9. Some older Fuzhou speakers today still maintain the distinction between /-ʔ/ and /-k/.

10. There are, in addition, syntactic constraints on the lenition which need not concern us here. What is of further interest is that in the case of final /-ʔ/ from earlier *-k, even its deletion in faster speech does not trigger consonant lenition. Thus, there remains clear distinctions between the glottal stop from *-ʔ versus *-k.

11. By "citation" syllables, reference is to syllables occurring in all positions except the preterminal, tone sandhi position. They are stressed syllables, occurring either as isolation forms, or as the terminal syllable of a tone sandhi span. They may also be followed by an atonic syllable.

12. The word occurs with two different tones: /12/ (Tone IIIa) in the literary pronunciation, and /342/ (Tone IIIb) in the colloquial pronunciation.

References

Chan, Marjorie K. M. 1977. "A study of tonogenesis in Chinese." University of British Columbia. Ms.

__________. 1980. Zhong-shan Phonology: A Synchronic and Diachronic Analysis of a Yue (Cantonese) Dialect. M.A. thesis, University of British Columbia.

__________. 1985. Fuzhou Phonology: A Non-linear Analysis of Tone and Stress. Ph.D. dissertation, University of Washington.

Chen, Leo and Jerry Norman. 1965a. An Introduction to the Foochow Dialect. San Francisco: San Francisco State College.

__________. 1965b. Foochow-English Glossary. San Francisco: San Francisco State College.

Chen Xiwu. 1966. "Taishan fangyan teshu biandiao chu tan." (A preliminary investigation of the special changed tone in the Taishan dialect) Zhongguo Yuwen (1966) 1:34-36.

Hartman, Lawton. 1944. "The segmental phonemes of the Peiping dialect." Language 20:28-42.

Hashimoto, Anne Yue. 1972. Phonology of Cantonese. Cambridge: Cambridge University Press.

Hashimoto, Mantaro. 1982. "The so-called 'original' and 'changed' tones in Fukienese: a case study of Chinese tone morphophonemics." Bulletin of the Institute of History and Philology 53.4:645-659.

Haudricourt, André-G. 1954. "De l'origine des tons en Vietnamien." Journal Asiatique 242:68-82.

He, Wei. 1984. "Luoyang fangyan jilue." (A sketch of the Luoyang dialect) Fangyan 6.4:278-299.

Hockett, Charles F. 1947. "Peiping phonology." Journal of the American Oriental Society 67:253-267.

Hou, Jingyi. 1980. "Pingyao fangyande liandu biandiao." (Tone sandhi in the Pingyao dialect). Fangyan 2.1:1-14.

Kratochvil, Paul. 1968. The Chinese Language Today: Features of an Emerging Standard. London: Hutchinson and Company.

Li, Rong. 1966. "Wenling fangyan yuyin fenxi." (Phonetic analysis of the Wenling dialect) Zhongguo Yuwen (1966) 1:1-9.

__________ 1978. "Wenling fangyande bianyin." (Tonal suppletion in the Wenling dialect) Zhongguo Yuwen (1978) 2:96-103.

__________. 1979. "Wenling fangyande liandu biandiao." (Tone sandhi in the Wenling dialect) Fangyan 1.1:1-29.

Norman, Jerry. 1971. A Guide to the Foochow Dialect. Princeton University, N.J.: Chinese Linguistics Project.

__________. 1977. "A preliminary report on the dialects of Mintung." Monumenta Serica XXXII:326-348.

__________. 1985. "Two early sources on the Shaowu dialect." Journal of Chinese Linguistics 13.2:331-345.

Pan Wuyun. 1986. "Gu Wuyude shengdiao gouni." (Reconstruction of the tones in Proto-Wu) Paper presented at the Conference on Languages and Dialects of China, Berkeley.

Tang, Zuofan. 1960. "Hunan Dongkouxian Huangqiaozhen fangyan." (The dialect of Huangqiao town in Dongkou county, Hunan Province) Yuyanxue Luncong. Volume 4. Shanghai: Shanghai Jiaoyu Chubanshe.

Tay, Mary Wan Joo. 1968. A Phonological Study of Hokkien. Ph.D. dissertation, University of Edinburgh.

Ting, Pang-hsin. 1982. "Some aspects of tonal development in Chinese dialects." Bulletin of the Institute of History and Philology 53.4:629-644.

__________. 1982(?). "Derivation by tone-change in Colloquial Min. Ms.

__________. 1983. "Wuyu shengdiaozhi yenjiu." (A study of the tones in the Wu dialects) Ms.

Yang, Paul F-M. 1960. "Hotei toryo hogen no seicho." (The tones in the Donglu, Baoding dialect). Chugoku Gogaku 97.4:9-13.

Ye, Xiangling. 1979. "Suzhou fangyande liandu biandiao." (Tone sandhi in the Suzhou dialect) Fangyan 1.1:30-46.

__________. 1984. "Guanyu Suzhou fangyande diaolei." (Tone-classes in the Suzhou dialect) Fangyan 6.1:15-18.

Woo, Nancy. 1969. Prosody and Phonology. Ph.D. dissertation, M.I.T. Reproduced in 1972 by the Indiana University Linguistics Club.

Hemispheric Specialization For Voice Recognition: Evidence From Dichotic Listening*

Jody Kreiman and Diana Van Lancker

Dichotic listening studies of voice recognition have failed to produce clear evidence for or against lateralization of this function to either hemisphere. A right ear advantage (Doehring & Bartholomeus, 1971), a left ear advantage (Riley & Sackeim, 1982) and zero ear advantages (Bartholomeus, 1974a, b; Tartter, 1984) have been reported, as has a sex-of-speaker by ear interaction effect (Buttet, Thuillard, Assal & Aubert, 1983).

The literature on voice recognition in brain-damaged patients provides a possible explanation for these varying results. Studies of familiar voice recognition and unfamiliar voice discrimination in this population suggest that recognition is lateralized to the right hemisphere (Van Lancker & Canter, 1982), while successful discrimination requires both hemispheres (Assal, Zander, Kremin & Buttet, 1976; Van Lancker & Kreiman, 1985; Van Lancker, Kreiman & Cummings, 1985). The inconsistent results produced by dichotic listening studies may thus be due to their exclusive use of <u>unfamiliar</u> voices: If both cerebral hemispheres are required for processing unfamiliar voice information, as data from clinical subjects suggest, then one would not expect a clear ear advantage to emerge when these are used in dichotic listening. In fact, the data on brain-damaged patients strongly suggest a right hemisphere specialization exists for familiar voice recognition only (Van Lancker & Kreiman, in press).

To examine hemispheric specialization for voice identification in normal subjects, we have therefore used a set of familiar voices as stimuli in the present study. To ensure the stimuli were uniformly familiar to an adequate number of subjects, famous male voices were selected. Our hypothesis was that individual voice quality is an auditory pattern, and that voices would therefore be better recognized at the left ear in a dichotic listening paradigm, even though carried in a speech stimulus. This hypothesis is based on a model of brain function that associates superior pattern recognition abilities with the right hemisphere (Bogen, 1969; Bradshaw & Nettleton, 1983; Levine & Koch-Weser, 1982), in contrast to the sequencing and analytic functions of the left hemisphere (Bever, 1975; Levy, 1974). This model further interprets superior performance at the left ear as a reflection of an underlying right hemisphere specialization for the genre of stimuli (e.g., Bryden, 1982; Kimura, 1967).

To test our hypothesis most stringently, listeners were required to identify both the linguistic and voice information that occurred at each stimulus presentation: Subjects monitored a single ear on each trial, and always identified both the word and the speaker. We hypothesized that the word identity of each stimulus would be better identified at the right ear (e.g., Kimura, 1967), while the specific identity of the speaker of the same stimulus would be better recognized at the left ear of the same subjects.

Method

Stimuli

From a set of 50 famous male voices previously studied (Van Lancker, Kreiman & Emmorey, 1985; Van Lancker, Kreiman & Wickens, 1985), the 16 voices familiar to the most listeners were selected for use in this experiment (Table I). Two of the voices were designated practice voices and were used only in training trials.

Table I

Stimulus Voices

Jack Benny	John F. Kennedy*
George Burns	Jack Klugman
Johnny Carson	James Mason
Maurice Chevalier*	Bob Newhart
Walter Cronkite	Richard Nixon
W.C. Fields	Vincent Price
Henry Fonda	Tony Randall
Alec Guinness	Ronald Reagan

* Practice voices

To prepare the stimulus tapes, four words were excised from each speaker's recorded voice sample. Care was taken to avoid words which could cue the speaker's identity. Editing was done on the PDP-11/23 computer in the UCLA Phonetics Laboratory. Speech samples were low-pass filtered at 4 kHz and sampled at 10 kHz. Expanded displays were used so that word onsets and offsets could be accurately located. All words had 2-4 syllables; durations ranged from 250 msec to 650 msec, with a mean duration of 454.77 msec (s.d. = 109.33).

In order to avoid effects of specific pairings of words and voices, each speaker was paired with four other speakers (one for each excised word), for a total of 56 stimulus pairs. Words were matched to within 40 msec in duration. Individual stimulus words were "merged" to produce dichotic pairs with one word on each channel. Stimulus onsets were exactly aligned; alignments were verified on two-channel oscillograms. After merging, dichotic pairs were played out onto stereo tapes. Recording levels were equalized during play-out. Stimulus pairs were separated by 13 seconds; each pair was preceded by its consecutive number spoken in a female voice.

Subject Selection

In order to ensure that subjects were genuinely familiar with all test voices, a listening pretest was prepared, made up of 4-second samples of continuous speech for each target speaker. No test stimulus words appeared in these samples, nor did samples contain speech trademarks or typical topics which could cue the speaker's identity. The pretest samples were randomly ordered and presented to candidate subjects, who were asked to match each voice sample to a list of 25 names (the 14 target voices, 2 practice voices, and 9 distractors). Only listeners who correctly identified all of the target voices participated in the dichotic listening study.

The first 15 listeners to successfully complete the pretest were given the dichotic listening test. All subjects were right-handed with no familial sinistrality. All reported normal hearing in both ears.

Design and Procedure

Dichotic stimuli were arranged into four blocks of 28 trials for a total of 112 trials. Headphone left-right orientation was switched after the first and third blocks, and initial orientation was alternated across subjects. Subjects were instructed to monitor one ear at a time and to ignore the competing stimulus in the other ear. To discourage subjects from developing a response set, the ear monitored was alternated every 14 trials. Within a block of 28 trials, each voice was monitored twice and ignored twice; a different word was used each time a voice occurred. In the next block, the two ignored words were monitored, and the monitored words ignored. Thus, across the four blocks of trials, each word was monitored twice and ignored twice, once through each side of the headphones.

Eight training trials preceded the test trials. For the first four trials, subjects were told in advance what voices and words they would hear; they then heard four trials for which post-trial feedback was provided. They were given the opportunity to repeat these training trials until they felt comfortable with the dichotic task.

For each test item, subjects reported both the word and the voice heard at the monitored ear by selecting one each of six words and six names provided on answer sheets. Names included the target voice, the voice in the unattended ear, and two foils for each (all from the target set). Care was taken to select foils which were similar enough in vocal quality to be plausible choices. Words included the target and the word in the unattended ear; word foils were selected by playing the tape to 4 naive listeners who wrote down the words they thought they heard in each ear. Their misperceptions were used as foils. No attempt was made to control order of report; however, for half the items, words were listed first on the answer sheet, and for the other half of the items, voices were listed first, alternated every other test item.

All instructions to subjects were recorded on the stimulus tape. The test was administered using an Aiwa HS-J02 tape player and Sony MDR-80 headphones. This apparatus was calibrated by observing the output (on an oscilloscope) when a 1 kHz tone was played through each channel; relative output at the two ears was equal. Testing took approximately 40 minutes.

Results

For each listener, the percent correct recognition for each task (word and voice recognition) at each ear, as well as the laterality index λ (lambda; (Bryden, 1982)) were calculated for each task[1]. Mean percent correct results are given in Table II[2]; chance for both tasks equals 16.7% correct.

A two-way (task x ear) ANOVA revealed significant effects of task (F (1,14) = 122.74, $p < 0.01$) and ear (F (1,14) = 10.77, $p < 0.01$) on percent correct: All subjects performed better on the word recognition task, and scores for both tasks were better overall at the right than at the left ear. A significant task x ear interaction also occurred (F (1,14) = 7.74, $p < 0.01$): Performance was significantly better at the right ear for the word recognition task (t (14) = 3.75, $p < 0.01$), but no significant difference between ears occurred for the voice recognition task (t (14) = 1.85, n.s.). Results of analyses using λ confirmed the above finding: A significant right ear advantage emerged for the word recognition task ($z = 5.24$, $p < 0.01$), while no significant ear advantage was observed for voice recognition ($z = 1.81$, n.s.).

Table II
Scores on Voice and Word Recognition Tasks

	Voices		Words	
	Left ear	Right ear	Left ear	Right ear
% correct	45.83	50.25	65.11	76.78
SD	7.96	7.05	9.48	9.06

To investigate individual patterns of lateralization, following Lauter (1983) we calculated for each subject the difference in performance at the two ears for each task and plotted these to reveal individual patterns of relative ear advantages (Figure 1). This analysis revealed that 5 of the 15 subjects showed essentially no difference in ear advantages for the two tasks, while the other 10 did show a difference in the extent to which the two tasks are lateralized: Although <u>absolute</u> ear advantages for the two tasks vary, for all but one subject in this group the word recognition task produced a large <u>relative</u> right ear advantage, and the voice task a relative left ear advantage. The significance of individual differences in performance was tested using the λ measure; most subjects had no significant ear advantage for the voice recognition task, and about half (7/15) had a significant right ear advantage for the word recognition task ($p < 0.05$).

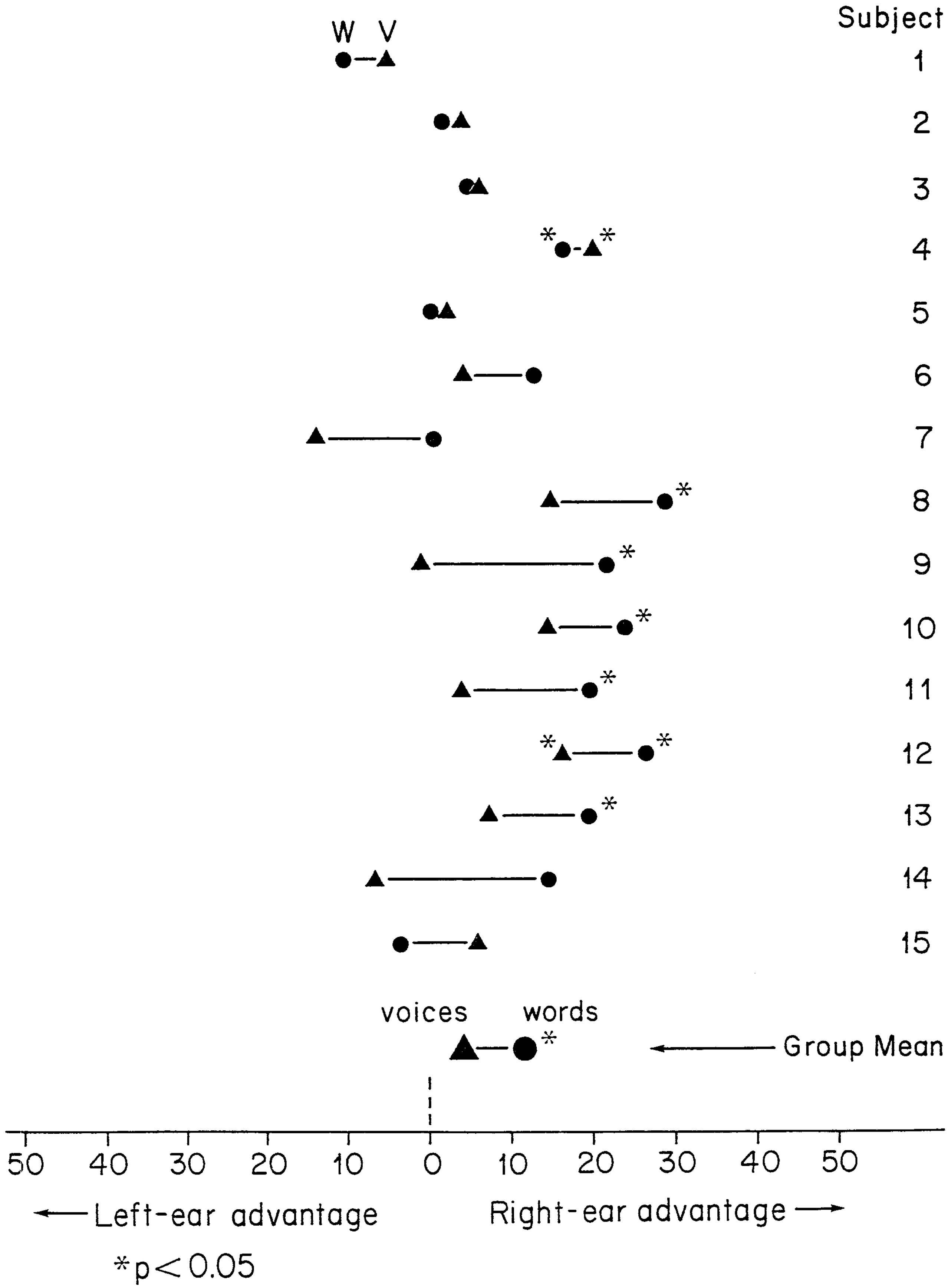


Figure 1. Relative ear advantages for individual subjects and for the group as a whole on the voice and word recognition tasks.

Discussion

As a group, subjects in this experiment showed the expected significant right ear advantage for word recognition, coupled with a zero ear advantage for voice recognition. The "strong" version of our initial hypothesis--that a left ear advantage for famous voice recognition (alongside a right-ear advantage for word recognition in these bidimensional stimuli) would be observed--is thus not supported. However, a weaker version of this hypothesis is supported, in that a relative left ear advantage for voice recognition compared to that observed for word recognition did emerge. This relative (and statistically significant) difference, found in a task where subjects were required to simultaneously process two different aspects of the same stimuli, suggests that voices and words are processed by different cerebral mechanisms.

A reasonable minimal conclusion from these findings is that words are processed in the left hemisphere, whereas voice information is not necessarily so. However, evidence from other research, indicating that prosodic information in speech is better processed in the right hemisphere in normal subjects (e.g., Ley & Bryden, 1982), and is deficient in right brain-damaged subjects (Ross, 1981; Van Lancker & Kreiman, 1985) points to a right hemisphere involvement underlying these results.

One explanation of the lack of a significant absolute left ear advantage for voices may come from examination of individual subjects' results. More than half our individual listeners did not produce the expected significant right ear advantage for word recognition. These results, along with the lack of a left ear advantage for voices, may both be due to the "bidimensional stimulus" design of this experiment. That is, expected right and left ear advantages for different "levels" of the same stimulus might have been attenuated by the task, which had listeners simultaneously attend to both aspects. The lack of a left ear advantage for voice recognition may also have been due to the salience of the linguistic content overriding the voice information. Subjects in fact reported that the word identification task was easier to focus on than was the voice identification task. This interpretation can be tested by dichotically presenting backward voice stimuli, which would be free of linguistic content ("one-dimensional stimuli"), and requiring only a voice identification response. If our interpretation is correct, a clear left-ear advantage should emerge from such a design.

Footnotes

*We thank Karen Emmorey for her help in preparing the stimuli used in this experiment.

1. λ is given by:

$$\lambda = \ln((\text{Correct R} \times \text{Wrong L})/(\text{Wrong R} \times \text{Correct L}))$$

and its significance is tested using z tables, where

z = \ /((1/Correct L)+(1/Correct R)+(1/Wrong L)+(1/Wrong R)).

\ has two advantages over other, more commonly used laterality indices: it does not fluctuate with differences in relative task difficulty, and significance levels may be calculated for individual subjects as well as for group data (see Bryden, 1982, for further discussion).

2. Intrusion errors were also examined, but the small number that occurred (an average of 6.33 for voices and 1.8 for words) could have been made by chance, and no formal analyses were undertaken.

References

Assal, G., Zander, E., Kremin, H., and Buttet, J. (1976). Discrimination des voix lors des lésions du cortex cérébral. *Archives Suisses de Neurologie, Neurochirurgie, et de Psychiatrie* 119.307-315.

Bartholomeus, B. (1974a). Effects of task requirements on ear superiority for sung speech. *Cortex* 10.215-223.

Bartholomeus, B. (1974b). Dichotic singer and speaker recognition. *Bulletin of the Psychonomic Society* 2.407-408.

Bever, T. G. (1975). Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: holistic and analytic. *Annals of the New York Academy of Science* 19.251-262.

Bogen, J. (1969). The other side of the brain. *Bulletin of the Los Angeles Neurological Society* 34.73-105.

Bradshaw, J. L., and Nettleton, N. C. (1983). *Human Cerebral Asymmetry*. Englewood Cliffs, N.J.: Prentice-Hall.

Bryden, M.P. (1982). *Laterality: Functional Asymmetry in the Intact Brain*. New York: Academic Press.

Buttet, J., Thuillard, F., Assal, G., and Aubert, C. (1983). Human voice recognition: A dichotic listening test. Presented at the International Neurological Society, Lisboa.

Doehring, D.G., and Bartholomeus, B. (1971). Laterality effects in voice recognition. *Neuropsychologia* 9.425-430.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex* 3.163-178.

Lauter, J. (1983). Stimulus characteristics and relative ear advantages: A new look at old data. Journal of the Acoustical Society of America 74.1-17.

Levine, S. C., and Koch-Weser, M. P. (1982). Right hemisphere superiority in the recognition of famous faces. Brain and Cognition 1.10-22.

Levy, J. (1974). Psychobiological implications of bilateral asymmetry. In Hemispheric Function in the Human Brain, S. J. Dimond and J. G. Beaumont (Editors). New York: John Wiley and Sons.

Ley, R.G., and Bryden, M.P. (1982). A dissociation of right and left hemisphere effects for recognizing emotional tone and verbal content. Brain and Cognition 1.3-9.

Riley, E.N., and Sackeim, H.A. (1982). Ear asymmetry in recognition of unfamiliar voices. Brain and Cognition 1.245-258.

Ross, E. (1981). The aprosodias: Functional-anatomical organization of the affective components of language in the right hemisphere. Archives of Neurology 38.563-568.

Tartter, V.C. (1984). Laterality differences in speaker and consonant identification in dichotic listening. Brain and Language 23.74-85.

Van Lancker, D., and Canter, G. (1982). Impairment of voice and face recognition in patients with hemispheric damage. Brain and Cognition 1.185-195.

Van Lancker, D., and Kreiman, J. (1985). Unfamiliar voice discrimination and familiar voice recognition are independent and unordered abilities. UCLA Working Papers in Phonetics 62.50-60.

Van Lancker, D., and Kreiman, J. (In Press). Preservation of familiar speaker recognition but not unfamiliar speaker discrimination in aphasic patients. To appear in Clinical Aphasiology 1986, R. Brookshire (Editor).

Van Lancker, D., Kreiman, J., and Cummings, J. (1985). Voice recognition and discrimination: New evidence for a double dissociation. Journal of Clinical and Experimental Neuropsychology 7.609.

Van Lancker, D., Kreiman, J., and Emmorey, K. (1985). Famous voice recognition: Patterns and parameters. Part I: Recognition of backward voices. Journal of Phonetics 13.19-38.

Van Lancker, D., Kreiman, J., and Wickens, T. (1985). Famous voice recognition: Patterns and Parameters. Part II: Perception of rate-altered voices. Journal of Phonetics 13.39-52.

"Tense" and "Lax" Revisited: more on phonation types and pitch in minority languages of China.

Ian Maddieson and Susan A. Hess

Paper presented at the 19th International Conference on Sino-Tibetan Languages and Linguistics, Columbus, Ohio.

A contrast between "tense" and "lax" vowels is a common feature of the minority languages spoken in Southwestern China and adjoining parts of Southeast Asia. The phonetic basis of this contrast has been studied in a previous paper (Maddieson & Ladefoged 1985). There, we found that, although there were often several measurable phonetic differences between a minimal pair of words contrasting tense and lax vowels, the consistent difference was one of phonation type - at least for the languages and speakers that we studied. It did, however, seem to us that the nature of the phonation type contrast differed somewhat between those languages in which the tense/lax contrast is traced back to historical syllable-initial voicing and those in which it is traced back to a checked/unchecked syllable distinction. Our phonetic analyses were based on limited data, sometimes only from a single speaker, and left a number of unanswered questions. For this reason we have been continuing research into this topic. With the assistance of Ren Hongmo, we obtained additional recordings of languages with tense/lax contrasts. In this progress report we will discuss some aspects of the contrast in three languages - Wa, Jingpho, and Liangshan Yi - and comment on some implications of these results, particularly with respect to the hypothesis that phonation type may play a role in tonogenesis or tone-splitting.

The acoustic measure which we have used to establish difference of phonation type is illustrated in figure 1. We obtain the power spectra of the vowels to be compared using a digital spectrograph, as in the upper part of the figure. The length of each spike in these spectra reflects the amplitude of the harmonic component at the frequency in question. For each vowel, we measure the difference in amplitude between the fundamental frequency (indicated as F_0) and the second harmonic (indicated as H_2) and compare the values for tense vowels with those for lax vowels. Several studies have shown that this measure successfully differentiates between different

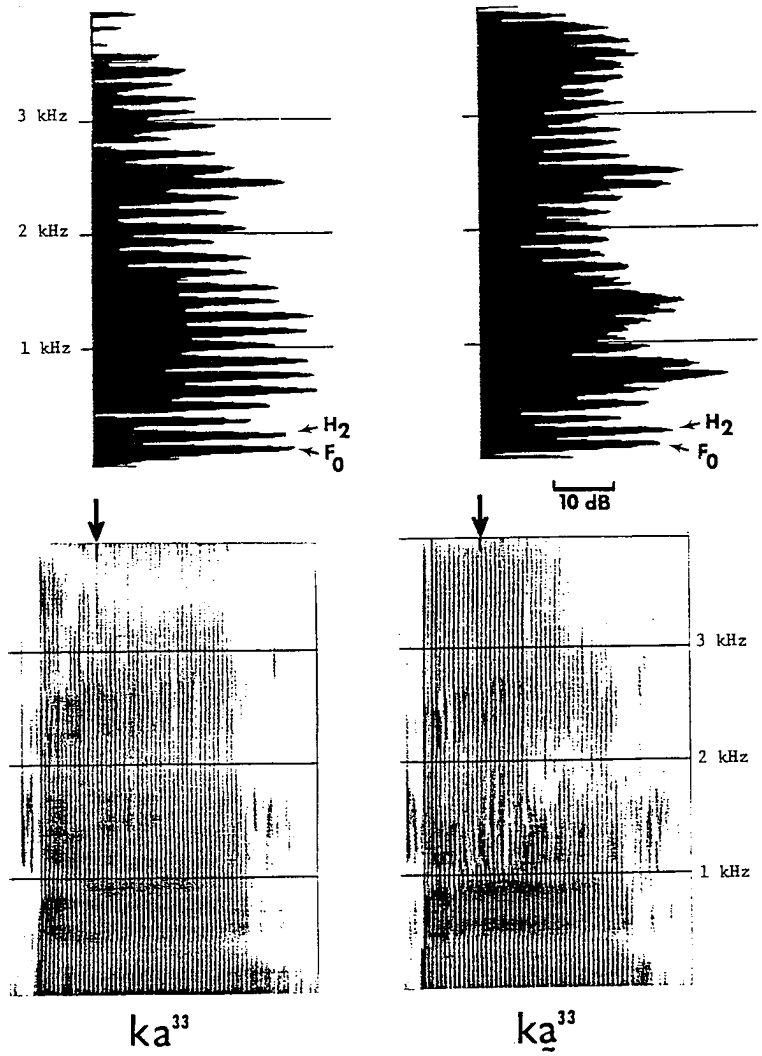


Figure 1. Wideband spectrograms (below) and spectral displays (above) of a pair of Jingpho mid-tone words with lax (left) and tense (right) vowels. The spectrum is of a 22 ms window preceding point indicated by the arrow on the wideband spectrograms.

phonation types, and that it correlates well with auditory judgements of degree of breathiness (see references in Maddieson & Ladefoged 1985).

First, Wa. This language, or more properly group of languages, belongs to the Mon-Khmer family (Diffloth 1980). Like most members of this family, Wa is non-tonal. Our speakers are speakers of the Parauk (Burao) dialect from Cangyuan and Gengma counties in Yunnan. This variety of Wa has nine vowel qualities, which occur as both tense and lax vowels (Qiu, Li & Nie 1980). The words we examined illustrate five of these pairs. The list is given in table 1. Note that in this and subsequent tables the tense member of a vowel pair is indicated by a stroke under the vowel.

Table 1. Wordlist illustrating tense and lax vowels in Wa.

ta̲	pale	ta	measuring cup
pɔ̲	psoas muscle	pɔ	don't leave
tɔ̲	besides	tɔ	carry
ka̲	bake	ka	gnaw
kɛ̲	mid-spring	kɛ	trap (n.)
rɤ̲	drag	rɤ	boat
ba̲	thigh	ba	drum
lɔ̲	excuse (n.)	lɔ	mule
kra̲m	raft	kram	indigo
la̲k	crow	lak	lakh (10,000)

Measurements of the difference in amplitude between H_2 and F_0 were made on 2 repetitions of these 10 pairs of words for 3 speakers. The mean value for $H_2 - F_0$ for "tense" vowels is 5.65 dB, and for "lax" vowels it is 3.68 dB. Although this difference is small it is statistically significant, as evaluated by a paired data t-test ($p < 0.005$, $t=2.94$, $df=59$). Maddieson and Ladefoged (1985) commented in the earlier paper that, for the one speaker examined there, "measurements of [vowel] duration showed a general trend for the tense vowels to be a little shorter than their lax counterparts." In the much larger body of data now available the vowel length difference appears negligible - an insignificant 7 ms. We will comment on pitch measurements in all three languages in a separate section below.

Yunnanese Jingpho as spoken in Yingjiang county has five vowel qualities which occur in both tense and lax forms (Liu 1964). The words we examined illustrate two of these pairs. The wordlist is given in table 2. Tones are indicated by raised numbers after the segmental transcription, with high pitch indicated by 5, low by 1.

Table 2. Words illustrating tense and lax vowels in Jingpho.

ka^{31}	speech	ka̲^{31}	dance
kaŋ^{33}	pull	ka̲ŋ̲^{33}	tense
kak^{31}	peck	ka̲k^{31}	caw
kʒa^{31}	sharp	kʒa̲a̲^{31}	reveal
pat^{55}	with a whip	pa̲t^{55}	stop up
ko^{33}	lay bricks	ko̲^{33}	give
tom^{31}	draw back	to̲m^{31}	conclude
po^{33}	heat	po̲^{33}	birth
pot^{31}	feed	po̲t^{31}	angry
kʒop^{31}	smash	kʒo̲p̲^{31}	hoarse

For 2 repetitions of these 10 words by 3 speakers of Jingpho the mean $H_2 - F_0$ for "tense" vowels is 9.17 dB and for "lax" vowels it is 2.01 dB. This difference is highly significant ($p < 0{,}0001$, $t=12.03$, $df=59$), and agrees well with the finding of a highly significant difference for two speakers in our earlier work. It also confirms the earlier indications that the tense/lax phonation contrast is more salient in Jingpho than it is in Wa. Again, no significant difference in the length of tense and lax vowels was found. Thus both Wa and Jingpho maintain the tense/lax contrast by phonatory differences, and not by vowel length, nor, to judge from auditory impressions, by salient differences in vowel quality. In both languages, the "lax" vowels have a somewhat more "breathy" quality than the "tense" vowels.

The third language is one which was not studied by Maddieson and Ladefoged (1985). We have called this language "Liangshan Yi" since that is the name used by Li and Ma (1983). Our 3 speakers are not specifically from Liangshan, but speak what is apparently the same language. Li and Ma divide the vowels of Liangshan Yi into 5 "lax" and 5 "tense" vowels. Two of the "tense" vowels are described as "tense glottal", the others seem to differ from their lax counterparts in height rather than by laryngeal features. But the two "tense glottal" vowels contrast with lax counterparts that are similar in quality. The Liangshan Yi wordlist, given in table 3, therefore consists of words contrasting only these vowels. Note that these vowels are extremely high in their articulation – one is front and is symbolized in very broad transcription by /i/, and one is back with a very close, but, we believe, not rounded lip position; this is symbolized with /u/. These "vowels" are barely vowels at all, being in some ways more like syllabic consonants. There is considerable allophonic variation induced by the preceding consonant. In particular, the lips may be close enough in /u/ to produce labial friction, trilling, or – in nasal contexts – complete occlusion.

Table 3. Words illustrating contrasts between tense and lax high vowels in Liangshan Yi.

pu	hedgehog	p_u_	rummage
phi	hot	ph_i_	fold
p^hu	price	$p^h\underline{u}$	renege
bu	mosquito	b_u_	write
mbu	yell at someone	mb_u_	sated
hmu	fungus	hm_u_	expand
fu	burn	f_u_	marry
vu	intestines	v_u_	kidneys
vi	buy	v_i_	grab
tu	small bell	t_u_	thousand

Although the difference between "tense" and "lax" vowel pairs is quite distinctive, with an auditorily "harsher" quality for the tense members, the $H_2 - F_0$ measure does not distinguish between the tense and lax members. The mean values of $H_2 - F_0$ for two repetitions of the words in table 3 for 3 speakers are −0.18 dB for tense vowels, and 0.64 dB for lax vowels ($p < 0.25$, $t=-1.16$, $df=59$). It may be that this measurement is simply not an appropriate one with which to detect phonatory differences in the range of phonetic segments that occur as allophones of these so-called "vowels". But we think that it is more likely that the tense/lax contrast is produced in a different way in this language. We speculate that the "tense" vowels employ a supralaryngeal mechanism like that used in producing the "pressed" or "strident" vowels found in some of the Khoisan languages of Southern Africa. This involves a narrowing between the base of the epiglottis and the upper part of the arytenoid cartilages which induces turbulence or vibration at this location. The use of this mechanism in !Xóõ is described in some detail by Traill (1985). Traill has listened to our Yi recordings and agrees that there is an auditory similarity between the strident vowels of !Xóõ and the tense vowels of Yi (Traill, personal communication). Obviously, the phonetics of this language deserve closer investigation, and we are making efforts to obtain more complete materials, on which we will report in due course.

In all three languages we also measured the fundamental frequency at the onset and the offset of the vowel in order to see whether the tense/lax difference correlated with some difference in pitch. This is of particular interest in view of the spreading belief that phonation type differences in vowels have played a role in the tonal evolution of Chinese (as well as some other languages), in particular, in the register split. For example, Pulleyblank, after considering both historical reconstruction and descriptive sources, concludes that:

> " ... in northern standard Chinese...the split into upper and lower registers was conditioned primarily by voiced aspiration, giving rise to breathy vowels, rather than simply voice...."
> (Pulleyblank 1978: 173).

The pitch lowering effect of prevocalic voiced consonants has been widely studied, but there seems little comparable data on the effect on pitch of phonation type differences in vowels. Our measurements of the frequency of F_0 in our present materials are given in table 4.

Table 4. F_0 measures on tense and lax vowels (in Hz).

	Wa		**Jingpho**		**Yi**	
	onset	offset	onset	offset	onset	offset
"tense"	146	112	157	128	157	153
"lax"	145	115	145	126	152	154
signific- ance level	n.s.	n.s.	$p<.0001$	n.s.	$p<.0001$	n.s.

Briefly, no pitch difference was observed in Wa, but a significant pitch difference at the vowel onset was observed in Jingpho. Wa is nontonal and might have been expected to permit F_0 to be influenced more than Jingpho, where pitch has a lexically distinctive function. But recall that the phonation contrast is less salient in Wa, and the different pitch effects in the two languages might be accounted for because of that. On the other hand, Jingpho might be doing no more than preserving a lowering of the onset pitch in syllables that had an original voiced initial consonant, i.e. in those syllables which are now lax syllables. We also see significantly lower pitch in lax syllables in Liangshan Yi, where initial consonant voicing is not the historical precursor of the tense/lax contrast. (In Yi tense vowels occur in syllables that earlier had final stops.) This, perhaps, is evidence that "lax" syllables are generally lower in pitch, regardless of historical origin. However, we are not sure that the tense/lax contrast in this language is based on phonation type! Hence we have at best ambiguous signals on the question of how phonetically well-motivated it is to suggest that a tonal register distinction may have its origin in a contrast of plain and breathy phonation.

References

Diffloth, Gerard. 1980. The Wa Languages (Linguistics of the Tibeto-Burman Area 5.2). California State University, Fresno.

Li Min and Ma Ming. 1983. Liangshan Yiyu yuyin gailun [Description of the sounds of the Liangshang Yi language]. Sichuan Minzu Chubanshe, Chengdu.

Liu Lu. 1964. Jingpoyu gaikuang [Brief description of the Jingpho language]. Zhongguo Yuwen 132: 408-417.

Maddieson, Ian and Peter Ladefoged. 1985. "Tense" and "lax" in four minority languages of China. Journal of Phonetics 13: 433-454.

Pulleyblank, E. G. 1978. The nature of the Middle Chinese tones and their development in Early Mandarin. Journal of Chinese Linguistics 6: 173-203.

Traill, Anthony. 1985. Phonetic and Phonological Studies of !Xóõ Bushman (Quellen zur Khoisan-Forschung 1). Helmut Buske, Hamburg.

Qiu Efeng, Li Daoyong & Nie Xizhen. 1980. Wayu gaikuang [Brief description of the Wa language]. Minzu Yuwen 1980.1: 58-69.

Morphemic Structure in the Mental Lexicon

Karen Emmorey

There is considerable evidence that the linguistic processing system differentiates between affixes and word stems. Fromkin (1971) presents speech error evidence which indicates that stem morphemes and affixes operate as independent speech units. For example, stem morphemes can be exchanged leaving behind a suffix -- "a floor full of holes" becomes "a hole full of floors". Garrett (1980) shows further that affixes and word stems exhibit different speech error patterns. Wordstems often participate in exchange errors (as in the above example), but affixes seldom if ever do. Affixes, on the other hand, often participate in shift errors but word stems (of major grammatical categories) rarely do.

Aphasia data also indicate that affixes and base forms are processed differently. Buckingham (1981) discusses patients who produce jargon forms which are appropriately inflected:

a) The leg [vIltəd] from here down.
b) This is the [kreibəkæ ks] where the [frəǰə ž] get out after the [čuw].

In these examples nominal suffixes attach to "nouns" in noun phrases and verbal suffixes attach to "verbs". The jargon primarily effects the lexical or stem morphemes. In addition, Broca's aphasics classically have difficulty producing grammatical suffixes but can produce the correct stem morpheme (Goodglass and Kaplan, 1972). These patterns of performance again indicate that grammatical affixes and lexical morphemes are treated differently within the linguistic system.

However, how morphological structure is represented and accessed in the mental lexicon remains unclear, and only recently have psycholinguistic models of the lexicon recognized the potential importance of morphological structure to word recognition and language comprehension. The Taft and Forster (1975) prefix stripping model has received much attention because it provides testable predictions about how morphologically complex words are recognized. According to their model, prefixed words do not have separate representations within the lexicon but are accessed by the stem, and prefixes must be recognized and removed prior to searching the lexicon for the stem. Taft (1979) extends the prefix stripping model to suffixes, and we will therefore refer to this model as the "Affix Stripping" model. Most of the evidence that bears on the question of whether morphologically complex words are decomposed prior to lexical access comes from experiments with visual word recognition. The experiment described below used an auditory lexical decision task to investigate some predictions of the Affix Stripping model proposed by Taft and Forster (1975) and Taft (1979) for auditory word recognition. Other factors that may influence how morphologically complex words are recognized were investigated as well.

First, if affix stripping occurs, it should slow the recognition of bimorphemic words. That is, if bimorphemic words must initially be parsed into stem and affix before the stem can be accessed in the lexicon, they might take longer to recognize than monomorphemic words which do not require such preprocessing. Note that Taft and Forster (1975) do not assume that the recognition and removal of prefixes prior to lexical search requires additional processing time. However, this assumption appears ad hoc and counter-intuitive. Psychological models using data from timed experiments generally assume that additional computations require additional processing time. For example, the Derivational Theory of Complexity (see Fodor, Bever, and Garrett, 1974) assumed that passive sentences take longer to comprehend than active sentences because more transformations must be undone. More recently, reading models assume that access to the lexicon via a grapheme-to-phoneme conversion process is slower than immediate access to the orthographic representation (Coltheart, 1980; Gfroerer, Guntler, and Weiss, 1984). We will assume that affix stripping does require some processing time and hypothesize that this procedure will slow recognition of bimorphemic words compared to monomorphemic words. Of course, if no difference between these word types is found, we cannot conclude that affix stripping does not take place since parallel processing may also occur.

Taft and Forster's model also predicts that pseudoaffixed words (e.g. invite; temper) will take longer to process than true affixed words (e.g. insane; winner) because the pseudoaffixed words are initially misparsed as affix+stem (or stem+affix). Since the "stem" of a pseudoaffixed word (e.g. vite) is not present in the lexicon, a second search occurs for the whole word; thus, recognition of pseudoaffixed words is delayed. Both true and pseudoaffixed words were included to test this hypothesis. Note that recognition of pseudoaffixed words should also be delayed in comparison to monomorphemic words. If these results are observed, it will argue for mandatory morphological decomposition prior to lexical access.

A third question addressed by the experiment was whether the linguistic distinction between derivational and inflectional morphology is relevant to processing mechanisms at the level of word recognition. It is clear that at some level inflectional morphology is processed differently than derivational morphology. Agrammatic patients fail to produce inflectional suffixes while (some) derivational affixes remain intact (Kean, 1977). Recent studies also suggest that free grammatical morphemes are accessed and stored differently than free lexical morphemes and that the access procedure for free grammatical morphemes is lost with agrammatism (Bradley, Garrett, and Zurif, 1980; but see Gordon and Caramazza, 1982). If free grammatical morphemes are processed differently during word recognition, are bound grammatical morphemes accessed differently as well? Psycholinguistic studies have used both inflectional and derivational morphology as stimuli (Manelis and Tharp, 1977; Henderson, Wallis, and Knight, 1984), and thus this question remains unanswered.

Word recognition of both prefixed and suffixed words was also investigated to determine if a processing difference exists between them. Cutler, Hawkins, and Gilligan (1985) propose an explanation for languages' preference for suffixation based in part on processing constraints. The authors argue that stems are processed prior to affixes, and thus stems are favored to be in the most salient position of the word -- the beginning. In determining the entire meaning of the word from its parts, the stem has computational priority over the affix. For prefixed words the stem is not in the most salient initial portion of the word, and the point at which the stem may be recognized will be delayed. Thus recognition of prefixed words may be delayed compared to suffixed words.

To summarize, the following questions are addressed:

1. Do bimorphemic words require more processing time than monomorphemic words?

2. Are listeners mislead by pseudoaffixed words (e.g. invite, motive) and forced to reanalyze the structure causing a delay in reaction time?

3. Will recognition response times to inflected forms differ from derived forms?

4. Will prefixed words take longer to recognize than suffixed words?

Method

144 words and nonwords were recorded on one channel of a tape, and a tone corresponding to the beginning of the stimulus was recorded on the other channel which began a reaction time counter. The interstimulus interval was 4 seconds. 20 Subjects wearing headphones heard the channel with the stimuli binaurally, decided if a given stimulus was a word, and pressed a telegraph key marked "yes" or "no" which stopped the timer. The speaker was a male native Californian.

Stimuli

The words formed the following categories (see Appendix):

Words		Example
1. Suffixed (24)	inflected (12)	chosen
	derived (12)	winner
2. Pseudosuffixed (12)	"inflected" (3)	garden
	"derived" (9)	temper
3. Prefixed (12)		insane
4. Pseudoprefixed (12)		invite
5. Control Monomorphemic (12)		rabbit

There was an equal number of words and nonwords. All words were matched for spoken word frequency (Dahl, 1979), number of syllables, and number of phonemes. Although the words were not matched for written word frequency, there was no significant difference in written frequency between word categories ($F(5,66)=.92$, $p>.1$) using Carroll, Davies, and Richman (1971). A pseudoaffixed word was defined as a word containing a phonological string which resembles an affix but is not an affix (e.g. temper) or a word that contains a historical affix which carries no semantic weight (e.g. imply). The pseudosuffixed words are generally of the former type and the pseudoprefixed of the latter. The stems of all affixed forms were free. Syllable structure was controlled to the extent that (except prestige) no word (or nonword) began with a consonant cluster. The stress pattern was held constant within a word category. It was impossible to control for stress placement across categories because prefixed (and pseudoprefixed) words almost always have second syllable stress. Cutler and Clifton (1984) have found that the reaction time difference between words with first and second syllable stress is due to acoustic length, and the effect of word length can be compensated for by using regression techniques.

The nonwords were constructed by changing one or two phonemes of a real word. The number of nonwords with first or second syllable stress was the same as for the words. Nonwords did not contain affixes.

Subjects

20 unpaid volunteers from undergraduate linguistics courses at UCLA participated in the experiment. Half of the subjects heard the stimuli in one order and half heard the stimuli in the reverse order.

Procedure

There were 18 initial practice items followed by the actual (randomized) list of stimuli with no break between the practice stimuli and the test stimuli.

Results

A one-way covariate analysis[1] was conducted with word length as the covariate. Reaction time was significantly correlated with word length ($r=.22$; $p<.001$), and there was a significant difference between the word categories ($F(5,1297)=2.49$; $p<.03$). Mean reaction times for each category are presented in Figure 1.

Bimorphemic vs. Monomorphemic Words

Planned Comparison analyses revealed no significant difference between bimorphemic and monomorphemic words ($F(1,867)=.11$, n.s.). Pseudoaffixed words were not included in this analysis although they are monomorphemic. The mean reaction time for monomorphemic words was 906 msec[2]; and the mean reaction time for bimorphemic words was 911 msec.

Pseudoaffixed vs. True Affixed Words

Pseudoaffixed words were **NOT** recognized more slowly than true affixed words but were in fact recognized more quickly ($F(1,1085)=4.49$; $p=.03$). The mean reaction times were 887 msec for pseudoaffixed words and 911 msec for true affixed words.

If the pseudoprefixed and pseudosuffixed results are considered separately, we find no difference between pseudoprefixed and true prefixed words ($F(1,435)=.14$, n.s.) and pseudosuffixed words are recognized faster than true suffixed words ($F(1,647)=11.99$, $p<.001$).

Inflected vs. Derived Words

There was no significant difference in reaction time between inflected and derived words ($F(1,433)=1.72$, n.s.). The mean reaction time for inflected words was 920 msec and 897 msec for derived words.

Prefixed vs. Suffixed Words

No significant difference in reaction time was found between prefixed words ($\overline{x}=919$ msec) and suffixed words ($\overline{x}=911$ msec) ($F(1,652)=.27$; n.s).

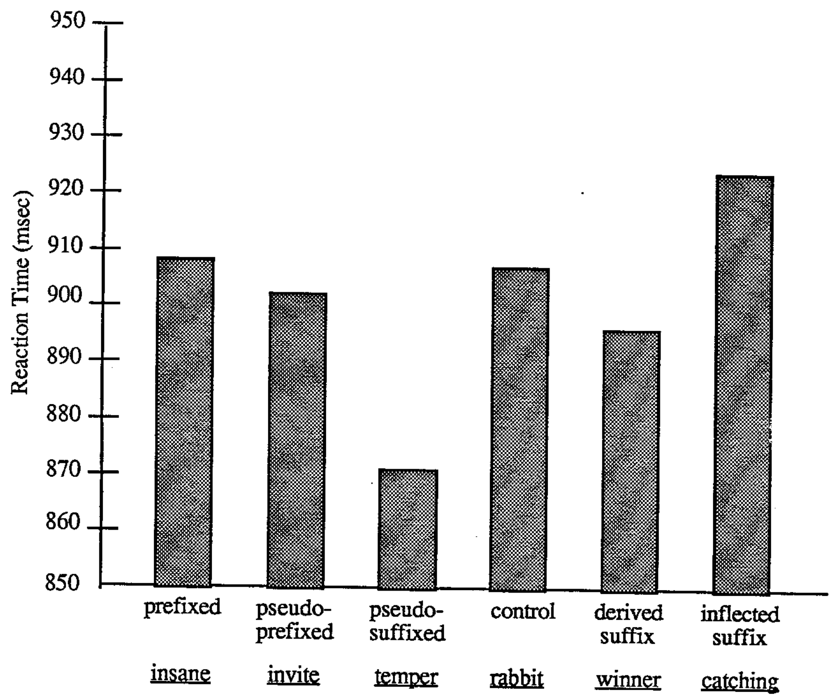


Figure 1. Mean reaction time measured from word onset.

Discussion

The results from the pseudoaffixed words will be discussed first. Recall that Taft and Forster (1975) hypothesized that any string resembling a prefix is stripped off prior to lexical access; thus, pseudoprefixed words (e.g. invite) should take longer to recognize than true prefixed or monomorphemic words because the "prefix" is erroneously stripped off. When no root is found (vite is not a root) a second search must occur for the recombined form causing a delay in recognition time. Our results do not support this serial search process -- the recognition of pseudoprefixed words was not delayed in comparison to true prefixed words or control (monomorphemic) words.

However, Taft and Forster (1975) and Taft (1979; 1981) considered historically prefixed words as prefixed words in their experiments. No distinction was made between words whose morphological analysis resulted in syntactic or semantic information relevant to the interpretation of the word and those words whose morphological structure was an historical relic. For example, Taft (1981) considered deprive and replica as prefixed bimorphemic words which were compared to devout and regime which were considered pseudoprefixed monomorphemic words.

The pseudoprefixed words used in our experiment are all historically prefixed (Klein, 1966), but of the pseudosuffixed words only motive is historically suffixed. If we assume that historically prefixed words are stored in the lexicon by their stem, then our results do not disconfirm the Affix Stripping model. Instead, the results suggest that neither semantic compositionality nor the boundedness of the root effect word recognition time. No difference in lexical decision time was observed between semantically compositional prefixed words with free roots (e.g. insane) and semantically noncompositional prefixed words with bound roots (e.g. invite).

However, if word recognition mechanisms do indeed distinguish between historical prefixes and phonological strings that simply resemble prefixes as the Taft (1981) results suggest, how does the child come to store the stems of historical prefixes in the lexicon but not "pseudostems" of words that sound/look as if they begin with prefixes? That is, how does the speaker/hearer come to store prive (from deprive) in the lexicon but not vout (from devout)?

As for the pseudosuffixed words, these were recognized faster than any other word category. The Affix Stripping model predicts that these words should be recognized more slowly than both control (monomorphemic) words and true suffixed words. On the other hand, a model containing no affix stripping predicts no difference between pseudosuffixed words and control monomorphemic words. In fact, the largest reaction time difference was between these words.

The cohort model for auditory word recognition proposed by Marslen-Wilson and colleagues (Marslen-Wilson, 1980; Marslen-Wilson and Tyler, 1980; Marslen-Wilson and Welsh, 1978) may provide an explanation for this result. According to the cohort model, word recognition is achieved in the following manner: the first one or two phonemes serve to "activate" all words in the listener's lexicon which begin with that initial sequence, and these words form the "word-initial cohort". A word is recognized (in isolation) by matching the word candidates in this cohort with the incoming sensory information. When a mismatch occurs between the sensory input and a word candidate, that word drops out of the cohort. This matching process continues until only one candidate remains which is consistent with the sensory input. At this point the listener is claimed to have recognized the word. Thus, according to this model, a word can be "recognized" even before the end of the word is heard if it is uniquely defined by the sensory information prior to the end of the word. For example, dwindle can be recognized when the /I/ is heard since at this point it is uniquely distinguished from dwell and all other English words.

It is possible that the uniqueness point for the pseudosuffixed words may have occurred earlier in the word than for the control words. If this is the case, then the pseudosuffixed words might have been recognized faster not because of some factor in their morphological structure, but because they are distinguished from their cohort earlier than the control words. This is not a property of pseudosuffixed words in general, but may have happened by chance in these stimuli.

To investigate this possibility, the uniqueness point was estimated using Kenyon and Knott (1953). The duration from the beginning of the word to the beginning of the phoneme which distinguished the word from the rest of the cohort was then subtracted from subjects' total reaction time. In this manner, reaction time from when the word became unique could be determined. For example, lobster becomes unique at /s/ where it is distinguished from lobby and lob; the duration from the beginning of the word to /s/ was subtracted from the reaction time of each subject. For suffixed words, the uniqueness point was determined for the stem. For example, the stem of jumpy becomes unique at the /p/ where it is distinguished from jumbo and jumble. Tyler and Wessels (1983) present evidence suggesting that only base forms are included in the cohort set used for recognition, therefore, we did not consider jumpy to become unique at /i/ when it is distinguished from jump. Furthermore, if suffixed words are considered to become unique when they are distinguished from the stem (i.e. jumpy from jump), then all other words must become unique word finally where the stem is distinguished from possible suffixed forms, e.g. insane from insanely or rabbit from rabbits.

In addition, to discover whether the remaining duration of the word after the uniqueness point was a factor, a correlation between reaction time and remaining duration was performed.

Results of Uniqueness Point Analysis

Reaction time was significantly correlated with remaining word length ($r=.40$; $p<.001$). A one-way covariate analysis with remaining duration as the covariate showed a significant difference between word categories ($F(5,1297)=10.13$; $p=.001$). Mean reaction times for each category are shown in Figure 2. The results suggest that the reaction time difference between the control words and the pseudosuffixed words (both of which are monomorphemic) was due to the pseudosuffixed words becoming unique earlier than the control words. Pseudosuffixed words were distinguishable an average of 71 msec earlier than control words. When this factor is adjusted by measuring reaction time from the uniqueness point the difference in recognition time disappears.

Furthermore, when reaction time is measured from the uniqueness point, several effects of word internal structure were observed:

Suffixed Words

Suffixed words took longer to recognize than both control words ($F(1,649)=16.57$; $p<.001$) and pseudosuffixed words ($F(1,647)=6.27$; $p<.01$). Again, no lexical decision time difference was found between words with inflectional and derivational suffixes.

Prefixed Words

Both prefixed and pseudoprefixed words were recognized

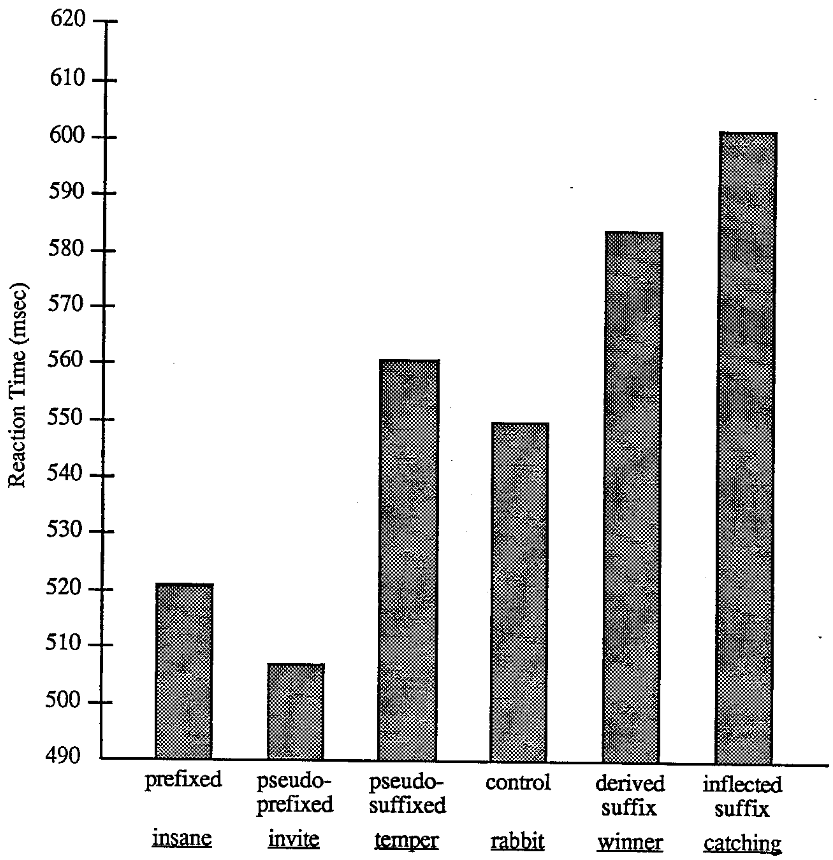


Figure 2. Mean reaction time measured from the uniqueness point.

FASTER than control words ($F(1,651)=9.37$; $p<.002$). In fact, prefixed and pseudoprefixed words were recognized faster than all other word categories. Again, no reaction time difference was observed between the prefixed and pseudoprefixed words themselves.

Discussion and Conclusion

The fact that pseudoaffixed words were not recognized more slowly than monomorphemic words (under either analysis) is evidence against the Affix Stripping model as proposed by Taft and Forster (1975) and Taft (1979). Our results using an auditory lexical decision task agree with those found by investigators using visually presented words (Henderson et al., 1984). Henderson et al. also found no difference between pseudoaffixed words and control words and proposed that the Affix Stripping model be revised such that both the unanalyzed form (e.g. massive; motive) and the stem of the decomposed form (e.g. mass; mot) are searched for in parallel. The earlier model assumed that a serial search occurred first for the stem, and if this search failed the entire word was searched for in the lexicon. Parallel processing predicts no difference between pseudoaffixed and control monomorphemic words. If a parallel search occurs, one analysis may fail (e.g. mot+ive), but the other (whole word) search will succeed (e.g. motive), and no processing cost for misanalysis will be incurred.

If lexical access for bimorphemic words normally occurs using an affix stripping procedure, these words should take longer to recognize than monomorphemic words -- this was in fact observed, at least for suffixed words. Again, we are assuming that the decomposition of a complex word into the stem and affix prior to or during lexical access is a slower procedure than the immediate search for the whole word. Caramazza, Miceli, Silveri, and Laudanna (1985) have made a similar assumption for their model of visual word recognition. Both Henderson et al. and Caramazza et al. propose that if visually presented words are parsed into their morpheme components, this analysis occurs *prior* to accessing an orthographic lexicon containing morphologically decomposed forms. We present evidence here that if *spoken* words are analyzed morphologically, this analysis occurs *during* access to a phonologically organized lexicon.

Models of auditory speech perception have only recently addressed how morphologically complex words are stored and accessed in the lexicon. Tyler and Wessels (1983) hypothesize that "words are represented in the mental lexicon as base morphemes with inflectional and derivational markers attached to them" (p. 418). Within the cohort model of lexical access, we hypothesize here that once the stem of a suffixed word is recognized, the suffixes stored with the stem may be accessed and matched with the incoming sensory information. This additional matching procedure may delay recognition for suffixed words compared to monomorphemic words. Another possibility is that suffixes are not stored with the stems but are stored in a separate component and are recognized after the stem has been identified. These two possible representations are illustrated in Figure 3:

Suffixes stored with the stem

catch
    ing
    er
    s
rabbit
    s
lobster
    s

Suffixes stored in a separate component

catch rabbit lobster	ing er s

Figure 3. Two possible models of the lexicon.

If only stems are contained in the cohort set used for recognition, then suffixes must be accessed and recognized after the stem is identified.

For monomorphemic words (e.g. *lobster*, *rabbit*), the word is potentially recognizable at the uniqueness point (e.g. at /s/ for *lobster*), and the access mechanism merely matches the remainder of the word with the incoming stimulus. Note that reaction time was not constant from the point at which the word could be recognized but varied with the remaining duration of the word. Apparently subjects waited to confirm that the stimulus was in fact the word they expected. This result is not surprising in light of the requirements of the lexical decision task. Words may "become nonwords" with a change in the final segments of the word. When subjects were asked about their false-positive errors, they often replied that they expected a particular word and had decided too quickly. Reaction time from the uniqueness point was correlated with remaining word duration because subjects may have waited to confirm the rest of the word before making their lexical decision.

We suggest that lexical decision times for suffixed words were longer than for the monomorphemic control words and pseudo-suffixed words because after the uniqueness point monomorphemic words only need to be confirmed with the incoming stimulus. That is, lobster is recognizable at /s/ and subjects simply match the remaining incoming stimulus /tər/ with the expected stored representation. For suffixed words, however, after the stem is identified the suffix must then be recognized and accessed. The suffix may be -ing, -y, -s, -er, etc. and is not predictable in isolation. Recognition of the suffix is a separate process and takes longer than simply confirming the expected word.

Suffixes may be stored with the stem as suggested by Tyler and Wessels (1983) or in a separate component in the lexicon. The latter proposal finds more support in the literature and is favored here. Based on evidence from speech errors, aphasia, and psycholinguistic data, several investigators have proposed that affixes are stored separately in the lexicon (Bradley, Garrett, and Zurif, 1980; Garrett, 1980; Caramazza et al. 1985). Some computational models of the lexicon also contain separate sub-lexicons for suffixes and stems (Karlsson and Koskenniemi, 1985; Hankamer, 1986).

Furthermore, we suggest that for auditory word recognition, access to the suffix component is constrained by the temporal order of speech perception. A stem must be found in the stem component before access to the suffix component is possible. This procedure prevents misanalysis of some pseudosuffixed words because the stem (i.e. the pseudosuffixed word itself) is encountered first, preventing access to the suffix component. For example, the entire stem heaven is initially found in the lexicon which prevents the misanalysis heav+en. No stem heav ([hɛv]) exists which might cause the parsing mechanism to search for a suffix. However, of the 12 pseudosuffixed words 8 contained an initial string that corresponds to a word stem (e.g. lob in lobby and lobster). The parser may be mislead by a "false" stem and attempt to locate a suffix, but should this misparse cause a delay in recognition? If analyses were attempted serially, one might predict a delay -- for example, lob in lobster can be parsed as a stem, and when no appropriate suffix is heard, the incoming stimulus may be then reanalyzed as a monomorphemic stem.

However, the recognition of pseudoaffixed words was not delayed in comparison to control monomorphemic words which is predicted by such an analysis. Therefore, we propose, along the lines of Henderson et al., that parallel analyses are conducted, one fails (lob+ster) and one succeeds (lobster). The search conducted in the stem component succeeds, but the parsing analysis fails. For true suffixed words only the (slower) parsing analysis succeeds by accessing the stem *jump* in the stem component and subsequently identifying the suffix *y*. The parallel search in the stem component for a stem "jumpy" fails because *jump* not *jumpy* is stored there. Pseudosuffixed words and control monomorphemic words can both be recognized in the stem component when they are distinguished from other possible stems; thus, when reaction time between these words is measured from uniqueness point no difference in recognition time is observed. Lexical decision times to suffixed words are longer because suffixed words must be recognized by the slower parsing procedure. The stem must first be identified, and then the suffix can be accessed and recognized.

Finally, although prefixed words are bimorphemic, response times to both prefixed and pseudoprefixed words were faster than for any other word category. Subjects may have been biased to respond "word" when they recognized the prefix. None of the nonwords contained prefixes, thus when the subject heard a prefix he/she may have been more biased to respond "word" compared to the other word types. This advantage for the prefixed words may have been masked when the uniqueness point of the word was not considered. Prefixed (and pseudoprefixed) words became unique later than other word categories (by about 75 msec), and this delay may have obscured the bias to decide that the item was a word. We hypothesized in the introduction that prefixed words might be recognized more *slowly* because their stems may have to be accessed first, thus *slowing* the identification of prefixed words. The potential identification of the prefixed words used here *was* delayed by about 75 msec, but this was compensated for by a lexical decision strategy. The data suggest that prefixes must have been recognized before the entire word was heard in order to bias the response. Apparently, subjects did not wait until the end of the word before analyzing it morphologically. However, the data presented here do not allow us to choose between a model of the lexicon which stores prefixed words separately from their stems (with their morphological structure marked in some way, e.g. *in+sane*) and a model which contains a prefix stripping process *prior* to access of the stem in the lexicon. In both models, prefixes can be recognized early and thus bias the subject to respond "word".

Additional support for the hypothesis that listeners do not wait until the end of the word to begin the recognition process is found in the correlation data from reaction time and word duration. Reaction time was positively correlated with word duration when measured from word onset ($r=.22$, $p<.001$) indicating that the longer the word, the longer the decision time. However, reaction time was *negatively* correlated with word duration when measured from word offset ($r=-.52$, $p<.001$). Jarvella and Meijers (1983) found the same pattern of correlations for Dutch. Listeners appear to begin processing words before their offset since for longer words they have more time to process the word, thus when recognition time is measured from word offset longer words produce faster decision times.

In summary, the results suggest a model of auditory word recognition which parses the morphological structure of words online. Lexical decisions to suffixed words were longer than monomorphemic words when reaction time was measured from the uniqueness point of the word. Under this analysis, what is essentially being measured (in addition to decision time) is the time to recognize the suffix in suffixed words compared to the time to match the end of the expected word with the remaining stimulus for monomorphemic words. However, some suffixed words became unique at the suffix, for example the stem harm in harmful cannot be uniquely identified at the stem, but at /f/ where it is distinguished from harmony. Because the uniqueness point for the stem of suffixed words varied, another experiment is currently underway in which the stems of all suffixed words become unique prior to the suffix (e.g. jump in jumpy becomes unique at /p/ where it is distinguished from jumbo and jumble). With this control we can be sure that recognition time for the suffix is measured because the stem is at least potentially recognizable prior to the suffix. Preliminary results (N=16) support our findings here -- lexical decision times are longer for suffixed words than for monomorphemic words. We propose that the accessing and recognition of the suffix delays lexical decision time compared to monomorphemic words which do not require this additional process. Evidence was also found suggesting the internal structure of prefixed words is recognized during lexical access as well. Subjects may have been able to recognize prefixes before the word offset and use this knowledge to bias their lexical decision. There was no evidence suggesting that the parsing procedure was mislead by pseudoaffixed words. We hypothesized that if misparses occur during auditory word recognition, they occur in parallel with the correct parse.

Finally, a caveat is required for the model of morphological analysis proposed here. We have offered no account of how non-concatenative morphology is parsed, accessed, or represented. The model tentatively accounts for how suffixed words are stored and accessed, and two possible access procedures for prefixed words were presented. However, there is currently no mechanism that deals with allomorphy or suffixation that effects the phonological form of the stem (e.g. trisyllabic shortening in English or vowel harmony in other languages). However, the model is at least a first attempt at addressing how morphological structure is parsed and represented within the mental lexicon.

Notes

[1]The conventional F statistic was employed here to avoid the substantial negative bias associated with quasi-F tests. Our materials were matched on a number of relevant dimensions (i.e. well balanced), and Wickens and Keppel (1983) have shown that balancing materials "is very effective in reducing the positive bias associated with a [fixed-effect] F_1 statistic" (p. 307).

[2]Means given in the planned comparison analyses vary slightly from the means given in the overall F because of the different

covariate duration adjustments.

APPENDIX

The words used in the experiment reported here are listed below:

Inflected	Derived	Prefixed
wasting	weakness	unfair
joking	vaguely	dislike
gambling	worthless	nonsense
curtains	safer	unsure
dining	shorter	insane
matches	jumpy	unlike
cookies	harmful	replace
tended	peaceful	unreal
lasted	filthy	renew
catching	winner	mislead
loosen	massive	distrust
chosen	thankful	unclear

Control Monomorphemic	Pseudosuffixed	Pseudoprefixed
rabbit	monster	invite
ticket	temper	retreat
victim	garden	reprieve
limit	bitter	release
soda	lobby	device
garbage	motive	reward
cancel	hunger	expense
campus	kidney	imply
cocktail	monkey	remote
hollow	token	discrete
robot	lobster	prestige
costume	heaven	design

References

Bradley, D., Garrett, M., & Zurif, E. (1980) Syntactic deficits in Broca's aphasia. In D. Caplan (Ed) Biological Studies of Mental Processes, The MIT Press: Cambridge, Massachusetts.

Buckingham, H. (1981) Where do neologisms come from? In J. Brown (Ed.) Jargon Aphasia, Academic Press: New York.

Caramazza, A., Miceli, G., Silveri, C., & Laudanna, A. (1985) Reading mechanisms and the organization of the lexicon: evidence from acquired dyslexia. Cognitive Neuropsychology, 2(1): 81-114.

Carroll, J., Davies, P., & Richman, B. (1971) Word Frequency Book. American Heritage Publishing Co.: New York.

Coltheart, M. (1980) Reading, phonological recoding, and deep dyslexia. In M. Coltheart, K. Patterson, J. Marshall (Eds)

Deep Dyslexia, Routledge & Kegan Paul: London.

Cutler, A. and Clifton, C. (1983) The use of prosodic information in word recognition. In H. Bouma & D. Bouwhuis (eds.) Attention and Performances X, Lawrence Erlbaum Associates: New York.

Cutler, A., Hawkins, J., & Gilligan, G. (1985) The suffixing preference: A processing explanation. Linguistics, 23: 738-758.

Dahl, H. (1979) Word Frequency and Spoken American English. Verbatim: Connecticut.

Fodor, J., Bever, T., and Garrett, M. (1974) The Psychology of Language: an Introduction of Psycholinguistics and Generative Grammar. McGraw Hill: New York.

Fromkin, V. (1971) The non-anomalous nature of anomalous utterances. Language 47:27-52.

Garrett, M. (1980) Levels of processing in sentence production. In B. Butterworth (Ed.) Language Production, Vol. 1, Academic Press: New York.

Gfroeney, Guntler, Weiss, (1984) Zur Frage des visuellen und phonologischen Zugange zum mentalen Lexikon. Phonetica 41:41-49.

Goodglass, H. and Kaplan, E. (1972) The Assessment of Aphasia and Related Disorders. Lea and Febiger: Philadelphia.

Gordon, B., & Caramazza, A. (1982) Lexical decision for open and closed class items: Failure to replicate differential frequency sensitivity. Brain and Language 19: 335-345.

Hankamer, J. (1986) Finite state morphology and left-to-right phonology. Paper presented at the 15th Annual Linguistics Symposium at the University of Wisconsin, Milwaukee, April 6-8.

Henderson, L., Wallis,J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouwhuis (Eds) Attention and Performance X: Control of Language Processes Lawrence Erlbaum Associates: London.

Jarvella, R.J. & Meijers, G. (1983) Recognizing morphemes in spoken words: Some evidence for a stem organized mental lexicon. In G.B. Flores d'Arcais & R.J. Jarvella (Eds) The Process of Language Understanding John Wiley & Sons Ltd: New York.

Karlsson, F. and Koskenniemi, K. (1985) A process model of morphology and the lexicon. Folia Linguistica 19: 207-231.

Kean, M-L. (1977) The linguistic interpretation of aphasic syndromes. In E. Walker (Ed.) Exploration in the Biology of Language. Bradford Books: Montgomery, Vermont.

Kenyon, J.S. & Knott, T. (1953) A Pronouncing Dictionary of American English. G. & C. Merriam Co.: Boston.

Klein, E. (1966) A Comprehensive Etymological Dictionary of the English Language Elseviser Publishing Co.: Amsterdam.

Manelis, L. & Tharp, D.A. (1977) The processing of affixed words. Memory and Cognition. 5: 690-695.

Marslen-Wilson, W. (1980) Speech understanding as a psychological process. In J.C. Simon (Ed.) Spoken Language Generation and Understanding. Reidel: Boston.

Marslen-Wilson, W. & Welsh, A. (1978) Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology. 10:29-63.

Marslen-Wilson, W. & Tyler, L. (1981) Central processes in speech understanding. Philosophical Transactions of the Royal Society of London B 295: 317-332.

Taft, M. (1979) Recognition of affixed word and the word frequency effect. Memory and Cognition 7(4): 263-272.

Taft, M. (1981) Prefix stripping revisited. Journal of Verbal Learning and Verbal Behavior. 20: 289-297.

Taft, M. & Forster, F. (1975) Lexical storage and retrieval of prefixed words. Journal of Verbal Learning and Verbal Behavior 14:635-647.

Tyler, L. & Wessels, (1983) Quantifying contextual contributions to word recognition processes. Perception and Psychophysics 34(5): 409-420.

Wickens, T. & Keppel, G. (1983) On the choice of design and of test statistics in the analysis of experiments with sampled materials. Journal of Verbal Learning and Verbal Behavior. 22: 296-309.

Rate-dependent variability in English and Japanese complex vowel F2 transitions

William B. Dolan and Yoko Mimori

Introduction:

A great deal of phonetic research has been devoted to the search for invariance in the speech signal, based on the assumption that motor coordination of the vocal tract is geared primarily toward producing certain constant gestures which underlie acoustic constancy (Kent and Moll,1972; Fujisaki and Higuchi, 1979; MacNeilage, 1970). Once these fundamental targets are determined, other articulatory (and acoustic) facts will presumably fall out from physiological factors, timing considerations, etc. Most phonetic research aimed at discovering invariant targets has examined static postures of articulation; we are concerned with articulatory dynamics. The control and production of diphthongs constitutes a relatively simple type of dynamic articulation.

Gay (1968), examined variation in the production of American English dipthongs at three different rates of speed; "slow", "moderate", and "fast". An earlier synthetic-speech perceptual study (Gay,1967) had suggested that the rate of F2 change, rather than absolute F2 frequency values, served as the primary auditory cue for diphthong identity. In his production experiment, Gay concentrated on the behavior of the first and second formants in tokens of the five English diphthongs /ai/, /ou/, /ei/, /au/, and /ɔi/ (hereafter written as /oi/) uttered by 5 speakers of New York English. Mirroring the results of the perception experiment, his findings supported the claim that the F2 transition rate is the cue which speakers must crucially control during diphthong production. Gay found that both onset frequency and the rate of F2 transition (in Hz/milliseconds) remained constant, uninfluenced by overall speech rate. Transition length, steady-state durations, and F2 offset frequency, on the other hand, all proved variable as speech rate changed. A similar claim for constancy of F2 across rate changes has been made for Argentine Spanish (Manrique,1979).

Gay views English diphthongs as separable into components, with duration-sensitive steady-states framing a fixed-rate transition portion. As durations increase, speakers fill time by either adding steady-state vowels or increasing the length of the transition. Transition offsets, Gay claims, are not fixed--shorter diphthongs will have shorter transitions which never reach their potential target values.

Such a model suggests that the task of adjusting the duration of an diphthong articulation is a relatively simple one for the English speaker, accomplished by shortening or lengthening a transition which has a predetermined rate. Similarly, the presence or absence of "prefabricated" steady-state sections at the edges of this transition is determined by the length of time available for the articulation. Diphthongs are treated by this approach as fixed V1-Transition-V2 sequences which are clipped at the edges rather then compressed as speech rate is increased. This production model thus supports a view of diphthongs as unit phonemes, rather than sequences of two vowel-like articulations connected by an incidental transition. According to Gay's view, duration modification is not a process of squeezing or stretching a transition between fixed-frequency endpoints; what is traditionally written /ei/, for example, may contain neither a prototypical /e/ nor /i/ portion. Instead, the transition itself is the fundamental feature of the diphthong, with steady-state portions (when they exist) and onset/offset frequencies merely falling out from the lengthening formula. Partial support for such a motor planning model for diphthongs is provided by cinefluorographic data for English (Kent and Moll, 1972) which indicates that initial and terminal tongue positions are not fixed characteristics of diphthong production.

The Problem:

If the relationship between fast- and slow-speed utterances of a diphthong is really this direct and simple, then neural programming models for articulation can be expressed relatively easily. Vocal tract motor planning could be handled in much the same way that many speech synthesizers organize production, with short, fixed units being added to the basic transition to produce a sound of the necessary length. At the other extreme is a segment-based model, in which the transition is simply derived to connect the targets of two steady-state vowels, with the actual slope of the transition being dependent upon the target values and the length of time available for the transition. Unlike the incremental model suggested by Gay's results, any model involving such duration-based distortions of formant transition slopes would lead us to conclude that speakers routinely reorganize their motor-coordination plans for particular utterances depending on factors like speech rate.

Such a conclusion would not be particularly surprising; a great deal of neurophysiological evidence suggests that speech and other complex neural behaviors are not produced simply by concatenating small preprogrammed neural commands chosen from limited repetoire (MacNeilage, 1970). If neural reorganization is the norm for each new production of a complex vowel, it is likely that this is governed by certain universal neurophysisological principles. One such principle, which has been proposed in the articulatory domain, is "the further the faster" (Kent and Moll, 1972), which directly links speed of articulator movement to the magnitude of the articulatory gesture.

Can the acoustic behavior of English diphthongs at different speech rates be explained in terms of this general constraint on speech production, with the velocity of the transition increasing as the distance traversed by F2 on the frequency scale does? Or must these facts also be treated as "phonologized" features of particular languages, with the rate of the transition remaining constant across different speeds and thus serving as a primary cue to diphthong identity? If it is assumed that F2 slope/offset/onset shapes (and, presumably, the articulatory gestures underlying these formant movements) are language-specific features of complex vowels, then we would expect these shapes to be completely idiosyncratic across languages and diphthong types.

Problems for such a model, however, are presented by data suggesting that transition slopes may be dependent on such external factors as speech rate. Ren (forthcoming) finds substantial variation in Mandarin diphthong slopes across rate changes, variation which he attributes to the influence of several factors: speech tempo, duration of steady-states, and offset target (which varies with rate). These results suggest that Gay's finding that F2 slope is invariant across rate changes in English is not a general phonetic principle governing diphthong production, but rather a parameter set within a given language. Lindau-Webb et al. (1985) suggest that transition duration is another parameter which is subject to non-random variation. Their examination of /ai/ and /au/ in Hausa, Arabic, Chinese, and English indicates that transition durations for some diphthongs in some languages follow the principle "the further to go [on the frequency scale] the longer it takes".

The Experiment:

The following is an investigation into the claims of these competing models of complex vowel production. Is the notion of the transition as primary target the correct one, or would it be preferable to model English diphthongs as sequences of two vowel targets connected by a transition whose properties fall out from the endpoint targets and timing factors? Can rate-dependent variability in diphthong production be ascribed to general phonetic principles which operate across all languages?

Our experiment is in part an attempt to replicate Gay's (1968) results for English diphthongs; but in order to address the question of cross-linguistic validity, Japanese vowel sequences were also examined. Although both English diphthongs and Japanese vowel sequences involve articulator movement between two vowel targets, the two are phonologically very different, and for this reason Japanese was chosen as an interesting counterpart to English. While English diphthongs are generally treated as tautosyllabic combinations of vowel plus a glide, Japanese vowel sequences are "non-unitary", two distinct elements which constitute two "morae" linked by a glide. Any of the five Japanese vowels can be combined with itself or any other vowel to form one of these sequences, which can then be produced with one of two pitch accent patterns, /V'V/ (pitch accent on the first syllable) or /VV/ ("flat-type word accent").

Procedures:

Word lists and speakers:

An English word list was devised, consisting of CDC syllables, where "D" was one of the five American English diphthongs /oi/, /ai/, /ou/, /au/, /ei/. Alveolar stops and /h/ were chosen as framing consonants for the diphthongs, since these tend to have less effect on formant transitions than other consonants. A complete set of tokens with the shapes /hDd/ and /dDd/ provided a standardized environment for the diphthongs. Most of these were real English words, but the consonant frame necessitated the inclusion of a number of nonsense words-- /doid/, for example. Each word was embedded in the frame sentence "Please say ______ again" in order to provide a constant, non-final stressed environment for each token. Ten male speakers of California English (all from either the Los Angeles or San Francisco area) were recorded reading this list.

The Japanese word list consisted of four diphthongs, /oi/, /ai/, /au/, /ui/ embedded in the words /hai/, /kau/, /toi/, /tsui/. Each token was read with two accent patterns, /V'V/ and /VV/, providing a total of total of eight. All were meaningful Japanese words. Tokens were read in the carrier sentence "kore wa _____ to iimasu" ("this is said CVVC"). Eight male speakers of the Tokyo dialect of Japanese were recorded reading this list.

All speakers were recorded using a high-quality reel-to-reel tape recorder and a sound booth. Each speaker was asked to read through the word list at three different rates--"slow", "normal", and "fast". Before being recorded, speakers were asked to read through the list and familiarize themselves with the items, especially the nonsense words. The "normal" speed reading was elicited first, in order to provide speakers with their own benchmark for deciding what appropriate "fast" and "slow" rates should be. This set was repeated twice, with the first reading serving as a trial run. Instructions regarding speeds were kept deliberately vague, with speakers being told to produce what was for them a "fast" or "slow" but still natural rendition of the word list.

Speakers were told to be as consistent as possible in sentential intonation, stress, and speed while reading through the list. Any mistakes they noticed as they spoke were to be simply repeated correctly, without breaking the rhythm. Corrected versions of other errors were elicited after each set. Rest periods were allowed between sets, and if desired, between tokens (particularly helpful at very fast speeds.) Individuals varied widely in their ability to produce three different rates; for some, the task seemed alien and difficult, while others exhibited excellent control over speed. In general, it seemed easier to elicit slower speech from a normally fast talker than vice versa. Several versions of each rate were elicited, in order to provide a choice for the experimenter. Speakers were re-recorded when sample spectrographic checks of durations indicated that they had not spoken quickly or slowly enough to distinguish the extreme rates from the "normal" reading.

<u>Measurements</u>

Wide-band digital spectrograms with a 0-5000Hz frequency range were made for each token. In cases where a speaker's "fast" vs. "normal" or "normal" vs. "slow" tokens for a single word proved durationally indistinct, a different token of that word was selected from one of the extra renditions. Linear Predictive Coefficient (LPC) analysis was performed on each token using a signal processing program (WAVES) on the UCLA Phonetics Lab PDP-11computer. Sampling at 10,000 Hz and low-pass filtering at 4,500 Hz, formant frequencies and bandwidth values were calculated over a 25.6 millisecond window at 10 millisecond intervals. A plot of F1-F4 formant frequencies at 10-msec intervals was printed for each token, as illustrated in **Figure 1**. Using these formant plots as a guide, the file of formant values for each token was edited by hand to remove initial and final consonant closures as well as initial /h/. Where the LPC program occasionally failed to correctly identify particular formants, appropriate values were interpolated between the values of the preceding and following points. Such interpolations were restricted to cases where the course of the formant was otherwise unambiguous, with the artifacts representing clear mistakes made by the formant-tracking program. The spectrograms also served as useful checks in this task. In no instance were more than 2 consecutive values interpolated; tokens with a great many missed poles were simply replaced with another token from the same speaker. Data for one English speaker was judged unuseable, since the formant-tracking program failed to consistently pick out his F2; data from a second English speaker was discarded because many of his slow/normal/fast rate tokens proved durationally indistinct.

The edited formant files were then run through a specially-written computer program which determined measurements for each F2 transition. This slope-measuring program was used to objectively identify steady-state vs. transitional portions of a complex vowel, as well as to calculate the rate of change during each transition. For English, the transition was defined as starting at the first window in which the frequency of F2 differed by more than 15 Hz from the frequency of F2 in the previous window, which had its center 10 milliseconds earlier. As long as the F2 varied less than this 15 Hz figure, the program identified those portions as belonging to a steady-state portion of the vowel. In deciding how to best select the transition, the 15 Hz figure was chosen somewhat arbitrarily --insisting on at least a 15 Hz change prevented small glitches in the F2 from being selected as onsets/offsets, while a larger number caused the measuring program to pass over what seemed to the operator to be clear transition boundaries. After some experimentation with the program, a 20 Hz figure was chosen to define Japanese transition boundaries.

Figure 1. Plot of formant structure (F1-F4) for Japanese token /hai/. Note irregular formant structure during the voiceless /h/ portion.

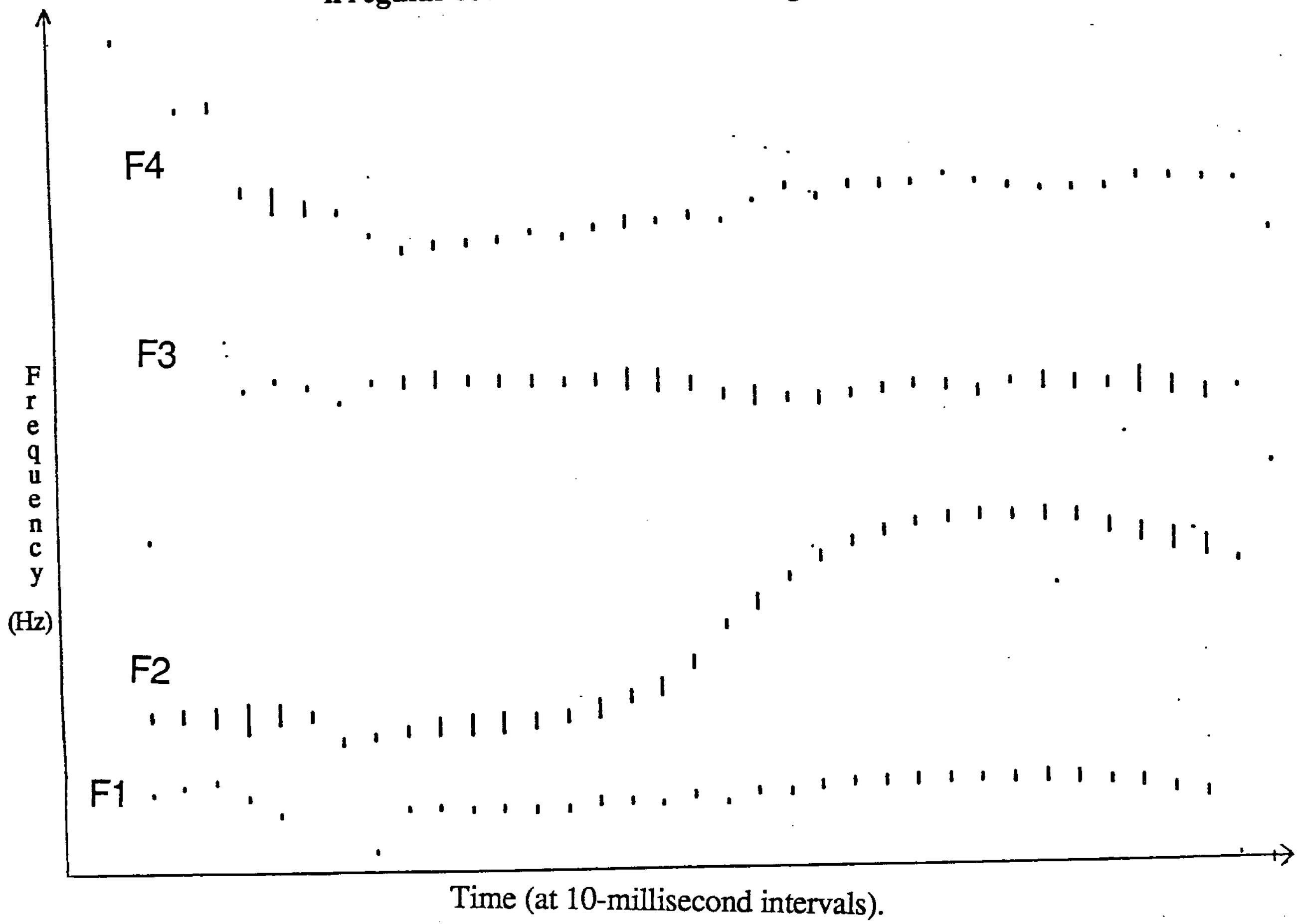
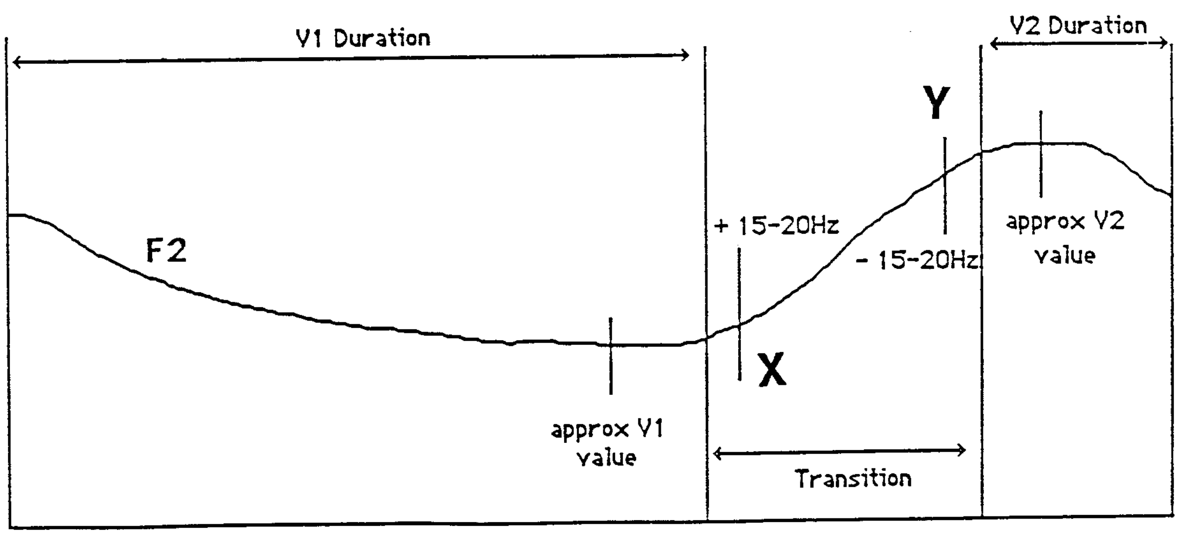


Figure 2. Idealized F2 curve for English token /daid/, illustrating the measuring algorithm's treatment of initial and final consonant transitions.



The algorithm described above provides a good way of defining the start of the transition in many CDC/CVVC sequences, but it is inadequate if there is a large F2 transition associated with the initial consonant. For example, in the token /doid/, F2 typically falls sharply after the release of the /d/ until an appropriate value for the /o/ part of the diphthong is reached. The upward movement for the transition then begins, resulting in a U- or S-shaped curve. **Figure 2** illustrates how the formant-measuring program treated such cases. In order to prevent the algorithm from picking the consonant-vowel transition as the actual diphthong transition, the program types out F1 and F2 values for each point during the first 200 ms of each vowel, and asks the user to decide whether a significant initial consonant transition exists. If so, it requests an appropriate F2 value for the initial steady-state (ie, approximate V1 value): for both /ai/ and /oi/, such decisions are quite straightforward, with the highpoint of F1 and lowpoint of F2 signaling the first part of the diphthong or vowel sequence. The V1 portion for /ui/ normally coincided with the lowpoints of both the first and second formants. For /au/, /ei/, /ou/, this point was normally defined by the course of F1, which rises out of the transition before beginning to fall for the diphthong. This early peak coincides with the vowel starting point. Again, the formant printouts were used as a guide in determining appropriate values. Segmentation of the initial consonant from the vowel in the Japanese token /tsui/ was made quite difficult by the tendency of speakers to devoice the /u/ portion of the vowel sequence after /ts/. In these cases, measurements were based on examination of the corresponding spectrogram as well as cues in the formant printouts such as bumps in the F3 and F4 signifying the onset of the vowel.

Given an approximate value, the algorithm treats all following points as potential transition onsets, moving forward from this approximate V1 value until a 15/20Hz change across 10 milliseconds is found. **(Point X on Figure 2)** If no such change is found, ie, when the formant movement is very gradual, the program chooses as the onset the time-point whose F2 value is closest to the approximate V1 value supplied by the operator. Measurements for the end of the transition were calculated as the exact mirror image of the onset. Thus the offset value represents the point preceding the last 15/20Hz change before a specified approximate V2 value **(Point Y on Figure 2)**. Again, in cases with very gradual slopes, the V2 input value itself was selected by the program as the offset.

Whatever comes before the beginning of the transition is classed as Vowel 1 steady-state, or "V1", with everything after the end of the transition being "V2". The frequency value for the second formant of V1 was defined as the value at the point immediately preceding the transition start, if there was an initial consonant transition. Initial and final consonant transitions were treated as part of the durations of V1 and V2. The frequency value for the V2 was defined as the F2 value at the transition offset point. Durations for V1, V2 and the transition were determined.

Two different measurements of the transition rate of change were determined. For the first, "maxslope", an algorithm was used to determine the mean change (in Hz/milliseconds) for all sets of four points between the starting and ending points of the transition. Maxslope was defined as the absolute value of the rate of change over that set of four points showing the maximum change, ie, that part of the transition showing maximum velocity of F2 over thirty milliseconds. The second measure, "fullslope" reflects the F2 change across the entire program-defined transition, not just the most extreme thirty milliseconds: fullslope = range/duration (that is, the absolute value of the difference in Hz between F2 onset and offset divided by the transition duration in milliseconds.)

Statistics:

Five one-tailed paired t-tests compared the entire set of English /h_d/ tokens with the set of /d_d/ tokens. Another set of five t-tests compared the two Japanese accent patterns, pooling speakers, rates, and diphthongs. For both languages, the variables examined were maxslope, fullslope, range, transition onset, and transition offset.

Correlations between for each slope measurement with F2 range were determined and plotted for the whole of each dataset, as well as for each diphthong population.

Analyses of variance were performed using the SAS package (SAS, 1985), examining the effects of the variables rate, token-type (For English only: /h_d/ vs. /d_d/), specific diphthong, and individual speaker on the maxslope and fullslope measures, as well as interactions between rate/speaker, rate/token-type, and rate/diphthong for both slope measurements.

A stepwise regression was performed on maxslope and fullslope values for English, in order to measure the relative effects of variables on the slope measurements. The variables tested were transition, V1, and V2 durations, "percentage of transition" (the percentage of the total duration consisting of transition--transition/duration * 100) and F2 range. Because reliable V1 duration measurements were unobtainable for Japanese, this statistic was performed only on the English data.

The data presented below represents tokens from eight speakers of English and seven speakers of Japanese.

Results:

T-tests

The results of four of the five one-tailed paired t-tests comparing the two English token-types, /h_d/ vs. /d_d/, proved non-significant; only F2 range differed significantly between the two sets ($t=2.16$, $p<0.0330$). Results of the five t-tests are presented in **Table 1.** Since values for F2 maxslope, fullslope, onset and offset were found to be nondistinct based on whether the token was /h_d/ or /d_d/, these datasets were pooled for further analysis.

Table 1. t-tests measuring the effect of English and Japanese token-type on the variables maxslope, fullslope, F2 onset, F2 offset, and F2 range.

	English (/h__d/ vs. /d__d/)		Japanese (/V'V/ vs. /VV/)
Maxslope	t=1.23 p<0.2203		t=-1.57 p<0.1201
Fullslope	t=-1.23 p<0.2224		t=-0.60 p<0.5494
F2 Onset	t=-0.84 p<0.4020		t=1.91 p<0.0593
F2 Offset	t=-0.60 p<0.5511		t=1.25 p<0.2154
F2 Range	t=2.16 p<0.0330		t=-1.27 p<0.2087

Table 2. Mean values and standard deviations for the five English diphthongs and four Japanese vowel sequences.

English

Measure		/ai/	/au/	/ei/	/ou/	/oi/
Maxslope (Hz/msec)	F	6.43 (2.8)	-6.59 (3.02)	2.96 (1.9)	-4.34 (1.82)	15.74 (4.16)
	N	8.22 (2.27)	-6.56 (2.15)	3.26 (1.22)	-3.3 (1.53)	15.38 (3.75)
	S	7.76 (1.93)	-5.7 (2.86)	2.72 (0.93)	-3.29 (1.86)	12.54 (2.36)
Fullslope (Hz/msec)	F	4.98 (1.8)	5.37 (2.4)	2.16 (1.36)	4.14 (1.66)	10.82 (2.61)
	N	5.27 (0.98)	4.5 (1.6)	2.25 (0.57)	2.95 (1.29)	9.51 (2.4)
	S	4.43 (1.14)	3.17 (0.98)	1.73 (0.68)	2.28 (1.99)	6.63 (1.91)
F2 Onset (Hz)	F	1443.14 (164.13)	1508.33 (134.04)	1890.08 (259.42)	1259.45 (154.24)	1022.71 (164.4)
	N	1295.17 (102.16)	1506.19 (180.74)	1897.44 (192.14)	1218.32 (87.79)	928.17 (120.33)
	S	1206.27 (128.56)	1449.61 (132.64)	1981.28 (220.81)	1214.38 (145.4)	852.65 (95.5)
F2 Offset (Hz)	F	1762.43 (174.06)	1231.7 (119.6)	1954.7 (348.52)	1151.8 (120.89)	1840.93 (205.56)
	N	1881.6 (252.62)	1127.5 (143.47)	2058.96 (304.68)	1091.6 (90.7)	1919.87 (273.44)
	S	1960.94 (290.67)	1005.68 (144.64)	2170.03 (304.68)	999.93 (114.13)	1980.18 (251.6)
F2 Range (Hz)	F	319.29 (151.38)	276.63 (131.3)	100.68 (106.23)	107.66 (65.59)	818.22 (242.67)
	N	586.43 (215.5)	378.69 (169.8)	164.61 (106.22)	126.7 (93.93)	991.7 (323.54)
	S	754.67 (213.33)	443.93 (205.51)	188.75 (136.58)	214.45 (188.85)	1127.53 (272.3)

Japanese

Measure		/ai/	/au/	/oi/	/ui/
Maxslope (Hz/msec)	F	11.63 (1.57)	-4.57 (2.09)	22.86 (4.70)	12.37 (3.41)
	N	11.37 (2.28)	-4.25 (2.01)	24.83 (5.59)	11.37 (2.62)
	S	11.45 (2.24)	-4.22 (1.48)	25.61 (6.24)	11.09 (3.56)
Fullslope (Hz/msec)	F	7.75 (2.22)	2.34 (1.33)	13.63 (3.71)	8.76 (2.58)
	N	7.46 (1.40)	3.36 (1.89)	12.09 (2.33)	7.22 (1.9)
	S	6.86 (1.16)	4.36 (3.24)	12.04 (2.85)	6.68 (1.31)
F2 Onset (Hz)	F	1508.14 (214.49)	1343.43 (97.21)	1085.58 (103.12)	1617.68 (112.2)
	N	1323.19 (147.8)	1307.42 (123.77)	984.96 (92.05)	1514.04 (54.17)
	S	1344.53 (71.43)	1273.74 (105.15)	924.96 (76.59)	1498.41 (83.43)
F2 Offset (Hz)	F	2245.14 (124.57)	1144.75 (106.07)	2179.78 (121.84)	2241.07 (112.98)
	N	2259.74 (128.86)	1094.46 (109.49)	2179.93 (135.61)	2219.71 (129.48)
	S	2242.18 (92.11)	1052.29 (77.27)	2208.75 (110.69)	2244.35 (135.90)
F2 Range (Hz)	F	744.47 (211.12)	198.68 (105.45)	1094.21 (111.79)	623.39 (70.78)
	N	936.55 (203.07)	212.95 (89.77)	1194.96 (183.24)	705.67 (145.45)
	S	897.65 (108.9)	221.45 (105.34)	1283.79 (140.14)	745.95 (162.44)

The effect of Japanese accent pattern proved nonsignificant for all five variables. Only transition onset showed any hint of an accent effect, but this was not signficant ($t=1.91$, $p<0.0593$). Tokens from the two accent types were pooled for further analysis.

Means:

Mean values and standard deviations for English and Japanese for the measures maxslope, fullslope, transition onset, transition offset, and F2 range are presented for each diphthong at each speed condition in **Table 2.**

For English, F2 range in particular showed substantial variation in means across the three rates for certain dipthongs. F2 range for /ai/, for example, varies from a "slow" 754 Hz to a "fast" value of 319 Hz, reflecting a much less extreme F2 movement at the faster rate. Mean values for maxslope, fullslope, onset and offset also exhibit apparent rate-dependent changes, although this effect varies with particular tokens and variables. F2 onset, for example, changes only 22 Hz across the "fast-slow" means for the token "hoed" (1157-1135 Hz), while the means for "hide" change 237 Hz as the rate changes (1420-1183 Hz).

On the whole, the Japanese data exhibits far less variation in means across speeds than English. This is particularly striking in the slope measurements (maxslope for /au/, for example, changes very little, from a "slow" value of -4.22 to a "fast" value of -4.57). Means for F2 range also change little with increased speaking rate: whereas the mean F2 range for English /ai/ is more than halved as speech rate increases (from "slow" 754 Hz to "fast" 319 Hz), range for the equivalent diphthong in Japanese decreases less than 20% (from "slow" 897 Hz to "fast" 744 Hz). Like English, rate-dependent variation in onset and offset values seems to vary with diphthong and token. For example, /ui/'s F2 onset rises from a "slow" value of 1498 Hz to 1617 Hz for "fast", but its offset remains almost constant (2241-2244 Hz).

Correlations:

Correlation coefficients (r) among F2 range, maxslope, and fullslope are presented for the English and Japanese datasets in **Table 3.**

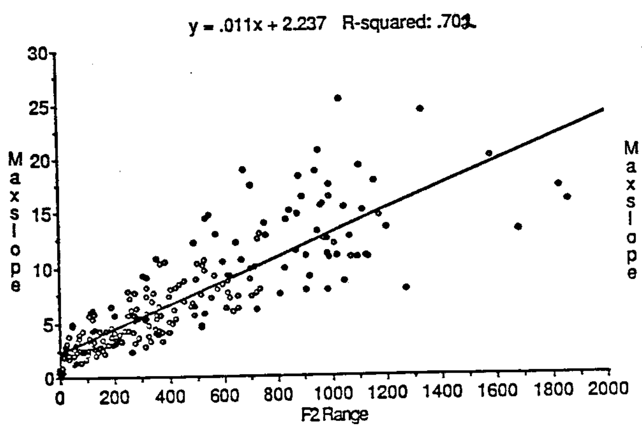
Table 3. Correlation coefficients. (All tokens within each language pooled).

	Maxslope-Range	Fullslope-Range
English	r=0.838 p< 0.0001	r=0.859 p < 0.0001
Japanese	r=0.717 p< 0.0001	r=0.778 p< 0.0001

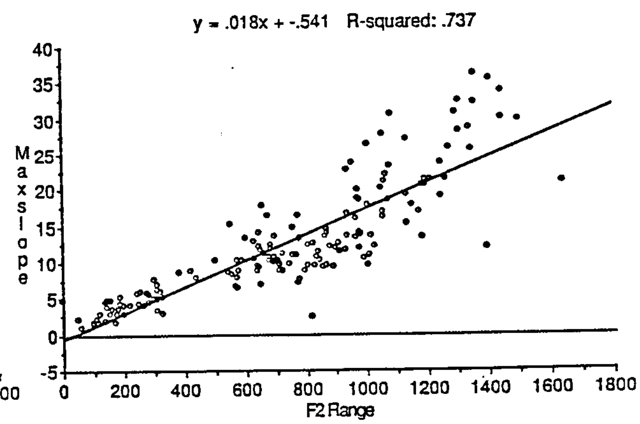
F2 range is highly correlated with both maxslope and fullslope for the five English diphthongs and four Japanese vowel sequences. Of the two slope measures, maxslope provides the better correlation in both languages (English r=0.838, p<0.0001 and Japanese r=0.859, p<0.0001; as opposed to fullslope/range values of r=0.718, p<0.0001 for English and r=0.778, p<0.0001 for Japanese.) Plots of maxslope/range and fullslope/range correlations for both datasets are presented in **Figure 3**.

Examination of each separate diphthong or vowel sequence population reveals very different degrees of correlation between maxslope/range and fullslope/range. These figures are presented in **Table 4.** Maxslope again yields consistently better correlations with F2 range than does fullslope. Fullslope and range are quite poorly correlated for English /ai/, for example (r=0.281, p<0.10), while maxslope and range for the same diphthong yield a much better correlation (r=0.681, p<0.0001). In English, the diphthongs /ai/ and /ei/ exhibit the best correlations for maxslope/range, followed by /ou/ and /au/. The diphthong /oi/ differs dramatically from the the other four, showing a distinct lack of correlation between range and either slope measurement.

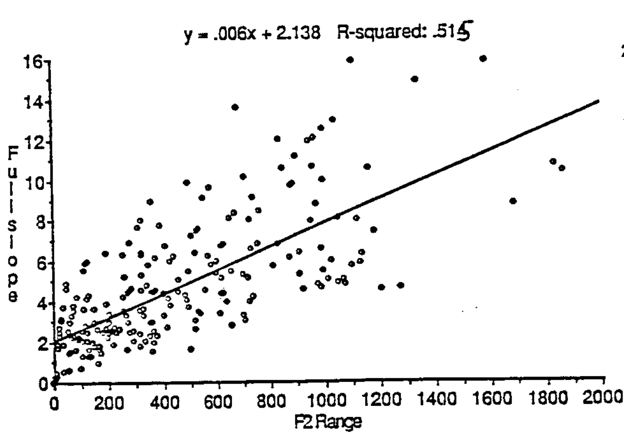
Figure 3. Plots representing the correlations between F2 range and fullslope and F2 range and maxslope for the pooled datasets of English (3a-b) and Japanese (3c-d).



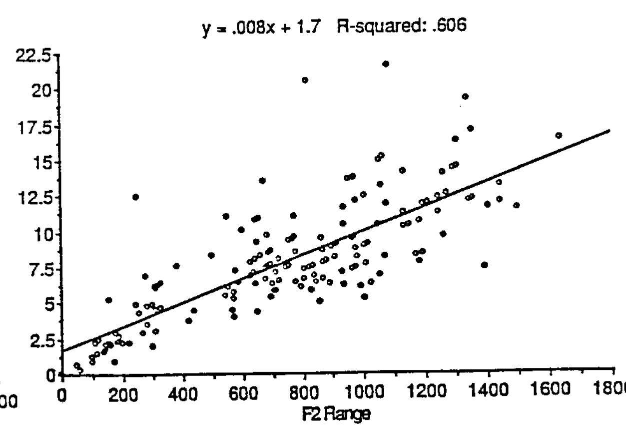
3a.



3c.



3b.



3d.

Table 4. Correlation coefficients (r) among F2 range, maxslope, and fullslope broken down by diphthong/vowel sequence for both English and Japanese.

		Maxslope-Range	Fullslope-Range
English:	/ai/	r= 0.678 p< 0.0001	r=0.281 p < 0.10
	/au/	r= 0.576 p< 0.0001	r=0.269 p < 0.10
	/oi/	r= 0.297 p < 0.05	r=0.165 p < 0.25
	/ei/	r=0.688 p< 0.0001	r=0.494 p < 0.005
	/ou/	r= 0.421 p < 0.01	r=0.588 p < 0.005

		Maxslope-Range	Fullslope-Range
Japanese	/ai/	r=0.309 p < 0.05	r=0.472 p < 0.005
	/au/	r=0.747 p< 0.0001	r=0.685 p < 0.0001
	/oi/	r=0.627 p < 0.0001	r=0.243 p < 0.25
	/ui/	r=0.593 p < 0.0001	r=0.262 p < 0.10

In general, the Japanese correlations also show a much stronger link between maxslope and range than between fullslope and range. F2 range for /oi/ and /ui/ is much better correlated with maxslope than fullslope. /au/ exhibits fairly good correlations with both slope measures (maxslope/range, r=0.747, p<0.0001; fullslope/range, r=0.685, p<0.0001), while /ai/ displays poor correlations between F2 range and both slope measures (maxslope/range, r=0.309, p<0.05;

fullslope/range, $r=0.472$, $p<0.005$).

Analyses of Variance: English and Japanese

The results of the analysis of variance are displayed in **Table 5.** For the English data, speech rate ("slow","normal", "fast") had a significant effect on both maxslope ($F(2,221)=3.42$, $p=0.034$) and fullslope($F(2,221)=24.02$, $p<0.0001$). Rate also had a significant effect on range ($F(2,221)=53.48$, $p<0.0001$), transition duration ($F(2,221)=130.43$, $p<0.0001$), and onset ($F(2,221)=5.62$, $p=0.004$). Rate did not strongly affect F2 offset values ($F(2,221)=2.52$, $p=0.084$). The interaction of rate and diphthong also proved significant (for maxslope, $F(8,215)=2.65$, $p=0.009$; for fullslope $F(8,215)=3.98$, $p=0.0002$), indicating that changes in speech rate had a stronger influence on transition rate on some diphthongs than on others. Neither the token type (/h__d/ vs. /d__d/) nor interaction between token type and rate had a significant effect on any of the variables tested. "Speaker" had a highly significant effect on all five variables tested: onset, offset, range, maxslope, and fullslope. The interaction between rate and speaker was more significant for fullslope than maxslope, and had no effect on either F2 onset or offset values. Rate * speaker interaction had a highly significant effect on F2 range.

The analysis of variance for Japanese indicates that speech rate did not have a significant effect on maxslope ($F(2,162)=0.54$, $p=0.582$), fullslope ($F(2,162)=1.93$, $p=0.15$), or offset values ($F(2,162)=0.87$, $p=0.421$). However, rate did significantly affect range ($F(2,162)=17.47$, $p=0.0001$) and onset values ($F(2,162)=25.27$, $p<0.0001$). Transition duration was also affected by rate: ($F(2,162)=7.76$, $p=0.0006$). As for English, "speaker" had a highly significant effect on all five variables, reflecting individual differences among speakers. The interaction of speaker and rate had a small effect on maxslope ($F(12,152)=2.57$, $p=0.0042$), but no effect on fullslope ($F(12, 152)=1.28$, $p=0.2346$), while the interaction of diphthong-type and rate had no effect on maxslope ($F(12,152)=1.46$, $p=0.1952$), but a slight effect on fullslope ($F(12,152)=3.36$, $p=0.0041$).

Stepwise Regression: English

Results of the stepwise regression for English indicate a high correlation ($r^2=0.703$, $p<0.0001$) between English maxslope and F2 range. Transition duration is also correlated with maxslope, although to a much lesser degree than range. The two-step regression combining these two variables produces an r^2 value of 0.798, $p<0.0001$, indicating that the rate of change

Table 5. English and Japanese Analysis of Variance results.

English

	Maxslope	Fullslope	F2 Onset	F2 Offset	F2 Range	Trans. Dur.
Speech Rate	F (2,221)=3.42 p= 0.0348	F (2,221)=24.02 p≤ 0.0001	F (2,221)=5.62 p= 0.0043	F (2,221)=21.6 p= 0.1177	F (2,221)=53.48 p≤ 0.0001	F (2,221)=130.4 p≤ 0.0001
Token Type	F (1,222)=1.10 p= 0.2951	F (1,222)=1.5 p= 0.2215	F (1,222)=1.41 p= 0.2361	F (1,222)=1.6 p= 0.2072	F (1,222)=3.59 p= 0.0598	F (1,222)=2.7 p= 0.1020
Diphthong Type	F (4,219)=191.7 p≤ 0.0001	F (4,219)=127.92 p≤ 0.0001	F (4,219)=318.23 p≤ 0.0001	F (4,209)=452.7 p≤ 0.0001	F (4,219)=276.97 p≤ 0.0001	F (4,219)=27.15 p≤ 0.0001
Speaker	F (7,216)=7.43 p≤ 0.0001	F (7,216)=3.37 p= 0.0022	F (7,216)=11.51 p≤ 0.0001	F (7,216)=40.62 p≤ 0.0001	F (7,216)=25.64 p≤ 0.0001	F (7,216)=11.12 p≤ 0.0001
Rate * Speaker	F (14,209)=1.62 p= 0.0762	F (14,209)=1.96 p= 0.023	F (14,209)=0.67 p= 0.7996	F (14,209)=1.24 p= 0.2510	F (14,209)=3.44 p≤ 0.0001	F (14,209)=3.0 p≤ 0.000-
Rate * Diphthong	F (8,215)=2.65 p= 0.0089	F (8,215)=3.98 p= 0.0002	F (8,215)=3.24 p= 0.0018	F (8,215)=5.64 p≤ 0.0001	F (8,215)=4.7 p≤ 0.0001	F (8,215)=2.23 p= 0.0273

Japanese

	Maxslope	Fullslope	F2 Onset	F2 Offset	F2 Range	Trans. Dur.
Speech Rate	F (2,162)=0.54 p= 0.5822	F (2,162)=1.93 p= 0.1497	F (2,161)=25.27 p≤ 0.0001	F (2,162)=0.87 p= 0.4214	F (2,162)=17.47 p≤ 0.0001	F (2,161)=7.76 p= 0.0006
Diphthong Type	F (3,161)=848.85 p≤ 0.0001	F (3,161)=527.18 p≤ 0.0001	F (3,161)=212.44 p≤ 0.0001	F(3,161)=1795.8 p≤ 0.0001	F (3,161)=548.7 p≤ 0.0001	F (3,161)=31.14 p≤ 0.0001
Speaker	F (6,158)=17.76 p≤ 0.0001	F (6,158)=10.97 p≤ 0.0001	F (6,158)=7.82 p≤ 0.0001	F (6,158)=21.03 p≤ 0.0001	F (6,158)=15.02 p≤ 0.0001	F (6,158)=7.6 p≤ 0.0001
Rate * Speaker	F (12,152)=2.57 p= 0.0042	F (12,152)=1.28 p= 0.2346	F (12,152)=1.75 p= 0.0628	F (12,152)=1.67 p= 0.0798	F (12,152)=2.54 p= 0.0048	F (12,152)=1.38 p= 0.1843
Rate * Diphthong	F (6,158)=1.46 p= 0.1952	F (6,158)=3.36 p= 0.0041	F (6,158)=1.75 p= 0.1140	F (6,158)=1.38 p= 0.2254	F (6,158)=2.65 p= 0.0182	F (6,158)=9.47 p≤ 0.0001

across the thirty fastest milliseconds of an English diphthong can be predicted with approximately 80% accuracy, given the F2 range and overall transition duration. This variable-slope correlation is improved only slightly by including the effects of the other three variables in the analysis. The r^2 value for the five-step analysis is 0.818, $p<0.0001$, with this minimal change from the two-step analysis reflecting the relatively poor correlations between maxslope and the variables percentage of transition, V1 duration, V2 duration, and total duration.

The results for English fullslope produce similar correlations, with F2 range again being the most influential variable. Transition duration plays a much greater role here than it does with maxslope, however. The r^2 value for the simple correlation between Fullslope-Range is 0.516, $p<0.0001$, a figure which improves dramatically to 0.814, $p<0.0001$ when the effect of transition duration is figured in.

Discussion:

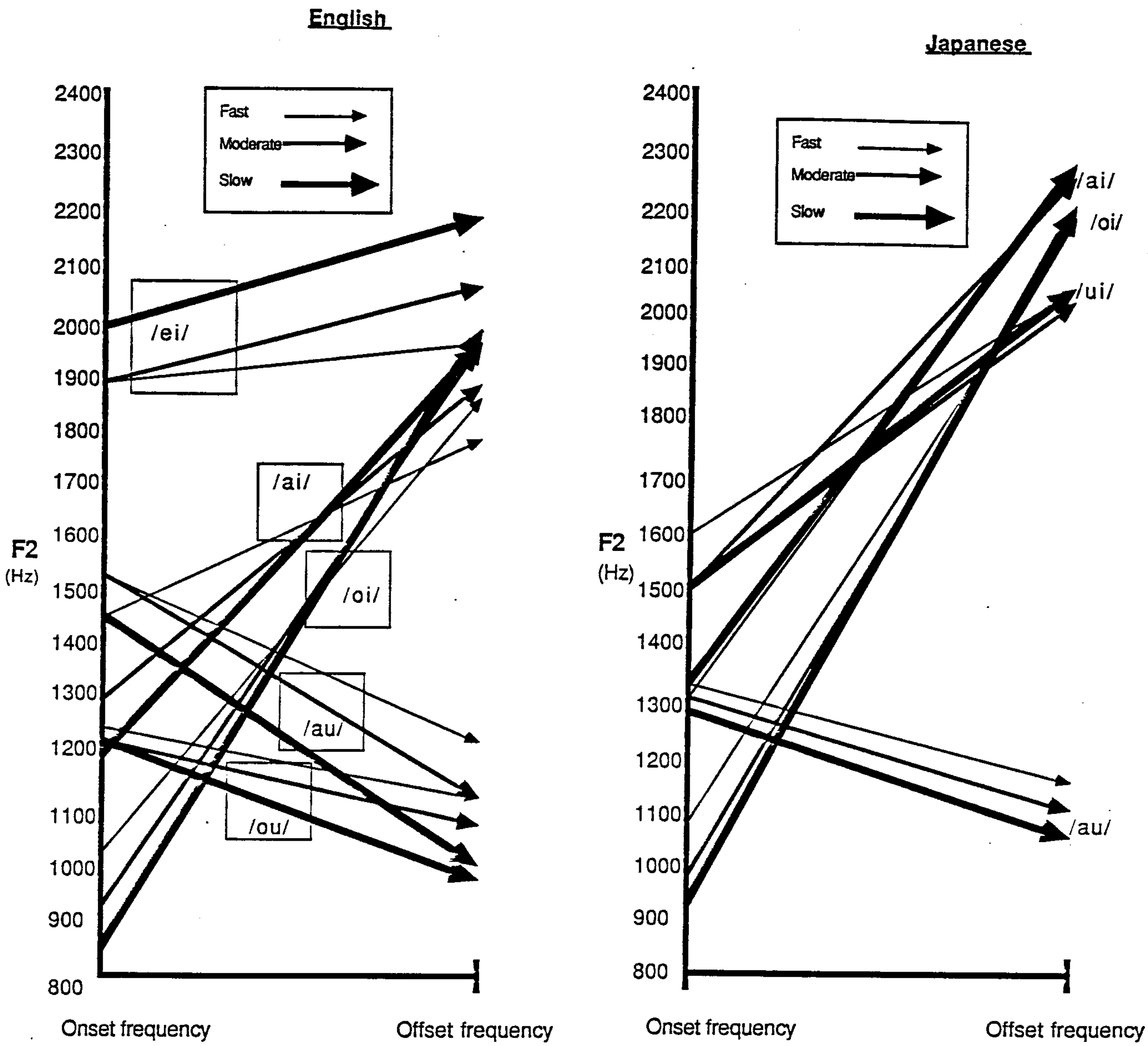
I. Targets:

Rate-dependent changes in F2 Onsets:

Figure 4 shows the change in mean frequency for each English and Japanese transition from F2 onset to offset. In both languages, F2 range tends to decrease as speech rate is increased, reflecting changes in both F2 onset and offset values. Interestingly, changes in F2 onset do not constitute a simple "headstart" on the transition in faster tokens, with the onset value climbing for diphthongs with rising F2 (/ai/, /oi/, /ui/,/ei/) and decreasing for those with falling F2 (/au/, /ou/). English /ei/, for example, displays a falling F2 onset with increased rate (from a "slow" 1981 Hz to a "fast" 1890 Hz), despite the fact that its overall F2 curve is a rising one. Onset values for all the other diphthongs tested in English and Japanese increase as speed increases (see **Figure 4**), including values for diphthongs with falling F2 slopes. Thus, the mean onset value for Japanese /au/ climbs from 1273 Hz at "slow" speed to 1343 Hz at "fast", even though the overall F2 transition slope is a falling one.

The regularity of this shift in F2 onset values, reflected in Japanese and English diphthongs with both rising and falling second formants, suggests that a change in diphthongal vowel quality may be a cross-linguistic effect of increased rate. Specifically, the first element of

Figure 4. Schematic comparison of English and Japanese change in mean frequency between transition onset and offset, broken down by diphthong/vowel sequence at three rates.



a complex vowel tends to move toward the center of the vowel space as rate is increased, resulting in a declining F2 onset for diphthongs which begin with a front vowel (/ei/), but a rising F2 onset for those which begin with a back vowel (/ai/, /oi/, /ou/, /au/, /ui/).

Offsets:

The rate-linked variability in F2 onset is not mirrored by consistent changes in offset values in either language. Mean values for Japanese offsets show remarkable constancy across rate changes, with offsets for /ai/, /oi/ and /ui/ barely changing. Mean values for English offsets display more variability, with relatively large changes as speech tempo increases (see figure). However, the results of the Analyses of Variance indicate that these changes cannot be directly linked to speech rate, since rate does not significantly affect offset values in either Japanese or English. This result is in sharp conflict with Gay's (1968) findings, which suggested that English offset values should fall out directly from duration information.

It would be oversimplification to claim that rate has no effect on offset values, however, and there are hints of support for Gay's position in our data. Mean offset values for the set of English diphthongs with rising F2--/ai/, /ei/, and /oi/ -- fall steadily as rate increases, reflecting the sort of undershoot of the final vowel target which Gay's model predicts. Offset for /ai/, for example, drops from a "slow" 1960 Hz to a "normal" 1881Hz, and finally to a "fast" 1762 Hz. English /ei/ and /oi/ exhibit similar progressions. The two English diphthongs with falling second formants also follow the general offset pattern predicted by Gay, with offsets rising as transition duration drops.

However, changes in English F2 offset values cannot be dependent simply on rate; diphthong type--whether the second formant is rising or falling--also plays a crucial role. Because a change in transition onset frequency seems to be a universal characteristic of fast-rate utterances, the F2 offset for "fast" tokens of diphthongs with a falling F2 must be additionally raised in order to counteract the rise in onset. Without such a compensatory effect, F2 range for these diphthongs would show an overall *increase* along with the increase in speech tempo. This does not occur in our data: mean F2 range decreases with increased speech rate for all the diphthongs tested. Further evidence for such a model of F2 onset/offsets is provided by the Japanese data, where the only diphthong to show any dramatic change in offset frequency is /au/, which rises from a "slow" 1052 Hz to "fast" 1175 Hz. This was the only Japanese diphthong with a falling F2 tested, and it is suggestive that /au/'s F2 offset rises in tandem with its onset, while offsets for the three Japanese diphthongs with rising F2's remain remarkably constant across rate changes .

II. Slopes:

Group Correlations--"The further the faster" as a general predictor of slope:

The correlations presented in **Table 3**, as well as the results of the English stepwise regression, show that maxslope and range are very closely linked for both the English and Japanese datasets. Given the F2 range produced by the articulation, it is possible to predict the slope of the second formant's thirty milliseconds of maximum velocity with a high degree of accuracy (English $r=0.838$, $p<0.0001$; Japanese $r=0.859$, $p<0.0001$). Fullslope also shows a strong correlation with F2 range for both languages (English $r=0.717$, $p<0.0001$; Japanese $r=0.778$, $p<0.0001$).

These facts support a description of production of both English diphthong and Japanese vowel sequence F2 transitions in terms of an acoustic equivalent of Kent and Moll's (1972) articulatory "the further the faster" model: as the distance traversed on the frequency scale increases, so does the rate of transition. In both English and Japanese, the maxslopes associated with particular diphthong or vowel sequence types are closely linked to the characteristic F2 ranges produced by these gestures. Toledo (forthcoming) finds similar results for eight diphthongs of Argentine Spanish, suggesting that languages may universally organize transition rates between members of complex vowel nuclei on the basis of distance traversed on the frequency scale.

English slope invariance refuted--individual diphthong correlations

Thus, for a range of languages, diphthongs involving a large change in F2 have sharper slopes than those with smaller F2 ranges. This finding is not necessarily related to the question of slope invariance across rate changes; change in speech rate need not affect speech transition slope. Allowing for the decrease in mean F2 range as speech rate is increased (an effect observed in all nine diphthong populations examined in English and Japanese), it is possible that a constant transition slope might be maintained by adjusting the transition duration--a shorter duration coupled with a smaller range could potentially yield the same slope as a longer transition with a greater range.

An examination of the individual diphthong populations in English shows that several display significant correlations between the F2 range and transition slope. English /ai/, for example, yields a correlation of r=0.687, p<0.0001 between maxslope and F2 range, while /ei/, /au/, and /ou/ also produce significant correlations between the two variables. For these four English diphthongs, then, the slope of maximum velocity portion of the transition is at least partly determined by the F2 range to be traversed, which is in turn partly determined by speech rate. Only English /oi/ exhibits a complete lack of correlation between range and either measure of slope. Looking at the mean maxslope/fullslope values for /oi/, however, suggests that this lack of correlation with range still does not indicate a constant transition velocity across different speech rates-- /oi/'s mean maxslope in the token /hoid/ increases from a "slow" value of 12.5 (sd=3.8) to "fast" 17.1 (sd=2.6).

Although /oi/ by itself displays no correlation between maxslope and F2 range, the set of /oi/ tokens <u>does</u> contribute significantly to the overall correlation between these two variables in **Table 3.** Removing /oi/ from the set of diphthongs causes the correlation between maxslope and range to drop substantially (from r= 0.838, p<0.0001 to r=0.780, p<0.0001). In effect, while individual tokens of /oi/ are not produced according to the "further the faster" principle, the set of /oi/ tokens as a whole does follow this pattern, with the great articulatory distance from target to target being reflected in faster transitions.

The analysis of variance for English indicates that for some diphthongs there is a highly significant effect of rate (ie, "fast"/"normal"/"slow") on both measures of slope, with faster speeds producing steeper slopes for four of the five diphthongs (/ai/ behaves somewhat anomalously, with its "normal" tokens producing the sharpest slopes). This effect is more pronounced for some diphthongs than others, as indicated by the highly significant interaction between speech rate and diphthong type (rate * diphthong interaction) on the slope measures. Thus, transition rate for a particular English diphthong may vary with rate and F2 range.

Influence of rate and transition duration: English:

Speech rate has a highly significant effect on transition slope for English, as indicated by the results of the stepwise regression and analysis of variance. Taken alone, F2 range is a better predictor of maxslope than fullslope. However, the results of the stepwise regression indicate that transition duration is also significantly tied to slope: adding the effect of transition duration to that of F2 range increases the maxslope correlation from $r^2=0.703$, $p<0.0001$ to $r^2=0.798$, $p<0.0001$.

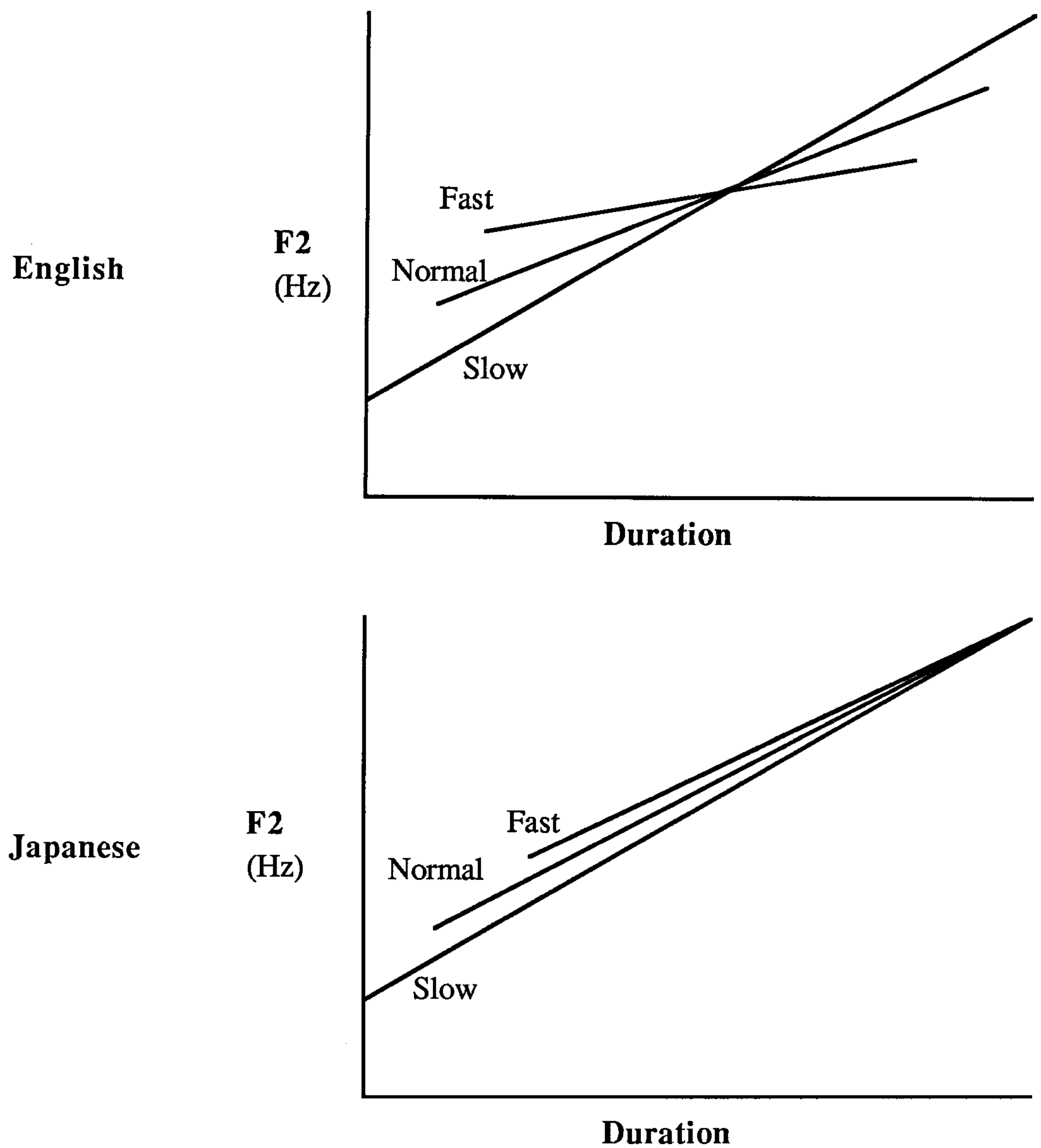
More dramatic is the improvement in the fullslope correlation when transition duration is added to F2 range, with the r^2 value jumping from 0.516, $p<0.0001$ to 0.814, $p<0.0001$. The minimal changes in r^2 values from the two-step analyses (transition duration and range combined with either maxslope or fullslope) to the full 5-step analyses reflect the relatively poor correlations between the slope measurements and the variables V1 duration, V2 duration, percentage of transition, and total duration. Only transition duration and F2 range have a strong effect on Japanese diphthong slopes.

Influence of rate and transition duration: Japanese:

In general, Japanese diphthongs are much less influenced by speech rate than English diphthongs. The Analysis of Variance indicates that both slope measures are unaffected by speech rate (maxslope $F(2,162)=0.54$, $p=0.5822$; fullslope ($F(2,162)=1.93, p=0.1497$), as is F2 offset ($F(2,162)=0.87$, $p=0.4214$). Support for a "further the longer" model for Japanese is provided by F2 range, which shows a highly significant effect of rate ($F(2,162)=17.47$, $p<0.0001$), and a significant correlation with transition duration ($r=0.43$, $p<0.0001$).

Thus, while English speakers routinely reorganize diphthong transition slopes to meet the demands of rate and range, Japanese speakers follow a different strategy, one much like that suggested by Gay for English speakers. While Japanese transition duration and F2 range decrease with increased speech rate, speakers do not adjust the transition's slope to fit new endpoint specifications. Instead, shorter duration and smaller F2 range are achieved primarily by adjusting the F2 onset (which shows a strong effect of rate ($F(1,161)=25.27$, $p<0.0001$), while the transition slope and F2 offset value are held constant. **Figure 5** represents an idealized comparison of English and Japanese slope patterns.

Figure 6. Idealized schematic of English vs. Japanese transition patterns under changes in rate.



III. Differential action of principles on transition slopes:

Diphthong-Specific Variation:

Interactions between speech rate and slope or rate and F2 range are clearly not an across-the-board phenomenon. In both English and Japanese, particular diphthongs display strong interaction between these variables, while others do not. The transition for Japanese /au/, for example, shows good correlations between F2 range and both slope measures (maxslope, $r=0.747$, $p<0.0001$; fullslope, $r=0.685$, $p<0.0001$), while English /oi/ exhibits poor correlations for both (maxslope/range, $r=0.297$, $p<0.040$; fullslope/range, $r=0.165$, $p<0.262$).

Edge Effects:

Even within a diphthong or vowel sequence population there may be a large gap between the maxslope and fullslope correlations with range. English /ai/, for example, produces $r=0.681$, $p<0.0001$ for the Maxslope-Range correlation, while the Fullslope-Range correlation for the same diphthong is only $r=0.281$, $p<0.10$. It thus appears that both speech rate and F2 range can interact in complex ways to affect certain vowel-vowel transitions but not others in a language, and even differentially affect portions of a particular transition.

F2 range is better correlated with maxslope than with fullslope for four of the five English diphthongs and three of the four Japanese vowel sequences. Such asymmetry implies a qualitative difference between the central portion of a transition and the slope as a whole, a notion at odds with Gay's view of diphthong transitions as homogeneous, pre-fabricated gestures whose angle is fixed without regard to transition duration. This constancy is crucial if primary perceptual salience is to be attributed to the slope angle (Gay, 1968; Manrique, 1979). Such a view predicts that a transition's length and F2 range will vary as rate changes, as will the occurrence of V1 and V2 steady-states, but that the rate of change itself will remain constant. F2 offset values will vary with rate due to the familiar phenomenon of target undershoot (Lindblom, 1963; Gay, 1981). Our data, however, suggests that in addition to alterations in onset and offset frequencies, the transition is internally complex. An "edge" effect is observable, with the slope in the central portion much more closely linked to the overall F2 range than are the "outlying" portions of the slope; neither English nor Japanese F2 transitions can be simply described in terms of a single governing principle.

IV. Modeling Production

Modeling diphthong behavior:

Transition production in diphthong seems to be affected by several general principles, notably "the further the faster", with transitions spanning a large F2 range being produced with a sharper slope than those with a smaller range. Fischer-Jørgensen (1964) advances a complementary hypothesis regarding duration, suggesting that "phonologized" segmental durations may in fact reflect a basic physiological constraint: bigger movements of the articulators result in longer durations. Fischer-Jørgensen, for example, attributes different characteristic durations for Danish stop consonants to this principle: "the extent of the articulatory movement is probably a decisive factor... a longer movement takes more time."

If F2 range can be equated with articulatory distance, ie. a larger range representing a greater movement of the tongue body, then diphthong behavior in both English and Japanese can be partly described in terms of this principle. The greater the articulatory/acoustic distance to be covered in the production of the transition, the longer its duration. Longer durations, in turn, contribute to larger F2 ranges, since less target undershoot occurs in the onset and offset. Thus, diphthong production can be affected by several general principles, with larger articulations resulting in more rapidly-changing, longer transitions which cover a greater distance on the F2 range.

English:

This is precisely what occurs in English: the data suggests a model of English diphthong coordination in which speakers produce a transition based on both the amount of time available for the transition and the distance on the F2 range which must be traveled. The slope of F2 for the central portion (maxslope) is very closely tied to F2 range, with speakers following a "further the faster" strategy, with some durational interaction. As speakers lengthen a diphthong around this portion, however, the "further the faster" strategy seems to be augmented by a formula for adjusting the slope of the transition edges according to speech rate, with faster rates resulting in a shorter, faster-changing transition. Both transition onset and offset values are subject to rate-dependent variation, as well as changes caused by the tendency to centralize diphthong-initial vowels.

Slopes in Japanese diphthongs do not vary directly with rate, but they are indirectly affected by changes in both rate-linked shifts in F2 range and transition duration. Since "the further the faster" explanation holds true for the Japanese dataset as a whole, as well as several of the individual diphthong populations, changes in F2 range will necessarily result in a change in the slope measures. (Part of the reason that these changes do not show up as directly linked to slope may lie in the behavior of the falling diphthong /au/, whose slope is altered by the centralizing increase in F2 onset and its resulting compensatory rise in F2 offset.) The interaction of rate and diphthong is significant for fullslope ($F(6,158)=3.36$, $p=0.0004$), indicating that for certain diphthongs, change in rate *does* directly result in an overall slope change.

Abstract vs. achieved endpoints:

Kent and Moll (1972, p.296) state that "the articulatory velocity during a vowel gesture probably depends more upon the vocal tract targets than upon the actual articulatory positions achieved during vowel production", implying that the undershoot associated with faster production of a complex vowel should have no effect on transition slope. This does not appear to be the case for diphthong production in either English or Japanese: the strong range/slope correlations produced by our data are based on *absolute* F2 ranges for these tokens, with slope values tied to the actual targets achieved rather than abstract targets associated with the diphthong.

Conclusions:

Characteristic F2 transition rates for diphthongs in both Japanese and English seem to be organized according to a "further the faster" strategy , with the range to be traversed on the frequency scale largely determining the angle of the transition. Thus /oi/, with its very large F2 range, has the most extreme slope in both languages, while English /ei/, with its much less extreme F2 range, has the shallowest transition.

In both languages, F2 range decreases as tempo is increased. Onsets climb with increased speech rate, for both diphthongs with both rising and falling second formants in English and Japanese. The resulting centralization of the initial element of a diphthong may prove to be a cross-linguistic effect of increased speech rate. Offset values, on the other hand, behave differently in English and Japanese. Japanese offsets for diphthongs with rising F2's are constant across rate changes, while for the falling diphthong /au/, offsets rise as speed increases. English displays variable F2 offsets as speech rate changes, although change in offset cannot be directly linked to rate. Diphthong type--whether the second formant is rising or falling--also seems to play a role in affecting offsets, with offsets for diphthongs with falling F2 climbing to counteract the raised onset.

Our data does not support the claim that English speakers produce an invariant slope at different speech rates. Instead, rate plays a significant role, with the amount of time available for the transition affecting both the range covered by the transition and its slope. In addition to "the further the faster" principle governing English diphthong production, "the further the longer" seems to play a role in determining the slope--transition duration increases along with F2 range. This suggests that F2 transitions in English diphthongs represent a sort of compromise between these two competing principles: the "further the faster" acts to increase transition rate as F2 range increases, while the "further the longer" decreases the rate of change by lengthening the transition as F2 range increases. The "further the faster" takes overall precedence; F2 range is much more closely linked to the slope measures than to transition duration.

Nor do the two principles act evenly across the whole transition--central portions of both English and Japanese complex vowels are more amenable to a "further the faster" analysis than are the overall slopes--indicating that the transitional component of a diphthong or vowel sequence cannot be considered a single, homogeneous slope.

As with English diphthongs, transition slopes for Japanese vowel sequences appear to be produced primarily on the basis of F2 range to be traversed. Unlike English, however, speech rate and transition duration play a far less significant role in determining the overall slope of the transition in Japanese. The differences between the two languages suggest that while temporal reorganization for different speech rates is affected by general physiological and phonetic principles, these principles may apply in unique ways to different languages, and even to different diphthongs within a single language.

References:

Fischer-Jørgensen, Eli (1964). "Sound duration and place of articulation in Danish". Zeitschrift für Phonetik Sprachwissenschaft und Kommunkationforschung. Band 17, Heft 2-4: 175-207.

Fujisaki, Hiroya and Norio Higuchi (1979). "Temporal Organization of segmental features in Japanese disyllables", Ninth International Congress of Phonetic Sciences. Proc., v.II: 275-282, Copenhagen.

Gay, Thomas (1967). "A Perceptual Study of American English Diphthongs," Status Rep. Speech Res., Haskins Lab. Suppl., 1-80.

Gay, Thomas (1968). "Effect of Speaking Rate on Diphthong Formant Movements," J. Acoust. Soc. Amer. 44:6, pp 1570-1575.

Gay, Thomas (1981). "Mechanisms in the control of speech rate", Phonetica 38:148-158.

Kent, R. D. and K. L. Moll (1972). "Tongue body articulation during vowel and diphthong gestures", Folia Phoniatrica, 24:278-300.

Lehiste, Ilse, and Gordon E. Peterson (1960). "Duration of Syllable Nuclei in English". J. Acoust. Soc. Amer. 32: 693-703.

Lindau-Webb, Mona (1985). "Hausa vowels and diphthongs", UCLA Working Papers in Phonetics, no. 60.

Lindau-Webb, Mona, Kjell Norlin and Jan-Olof Svantesson (1985). "Cross-linguistic differences in diphthongs", UCLA Working Papers in Phonetics, no. 61.

Lindblom, B. (1963). "Spectrographic study of vowel reduction", J. Acoust. Soc. Amer. 35:1773-1781.

MacNeilage, P. F. (1970). "Motor control of serial ordering of speech", Psychological Review, 77:182-196.

Manrique, Ana Maria Borzone (1979). "Acoustic Analysis of the Spanish Diphthongs". Phonetica 36:194-206.

Ren, Hongmo (forthcoming). "The Acoustic Structure of Diphthongal Syllables" PhD. dissertation, Univ. of Calif., Los Angeles, 1986.

SAS (1985). "SAS/stat Guide for Personal Computers, Version 6 edition." SAS Insitute inc. Box 8000 Cary, North Carolina.

Acknowledgements:

The authors are greatly indebted to Ian Maddieson, Mona Lindau-Webb, Peter Ladefoged, and Ren Hongmo for their help and support, and to Norma Antoñanzas for the programming. Financial support came from the National Institute of Health.

Comprehension of Familiar Phrases by Left- but not by Right-hemisphere Damaged Patients

by

Diana Van Lancker and Daniel Kempler

Overlearned, familiar phrases, also termed "automatic speech,"[1] are often selectively preserved in aphasia, and exemplary lists of residual phrases of this type have been included in clinical descriptions from the time of Broca. This phenomenon appears to be quite general, and cuts across aphasia type: fluent aphasic patients have been observed to use formulaic speech amidst otherwise paraphasic and neologistic output (Bay, 1964; Luria and Tsvetkova, 1968); anomia is often not extended to formulaic speech (Bay, 1963; Wepman, Bock, Jones and Van Pelt, 1956); and in conduction aphasia, repetition of familiar phrases was found to be significantly better than repetition of novel phrases (Canter, Coughlin, and Van Lancker, 1978). But the most striking incidence is in the nonfluent and global aphasias, where, despite severe impairment in production of words and novel utterances, these overlearned utterances are produced with normal prosody and fluent articulation, such that rudimentary communication is often achieved. In fact, the differences between formulaic and novel speech in fluency, articulatory precision, prosody, pragmatic characteristics, and communicative effectiveness are so striking as to suggest that different cerebral mechanisms are involved.

The theory that familiar phrases are processed differently from novel phrases agrees with data from psychological studies of normal subjects, which show systematic processing differences between novel and familiar expressions in children (Vihman, 1982; Pollio and Pollio, 1974; Wong Fillmore, 1979; Peters, 1977, 1983) and adults (Osgood and Hoosain, 1974; Horowitz and Manelis, 1973; Gibbs, 1980; Pickens and Pollio, 1979; Lieberman, 1963; Van Lancker, Canter, and Terbeek, 1981; Simon, 1974). In all these studies, novel expressions were observed to be processed--categorized, recognized, learned and/or remembered--in terms of their individual component parts and constituent structure, whereas familiar expressions presented in a same paradigm were processed as a unitary wholes, without comparable attention to constituency.

These observations in normal and aphasic speakers suggest that familiar phrases are unlike novel phrases (1) in their linguistic structure and (2) in their cerebral representation. Because of their formal similarities--both words and familiar phrases constitute single, grammatically unanalyzed units -- it has been proposed that familiar phrases are processed like single words. In support of this proposal, two studies using a lexical decision task, one of normal subjects (Swinney and Cutler, 1979) and the second conducted on aphasic patients (Dronkers, 1984) led to the conclusion that familiar phrases are processed more like single words than like syntactic phrases of comparable length and complexity.

The second point, that familiar phrases are represented differently in the brain from novel expressions, can be traced back to the writings of Jackson (1878). Two possible theories of the neurological representation of familiar phrases must be considered. First, parallel to the suggestion that familiar phrases are structurally similar to single words, it has been proposed that these share neurological substrates. That is, while the ability to concatenate words into syntactic strings appears to be mediated by the anterior left frontal lobe (e.g., Mohr, 1976), the production of single words can be disrupted by focal damage to widespread areas of the left hemisphere (Benson, 1979). Thus, given that the ability to produce single words is disturbed by damage to different areas of the brain, and that, therefore, no specific region of the left hemisphere has been thought to be the "locus" of naming (Knopman, Selnes, Niccum, and Rubens, 1984), it might be concluded that formulaic speech, like naming, is diffusely represented in the left hemisphere.

There are problems with a theory that proposes a parallel between single words and familiar phrases in cerebral representation. Unexplained, for example, is the fact that formulaic speech is frequently preserved--produced with ease and fluency--when single word production is severely impaired. Often when an aphasic patient cannot spontaneously name or refer to common objects, he/she can do serial counting, and use expletives and social interaction formulae. Therefore, it appears that abilities to produce formulaic speech are even less vulnerable to loss resulting from left hemisphere damage than the production of single words. These clinical observations point to a possible role of the right hemisphere in familiar phrase processing.

Several other clinical observations implicate the right hemisphere in the production of familiar phrases. First, as mentioned above, residual formulaic speech has been observed in all types of aphasia, implying their preservation following left hemisphere lesions of varying location and size. Secondly, studies using the Wada procedure reported continued aphasic output during anesthetization of the left (dominant) hemisphere (Kinsbourne, 1971; Czopf, 1981), implying that residual aphasic speech was represented in the right (unanesthetized) hemisphere. Further, a callosal sectioned patient was reported to speak via right hemisphere mechanisms (Levy, Nebes and Sperry, 1971), and several clinical cases have been reported in which the nondominant right hemisphere was the site of aphasic speech (Landis, Cummings and Benson, 1980; Cummings, Levine and Mohr, 1979; Benson, Walsh and Levine, 1979). Right hemisphere activation during automatic speech was observed in a blood flow study (Ingvar and Schwartz, 1974; Larsen et al., 1978). In addition, several left (dominant) hemispherectomized adults with little or no propositional speech were observed to produce fluent, normally intoned formulaic speech(Crockett and Estridge, 1951; Hillier, 1954; Smith, 1966; Smith, 1974; Bogen, 1973). These facts tend to support the alternate view of formulaic speech which associates its cerebral representation with the right hemisphere (Jackson, 1878; Van Lancker, 1973, 1975; 1986).

Most observations in preserved aphasic production of familiar phrases are anecdotal, as it is difficult, when dealing with a disorder as individual as aphasia and a phenomenon which is unique in every patient (i.e., no two patients utter the same set of overlearned pharases), to directly compare and contrast abilities to produce familiar and novel utterances. Similarly, the theory which attributes control over familiar phrases to the nondominant hemisphere is difficult to test since right hemisphere patients are not usually aphasic, and subtle variations, such as diminished used of familiar phrases, would be difficult to observe and quantify.

As well known as this differential in aphasic ability is in production, little information on aphasic comprehension of familiar phrases is available. In order to systematically compare novel and familiar phrase knowledge, a comprehension test was needed. Such a test could be used to examine the hypothesis that novel and formulaic expressions have different cerebral representations in the left hemisphere or, alternatively, that the right hemisphere plays a role in familiar phrase comprehension. First we could observe whether performance on familiar phrases was more like that on single words or on matched novel phrases. The finding that preserved ability for familiar phrase comprehension occurs alongside an impairment in syntactic processing in left brain damage (LBD) would support the hypothesis that formulaic expressions are represented differently from novel language. The hypothesis of a right hemisphere involvement would be supported if an impairment in familiar phrase comprehension were observed in right brain damage (RBD), but not if a preserved ability for familiar phrase recognition occurred in RBD. These hypotheses could be assessed by testing unilaterally left- or right-brain damaged subjects on comprehension of single words, familiar phrases, and novel phrases, matched to the familiar phrases in length, number of words, and surface phrasal structure[2].

METHODS

Subjects

Experimental subjects were 39 consecutively available, right-handed, native speakers of American English (5 females, 34 males). All subjects had normal vision with corrective lenses, and none had a hearing impairment sufficient to interfere with speech perception. Age ranged from 45 to 80 years (mean age = 62.8); education ranged from 10 to 19 years. There were 28 subjects in the LBD group, with a mean age of 62.3, and an average of 63.5 months since onset of lesion. All were aphasic: aphasia type was diagnosed by a speech-language clinician using standardized instruments (The Boston Diagnostic Aphasia Examination and the Western Aphasia Battery). Eight patients had fluent aphasia, 10 had nonfluent aphasia, 5 were primarily anomic, 3 were globally aphasic, and 2 were mixed. The 11 subjects in the RBD group had an average age of 63.4, and a mean time-post-onset of 19.5 months. All had unilateral, focal lesions as a result of a cerebral vascular accident, as determined by radiological reading of CT-scans and neurological examination. Side and etiology of lesion were further supported by EEG records, neuropsychological testing (Wechsler Adult Intelligence Scale, verbal and visual memory tests, and visuospatial construction) and the neurologist's report.

Normal control subjects were 50 adults (36 females and 14 males) ages 45-82 (mean age = 62.2), with 11-19 years of education.

Materials

Ten single words (e.g., concrete nouns such as bird, telephone), 20 familiar phrases (e.g., "He's turning over a new leaf," "While the cat's away the mice will play"), and 10 novel sentences comparable in length and surface syntactic structure, and comprised of words with similar text frequency-counts to the familiar phrases (e.g., "He's sitting deep in the bubbles," "When the happy girl pushes, the angry boy swings") were selected and 4 line drawings were prepared for each item. Foils (wrong answers) in the word test were semantically related items. Foils in the familiar phrase portion of the test included one literal response item (e.g., a man raking leaves for "He's turning over a new leaf"), one related or opposite in meaning (e.g., an angry, defiant convict sitting in a prison cell bunk) and an irrelevant item (e.g., people sitting in a movie theater). (See Figure 1). All 3 distractor items for the novel sentences were permutations of grammatical roles (e.g., agent & patient reversed) or adjective mis-assignments (e.g., "When the happy girl pushes, the angry boy swings" includes a picture of an unhappy girl pushing a boy, etc.).

Each patient received two practice items for each stimulus set and then a set of 10 words, 10 familiar phrases and 10 novel sentences. Two sets of materials were prepared for the familiar phrases to ensure that there was no effect of individual phrases and/or distractor items. The two sets were alternated between consecutive patients. Because no such effect was observed, the data were pooled for analysis.

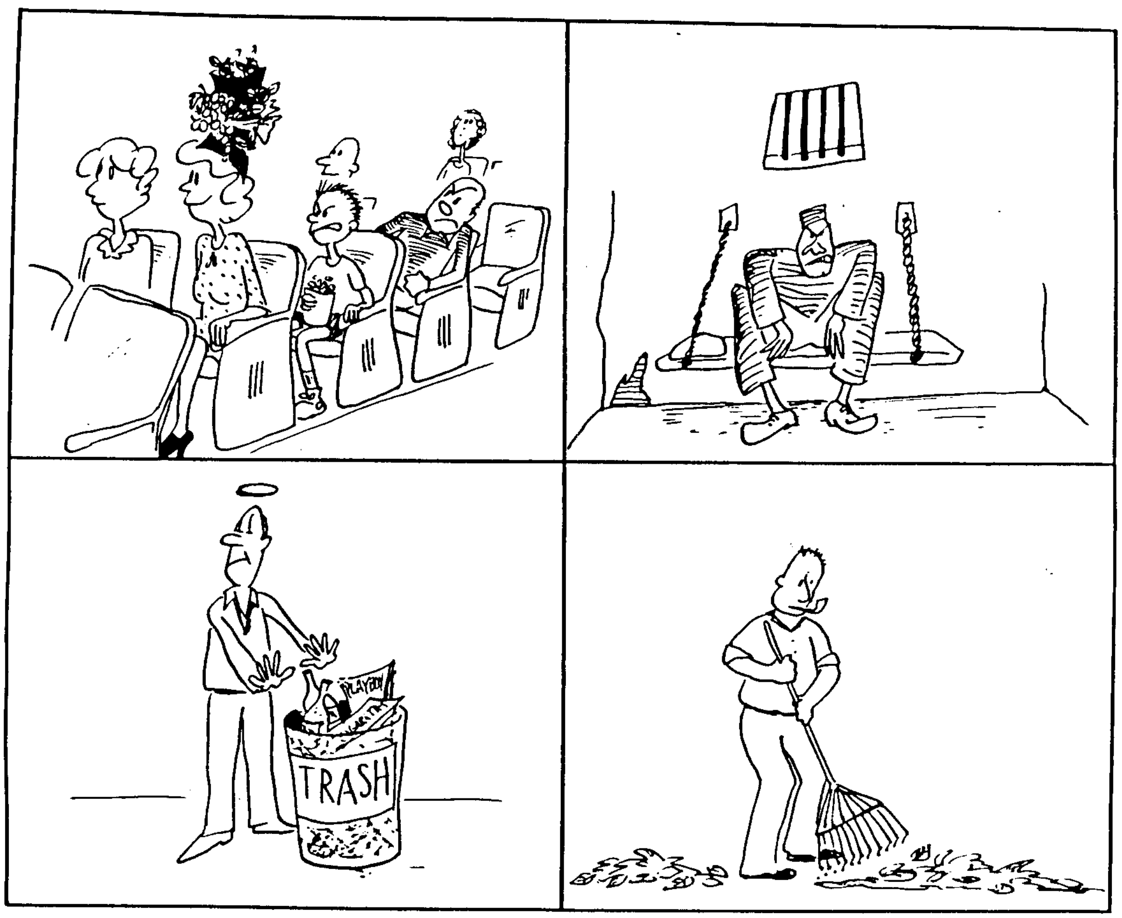


Figure 1. A response sheet for the familiar phrase "He's turning over a new leaf."

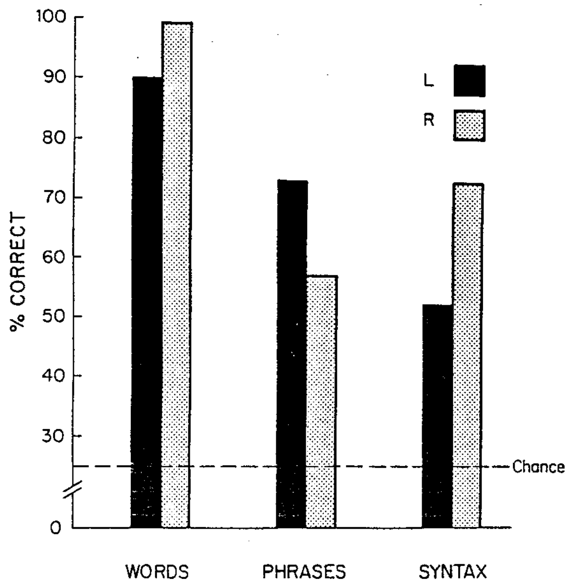


Fig. 2. Mean scores on words, familiar phrases, and novel sentences for LBD and RBD groups.

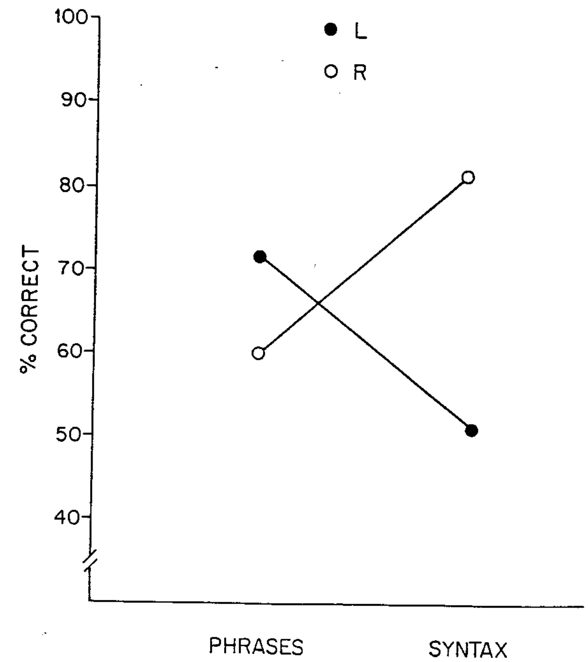


Fig. 3. Illustration of task differences for familiar phrases and novel sentences between LBD and RBD groups.

RESULTS

Results for LBD and RBD groups can be seen in Figure 2. As intended in the design of the protocol, normal subjects performed nearly perfectly on all three tasks, achieving 100% correct on single words, 97.3% correct on familiar phrases, and 99.8 percent correct on novel phrases. No differences in performance associated with sex, age, or education were observed. Statistical analysis comparing the clinical populations revealed that the two groups showed opposite patterns of performance. On an analysis of variance (ANOVA) of RBD and LBD groups with repeated measures on the three tasks (words, familiar phrases, and novel sentences), a nonsignificant trend for an effect of group ($F = 3.15$, $df = 1,37$, $p = .08$), a significant task effect ($F = 35.91$, $df = 2,74$, $p < .001$) and a significant group by task interaction ($F = 17.44$, $df = 2,74$, $p < .001$) were observed. Another ANOVA on repeated measures on the two tasks of interest, familiar phrases and novel sentences resulted in another significant task by group interaction ($F = 35.69$, $df = 1,37$, $p < .001$). The group by task interactions on both the 3-way and 2-way ANOVAS indicate different patterns of abilities for the two groups, with the LBD group more likely to preserve familiar phrase comprehension with impairment to syntactic processing, while the RBD group showed selective impairment of familiar phrase comprehension (Figure 3).

DISCUSSION

The finding that familiar phrase recognition is relatively less impaired than are syntactic abilities in LBD aphasic patients indicates that the preservation of abilities for formulaic speech production in aphasia extends to comprehension. That is, this selectively preserved speech ability is not limited to performance of overlearned motor patterns, but represents an intact language ability cutting across input and output modalities.

These results support the hypothesis, derived from observations in production, that familiar phrases are stored and processed in the brain differently from newly generated language. In particular, that LBD and RBD patients show opposite patterns of performance on familiar phrase versus novel sentence comprehension (Figure 3) implies that different neural substrates subserve these two language functions.

Third, these results show that RBD subjects have a language disturbance. The deficiency in familiar phrase comprehension associated with damage to the right hemisphere has not yet been described, although there are related observations. An impairment in comprehension of metaphors, many of which were frozen metaphors and therefore are very similar to familiar phrases in requiring an idiomatic and unitary interpretation ("He has a heavy heart") was reported by Winner and Gardner (1977), and a deficit in idiom comprehension associated with RBD was reported by Myers and Linebaugh (1981). It is not known whether this impairment in comprehension of familiar phrases extends to production in RBD patients. However, observed deficiencies in this population in conversational turn-taking, which requires use of social interaction formulas, suggest that the deficit may involve both input and output (Foldi, Cicone and Gardner, 1983; Jaffe, 1978; Myers, 1979).

The theory that familiar phrases are processed like lexical units is not wholly supported by these results, because performance scores on familiar phrases and words were not parallel. The differences might be attributable to the fact that whereas familiar phrases are like words in form (i.e., they both comprise unitary, non-syntactically analyzed wholes), they have different semantic properties. For example, familiar phrases are more semantically "complex" in containing a set of propositions (see footnote 3), evoking meanings not derivable from the words themselves, and mapping onto a complex "scenario" full of obligatory detail.

That familiar phrases are processed like unitary, non-syntactically analyzed elements is indirectly suggested by the findings in RBD subjects, if it can be inferred that the deficiency in familiar phrase recognition is attributable to the general impairments of pattern recognition associated with RBD (Bogen, 1969a, b; Levy, 1974; Bever, 1975; see review by Bradshaw and Nettleton, 1983). This modified version of the "familiar-phrase-as-single-word" hypothesis, which maintains that familiar phrases behave in performance like words formally (i.e., structurally) but not semantically is compatible with the Dronkers (1984) and Swinney and Cutler (1979) lexical decision findings (in which familiar phrases were responded to like single words), because in that study the task was to make a response to the overall form but not to the meaning of the phrases. In contrast, in the experiment reported here, the task required subjects to process both the holistic form and the complex meaning of familiar phrases. This led to different results in the two experimental groups (see Figure 3). Presumably, the LBD group was at an advantage because the familiar phrases did not require syntactic analysis, whereas the RBD group was at a disadvantage because the phrases, structurally, required apperception of an overall pattern.

It is interesting to note that although the meanings of the familiar phrase stimuli used in this study, as mentioned above, are quite complex, aphasics showed better comprehension of these than of novel expressions. This observation has implications for aphasic rehabilitation and family counseling. It may be demonstrated, in certain cases, that severely impaired aphasics have relatively preserved comprehension of formulaic speech, which is of considerable value and importance in everyday social interaction (Bolinger, 1976; Pawley and Syder, 1980; Fillmore, 1979; Jesperson, 1965; Tyler, 1978). Furthermore, the seemingly language-bereft individual may be able to comprehend quite complex "propositional" meanings[3], when they are offered in the right kind of phrase. This information will be of value to the clinician designing a plan for therapy, and to the family of the patient.

Finally, these results add a few more lines to the unfolding story of the role of the right hemisphere in language and communication. The association of impaired comprehension of familiar phrases with RBD strongly suggests a role of the right hemisphere in the normal comprehension of formulaic speech. Its role in formulaic speech production is not clear, although the observations in Wada testing in aphasia and adult hemispherectomy speech mentioned above suggest that the right hemisphere functions as the substrate for such speech behavior, at least in abnormal conditions. Recent reports have indicated right hemisphere involvement in paralinguistic parameters of the speech signal such as emotional meanings (Wechsler, 1972; Van Lancker, 1980; Heilman, 1975; Ross, 1981, 84; Kent and Rosenbek, 1982) and personal voice identity (Van Lancker and Canter, 1982; Van Lancker, Cummings, and Kreiman, 1985; Van Lancker and Kreiman, 1986). In addition, the findings reported here suggest that the right hemisphere plays an important role in processing formulaic language.

Footnotes

[1]We prefer the term "formulaic speech," and use "familiar phrases" to refer generically to the overlearned, holistic expressions that characterize this language behavior, including, *inter alia,* social interaction formulae (e.g., greetings), expletives, overlearned lists and serials (e.g., days of the week), song lyrics, proverbs, and idioms.

[2]The familiar-phrase-as-single-word hypothesis entails that familiar phrases do not have grammatical structure in the sense that novel expressions do, but that they are processed as unitary single elements. That is, applying standard grammatical processes to a familiar phrase, such as "She has him eating out of her hand" will result in a wrong interpretation. But for experimental purposes, the surface "structures" of the familiar phrase stimuli were matched to a parallel set of novel phrases.

[3]Note that although the phrases of interest here are called "formulaic," they map onto complex "propositional" meanings. For example, "When the cat's away, the mice will play," means that when a person who is viewed as an authority by two or more subordinates leaves the area, those subordinates can be expected to engage in activities held by the authority to be unacceptible or reprehensible, and that they will do this with an air of "getting away with something" and so on. These kinds of meanings were incorporated into our drawings as far as was possible, and aphasic patients were seen to make a significant number of correct responses on these items.

REFERENCES

Bay, E. 1963. Aphasia and conceptual thinking, in L. Halpern (Ed.), Problems of Dynamic Neurology, Jerusalem.

Bay, E. 1964. The history of aphasia and the principles of cerebral localization. In G. Schaltenbrand and C. Woolsey (Eds.), Cerebral localization and organization. Madison, Wisconsin: University of Wisconsin Press.

Benson, D.F. 1979. Neurologic correlates of anomia. In H. Whitaker and H.A. Whitaker (Eds.), Studies in Neurolinguistics, Vol. 4, New York, Academic Press.

Bever, T.G. 1975. Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: holistic and analytic. Annals of the New York Academy of Science, 263, 251-262.

Bogen, J. 1969a. The other side of the brain II: An appositional mind. Bulletin of the Los Angeles Neurological Societies, 34(4), 135-162.

Bogen, J. 1969b. The other side of the brain III: An appositional mind. Bulletin of the Los Angeles Neurological Societies, 34(4), 191-219.

Bogen, J. 1973. Movie of a hemispherectomy patient and what it all means. UCLA Conference on Cerebral Dominance. Los Angeles.

Bolinger, D. 1976. Meaning and memory. Forum Linguisticum, 1, 1-14.

Bradshaw, J.L. and Nettleton, N.D. 1983. Human cerebral asymmetry. Englewood Cliffs, New Jersey: Prentice Hall.

Canter, G., Coughlin, and Van Lancker, D. 1978. Observations on a conduction aphasic. Paper presented at the American Speech and Hearing Association, November.

Crockett, H.G. and Estridge, N.M. 1951. Cerebral hemispherectomy. Bulletin of the Los Angeles Neurological Societies, 16, 71-87.

Cummings, D.L., Benson, D.F., Walsh, M.J. and Levine, H.L. 1979. Left-to-right transfer of language dominance: a case study. Neurology, 29, 1547-1550.

Czopf, J. 1981. Uber die Rolle der nicht dominanten Hemisphare in der Restitution der Sprache der Aphasischen. Archives Psychiatrischen Nervenkrankhiten 216, 162-171.

Dronkers, N.F. 1984. The role of syntax in automatic speech processing. Paper presented at the 22nd Meeting of the Academy of Aphasia, Los Angeles, CA.

Fillmore, C. 1979. On Fluency. In C.J. Fillmore, D. Kempler and S-Y Wang (Eds.), Individual differences in language ability and language behavior. New York: Academic Press.

Foldi, N.S., Cicone, M. and Gardner, H. 1983. Pragmatic aspects of communication in brain-damaged patients. In S.J. Segalowitz (Ed.), Language functions and brain organization. New York: Academic Press.

Gibbs, R. 1980. Spilling the beans on understanding and memory for idioms in conversation. Memory and Cognition, 8, 149-156.

Heilman, K.M., Scholes, R. and Watson, R. 1975. Auditory affective agnosia. Journal of Neurology, Neurosurgery and Psychiatry, 38, 69-72.

Hillier, W.E. 1954. Total left hemispherectomy for malignant glioma. Neurology, (Minn.), 4, 718-721.

Horowitz, L.M. and Manelis, L. 1973. Recognition and cued recall of idioms and phrases. Journal of Experimental Psychology, 100, 291-296.

Ingvar, D.H. and Schwartz, M.S. 1974. Blood flow patterns induced in the dominant hemisphere by speech and reading, Brain, 97, 273-288.

Jackson, J.H. 1878. On affections of speech from disease of the brain, Brain, 1, 304-330. (Reprinted in J.Taylor (Ed.), Selected writings of John Hughlings Jackson, Vol. 2, New York: Basic Books, 1958).

Jaffe, J. 1978. Parliamentary procedure and the brain. In A.W. Siegman and S. Feldstein (Eds.), Nonverbal behavior and communication, Hillsdale, New Jersey: Lawrence Erlbaum.

Jespersen, O. 1965. The philosophy of grammar. New York: Norton.

Kent, R.D. and Rosenbek, J.C. 1982. Prosodic disturbance and neurologic lesion. Brain and Language, 15(2), 259-291.

Kinsbourne, M. 1971. The minor cerebral hemisphere as a source of aphasic speech. Transactions of the American Neurological Association, 96, 141-145.

Knopman, D.S., Selnes, O.A., Niccum, N. and Rubens, A.B. 1984. Recovery of naming in aphasia: Relationship to fluency, comprehension, and CT findings. Neurology, 37.

Larsen, L.B., Skinnoj, E. and Lassen, N.A. 1978. Variation in regional cortical blood flow in the right and left hemisphere during automatic speech. Brain, 101, 193-209.

Landis, T., Cummings, J.L. and Benson, D.F. 1980. Le passage de la dominance du langage a l'hemisphere droit: une interpretation de la recuperation tardive lors d'aphasie globales. Revue Suisse Romande, 100, 171-177.

Levine, D.N. and Mohr, J.P. 1979. Language after bilateral cerebral infarction: role of the minor hemisphere in speech. Neurology, (Minn.) 29, 927-38.

Lieberman, P. 1963. Some effects of semantic and grammatic context on the production and perception of speech. Language and Speech, 6, 172-187.

Levy, J. 1974. Psychobiological implications of bilateral asymmetry. In S.J. Dimond and J.G. Beaumont (Eds.), Hemisphere function in the human brain, New York: Johns Wiley and Sons, 121-183.

Levy, J., Nebes, R.D. and Sperry, R.W. 1971. Expressive language in the surgically separated minor hemisphere. Cortex, 7, 49-58.

Luria, A.R. and Tsvetkova, L.S. 1968. The mechanism of "dynamic aphasia." Foundations of Language, 4, 296-307.

Mohr, J.P. 1976. Broca's area and Broca's aphasia. In H. Whitaker and H.A. Whitaker (Eds.), Studies in neurolinguistics, Vol. 1. New York, Academic Press, 201-237.

Myers, P.S. 1979. Profiles of communication deficits in patients with right cerebral hemisphere damage: Implication for diagnosis and treatment. In R. Brookshire (Ed.), Clinical aphasiological conference proceedings, (Minn.) BRK Publishers.

Myers, P.S. and Linebaugh, C.W. 1981. Comparison of idiomatic expressions in right-hemisphere-damaged adults. In R. Brookshire (Ed.), Clinical aphasiological conference proceedings, (Minn.), BRK Publishers, 254-261.

Osgood, C. and Hoosain, R. 1974. Salience of the word as a unit in the perception of language. Perception and Psychophysics, 15, 168-192.

Pawley, A. and Syder, F.H. 1980. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J.C. Richard and R. Schmidt (Eds.), Communicative competence. London: Longmans.

Peters, A.M. 1977. Language learning strategies: does the whole equal the sum of the parts? Language, 55, 560-573.

Peters, A.M. 1983. The units of language. Cambridge: Cambridge University Press.

Pickens, J.D. and Pollio, H.R. 1979. Patterns of figurative language competence in adult speakers. Psychological Research, 40, 299-313.

Pollio, M.R. and Pollio, H.R. 1974. The development of figurative language in children. Journal of Psycholinguistic Research, 3, 195-201.

Ross, E.D. 1981. The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. Archives of Neurology, 38(9), 561-596.

Simon, H.A. 1974. How big is a chunk? Science 183, 482-488.

Smith, A. 1966. Speech and other functions after left (dominant) hemispherectomy. Journal of Neurology Neurosurgery and Psychiatry 29, 467-471.

Smith, A. 1974. Dominant and nondominant hemispherectomy. In M. Kinsbourne and L. Smith (Eds.), _Hemispheric disconnection and cerebral function._ Illinois: Chas C. Thomas.

Swinney, D.A. and Cutler, A. 1979. The access of processing idiomatic expressions. _Journal of Verbal Learning and Verbal Behavior._ 18, 523-534.

Tyler, S.A. 1978. _The said and the unsaid_. New York: Academic Press.

Van Lancker, D. 1973. Language, lateralization and grammars. In J.P. Kimball, (Ed.), _Syntax and semantics,_ Vol II. New York: Seminar Press, 1973.

Van Lancker, D. 1975. Heterogeneity in language and speech: Neurolinguistic studies. _Working Papers in Phonetics,_ 29, UCLA, Phonetics Laboratory, Los Angeles.

Van Lancker, D. 1980. Cerebral lateralization of pitch cues in the linguistic signal. _Papers in Linguistics: International Journal of Human Communications,_ 13(2), 201-277.

Van Lancker, D. 1986. Nonpropositional speech: Neurolinguistic studies. In Andrew Ellis (Ed.), _Advances in the Psychology of Language,_ Vol. 2. New York: Academic Press.

Van Lancker, D., Canter, G.J. and Terbeek, D. 1981. Disambiguation of ditropic sentences: acoustic and phonetic cues. _Journal of Speech and Hearing Research,_ 24, 330-335.

Van Lancker, D. and Canter, G.J. 1982. Impairment of voice and face recognition in patients with hemispheric damage. _Brain and cognition,_ 1, 185-195.

Van Lancker, D., Cummings, J. and Kreiman, J. 1986. Voice recognition and discrimination: New evidence for a double dissociation. Abstr. _Journal of Clinical and Experimental Neuropsychology,_ 7(6), p.69.

Van Lancker, D. and Kreiman, J. 1986. Preservation of familiar speaker recognition but not unfamiliar speaker discrimination in aphasic patients. In R.H. Brookshire (Ed.), _Clinical aphasiology._ Minneapolis, MN: BRK publishers.

Vihman, M.M. 1982. Formulas in first and second language acquisition. In L.K. Obler and L. Menn (Eds.), _Exceptional language and linguistics_. New York: Academic Press.

Wechsler, A. 1972. The effect of organic brain disease in recall of emotionally charged versus neutral narrative texts. _Neurology,_ Minn., 23, 130-135.

Wepman, J.M., Bock, R.D., Jones, L.V. and Van Pelt, D. 1956. Psycholinguistic study of aphasia: A revision of the concept of anomia.

Journal of Speech and Hearing Disorders, 21, 466-477.

Winner, E. and Gardner, H. 1977. The comprehension of metaphor in brain-damaged patients. *Brain,* 100, 719-727.

Wong Fillmore, L. 1979. Individual differences in second language acquisitions. In C.J. Fillmore, D. Kempler, and W. S-Y. Wang (Eds.), *Individual differences in language ability and language behavior.* London, Academic Press.

Zaidel, E. 1978. Lexical organization in the right hemisphere. In P.A. Buser and A. Rougel-Buser, (Eds.) *Cerebral correlates of conscious experience,* Amsterdam, North Holland: Elsevier.

Zollinger, R. 1935. Removal of left cerebral hemisphere: report of a case. *Archives of Neurology and Psychiatry,* (Chicago), 34, 1055-1060.