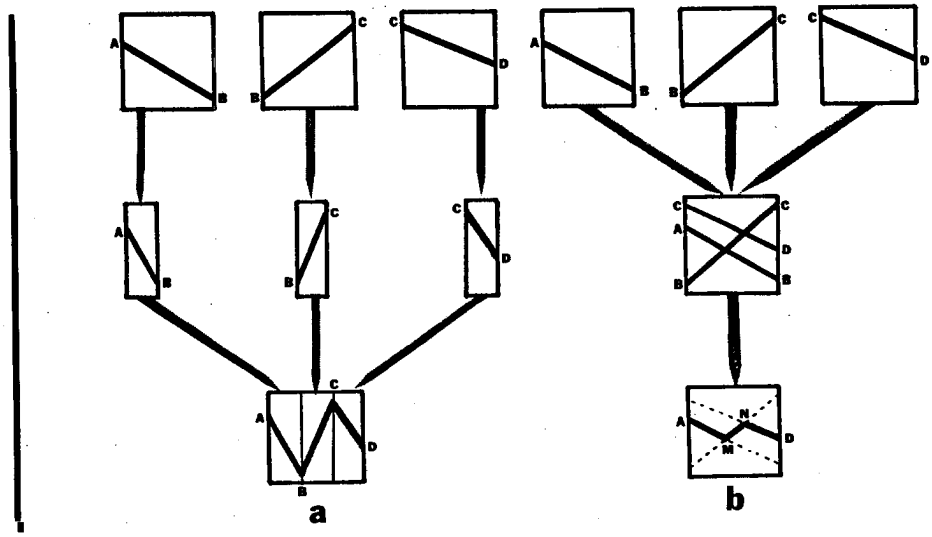
HONGMO REN

ON THE ACOUSTIC STRUCTURE OF DIPHTHONGAL SYLLABLES



ucla working papers in phonetics 65
December 1986

复合元音音节的声学結构

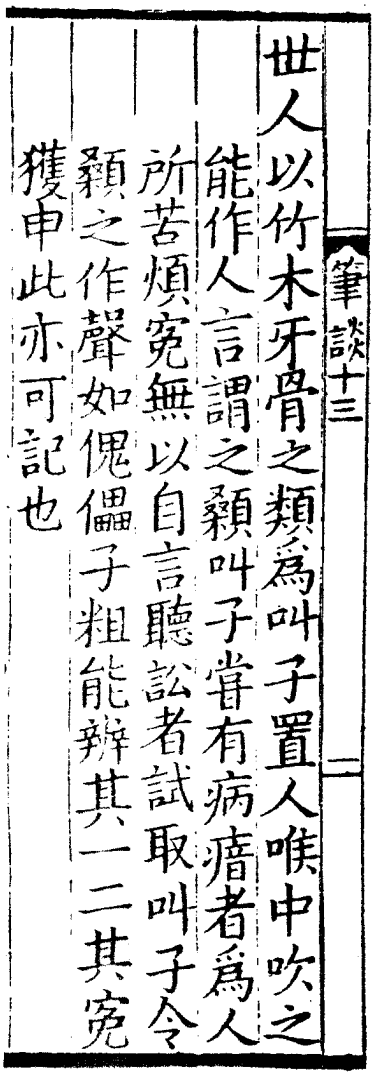
任宏谟 著

加利福尼亚大学（洛杉矶）
语音实验录 · 65

1986 ◆ 12

From an old Chinese book titled "Meng Xi Bi Tan" (Writings in the Mengxi Garden) by Shen Gua (1031-1095 A.D.)

There were people who made a sound generator using materials such as bambo, wood, ivory and bone. When putting it into the larynx and whistling through it, one could produce actual speech. This device was called a voice whistle. There was a man who had become mute due to illness. He needed to argue in court against someone who had bullied him. He was frustrated, because he was unable to present his case. The judge tried to help him by fetching a voice whistle and asking the man to articulate with it. The utterance sounded odd, like puppet talk, but people were able to roughly understand part of what he said. Finally, his sufferance of the injustice was redressed. I found this story worthy of being documented.


筆談十三 二

世人以竹木牙骨之類爲叫子置人喉中吹之能作人言謂之顋叫子嘗有病瘖者爲人所苦煩寃無以自言聽訟者試取叫子令顋之作聲如傀儡子粗能辨其一二其寃獲申此亦可記也

UNIVERSITY OF CALIFORNIA

Los Angeles

On the Acoustic Structure of Diphthongal Syllables

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Linguistics

by

Hongmo Ren

1986

© Copyright by

Hongmo Ren

1986

TABLE OF CONTENTS

LIST OF FIGURES

shifted in phase to maximize overlap). The dotted line on (b) represents the F2 trajectory of /au/ unadjusted for phase.

Figure 4.7. Scheme of two possible realization patterns of F2 in the syllable /au/.

Figure 4.8. LPC formant tracks and spectrograms of /ou/ and /iou/ read by six speakers at a moderate speech tempo. In the LPC graphs, some F2 values are provided for the discussion. The F2 trajectories of the /ou/ are also traced (with x) onto the formant pattern of the /iou/ read by the same speaker. The empty arrow indicates the earliest point where the two F2 trajectories coincide maximally so that in the remainder of the syllable the two F2 trajectories can be considered to overlap.

Figure 4.9. Mean measured values in /ou/ (a) and /iou/ (b), pooled across 6 speakers x 3 speech tempos. The letter X denotes the temporal point where the F2 trajectories in the two syllables are considered to begin to overlap (when shifted in phase to maximize overlap). The dotted line on (b) represents the F2 trajectory of /au/ unadjusted for phase.

Figure 5.1. Correlation between Δ F2 and transition rate.

Figure 5.2. Correlation between F2 transition duration and rate.

Figure 6.1. F2 transition samples of /ia, ua, ai, au/ from Speaker B1. The F2 frequency values (plotted with o) from the speech sample are given without parentheses; the F2 values calculated using Equation (6.1) are given in parentheses. The beginning, ending points and the mid-frequency points of the transitions are marked with a vertical dotted line. The mid-frequency values (mean of fi and fn), durations d1 and d2 are given as well.

Figure 7.1. Two different views concerning the acoustic realization patterns of a syllable consisting of four components A, B, C and D. Suppose that there is no post-nucleus component involved in the syllable.

LIST OF TABLES

ACKNOWLEDGMENTS

I wish first to thank Peter Ladefoged, my committee chair, for his guidance, advice, encouragement and support to complete my research work. He has influenced me directly and in many ways. His scholarship always set an example for me to follow. I have gained greatly and will continue to benefit from his contribution to the field, in which I am only a novice.

Ian Maddieson, a member of my committee, deserves my endless gratitude for teaching me every aspect of conducting the research for this dissertation. Without his guidance and help, which were always available---weekdays and weekends--this dissertation would not have been possible. His comments and discussions of my drafts were crucial to the completion of this dissertation.

I also benefit a lot from the other members of my committee. Patricia Keating gave many comments on this dissertation. Carlos Quicoli read through the dissertation and gave many valuable comments. Michael Goldstein provided suggestions and encouragement during the research.

Besides my committee members, I would also like to express my gratitude to all the teachers who taught me various aspects of the field of linguistics -- Stephen Anderson, Raimo Antilla, George Bedell, William Bright, Susan Curtiss, Gunnar Fant, Victoria Fromkin, Bruce Hays, Edward Keenan, Pamela Munro, Paul Schachter, Tim Stowell and Sandra Thompson.

Members of the UCLA Phonetics Laboratory also helped me in many ways. I would like in particular to thank the following Lab members: Alice Anderton, Norma Antonanzas-Barroso, Margie Chan, Bill Dolan, Susan Hess, Michel Jackson, Jenny Ladefoged, Mona Lindau, Yoko Mimori, Mika Spencer, Henry Teheranizadeh and Wanda White.

Many of my friends provided all kinds of support and help. These friends include: Beverly Afifi, Sean Boisen, Hao Cheng, Jean and Edwin Gillette, Thomas Lee, Rongrong Liao, Feng-hsi Liu, Barbara Magnus, Claudia Reed, Janice Stedman, Mac Thompson, Beverly Trupp and Eric Zee.

Finally, I wish to express my gratitude to the following institutions for their financial support during my graduate years at UCLA: the Chinese Ministry of Education, the Linguistic Institute of the Chinese Academy of Social Sciences, and the National Science Foundation (USA) for grants to the UCLA Phonetic Laboratory.

ABSTRACT OF THE DISSERTATION

On the Acoustic Structure of Diphthongal Syllables

by

Hongmo Ren

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 1986

Professor Peter Ladefoged, Chair

This dissertation seeks to develop a dynamic concept with which to approach the problem of the interface between linguistic transcriptions and physical properties of speech. The materials used are basically syllables containing a diphthong or triphthong in (Standard) Chinese. A set of selected syllables in English is also included. Chapter 2 is a brief introduction to Chinese phonology. In Chapter 3, I argue that the complicated extrinsic rules needed in traditional accounts of diphthongal syllables are only required because of a misunderstanding of how a phonological segment sequence is realized in real speech. I propose a new model--the truncation model-- that can conceptually explain and quantitatively predict the acoustic structure of syllables with complex vocalic components. This model claims that, at speech plan level, all the underlying tauto-syllabic targets corresponding to pre-nucleus element(s) and nucleus element are actually located at the same temporal position, namely, at the syllable initiation. Each phonologically adjacent pair of targets has a

transition of specified rate. The acoustic realization of this type of speech plan is achieved by a programmed process--a truncation process between two phonologically adjacent transitions. That is why the static (sequential) correlations between phonological units and acoustic properties look so complex from a traditional viewpoint. In Chapters 4, 5 and 6, the basic truncation model is modified in order to incorporate further complexities such as the details of the rate specifications for the F2 transitions between adjacent phonological targets. Gay's theory of a constant F2 transition rate under different tempo conditions (1968) and Kent and Moll's theory of "the further the faster" (1972) are tested against the data on Chinese diphthongs. Chapter 7 is a summary of the components of a model which can predict the acoustic structure of diphthongal syllables based on the phonological transcriptions and a given speech tempo.

CHAPTER 1: INTRODUCTION

1.1. Aim and Scope

Linguistic phonetics is a field which provides physical (acoustical, articulatory, and other) specifications for linguistic phenomena by looking at speech behavior. The field bears the brunt of the long-standing problem of the complex relations between linguistic transcriptions and physical properties of speech.

We must acknowledge that "almost every insight gained by modern linguistics from Grimm's Law to Jakobson's distinctive features depends crucially on the assumption that speech is a sequence of discrete entities" (Halle, 1964); not even applied speech research can ignore these entities. Much phonetic research has attempted to characterize this type of linguistic unit in terms of measurable physical properties. Basic to this kind of approach is the belief that the phonetic description of an utterance consists of a linear sequence of physical entities, either articulatory configurations or acoustic patterns. However, we also must recognize that researchers have been relatively unsuccessful in identifying properties of speech which uniquely match the units derived from linguistic analysis. A single acoustic segment of speech waveform often contains information about several neighboring linguistic segments, and conversely, the same linguistic unit is often represented acoustically in the speech waveform in quite different ways depending upon the surrounding phonetic context, speech tempo and higher linguistic levels, as well as the physical characteristics of speakers. Furthermore, speech waveform cannot always be unambiguously segmented into temporally non-overlapping portions corresponding to a linearly ordered sequence of phonological units. The lack of invariance and the difficulties in segmentation in speech have long been recognized as basic problems in the field.

It is unlikely that we can find a comprehensive solution to this problem based on current knowledge. However, by investigating different aspects of language and speech, we can come to a deeper understanding of the complex relations between linguistic units and the physical properties of speech. Taking the acoustic analysis of one sound class, diphthongs, as its topic, the present study seeks to develop a dynamic concept with which to approach this interface problem. A 'downstream' orientation, namely, an approach from linguistic transcriptions to physical properties, will be adopted. That is, we will examine phonetic properties of speech based on phonological transcriptions.

In dealing with the "inappropriateness of conceptualizing the dynamic processes of articulation itself in terms of discrete, static, context-free linguistic categories such as phonemes and distinctive features," MacNeilage and Ladefoged (1976) suggest "a need for new concepts to characterize articulatory function, concepts more appropriate to the description of movement process than of stationary state". An efficient set of articulatory parameters has been advanced, which can account for the contrasts among sounds in one language and the differences among similar sounds in different languages (Ladefoged, 1980). Likewise, there is a set of acoustic parameters which can be manipulated to synthesize sounds in many languages (Liljencrants, 1968; Ladefoged, 1980). Much evidence suggests that parameters of this sort can specify well the sounds used in human languages. With these time-varying parameters, we can orient our research away from the traditional approach of investigating static articulatory positions or acoustic steady states and direct our research toward a dynamic approach for both articulatory and acoustic domains.

For a study of dynamic properties of speech sounds, diphthongs as a sound class are of particular interest, since timing is a critical property of a diphthong. A diphthong is a dynamic linguistic unit, distinguished from many other sounds in languages in that it is dynamic both as a linguistically canonical form and as a physically executed state. Many other sounds may be dynamic only when they are physically executed and are static as a linguistically canonical form. It is also an advantage that diphthongs (or triphthongs) can be characterized by some well-defined and unambiguously observable parameters, such as the F2 trajectory.

There is another reason for choosing diphthongs as the topic. Diphthongs supplement the range of phenomena covered by some well-known models of speech production and coarticulation, such as the articulatory syllable model (Kozhevnikov and Chistovich, 1965) and the coproduction model (Fowler, 1980). The phenomena examined in this study of diphthongs involve complex vocalic sequences in a syllable, and in many cases result from competing or antagonistic articulatory movement.

1.2. Diphthongs

Originally, the Greek word 'phthong' meant 'vowel' or 'voiced sound'; diphthong thus refers to a two-vowel, or two vocalic sound, combination. However, like other classes such as stops, fricatives, liquids and so on, diphthongs behave as a single sound class, even though they may be described in some cases as one event with internal quality change and in other cases as two connected events. For example, some phoneticians (Malmberg, 1963; Mose, 1964; Delattre, 1965; Abercrombie, 1967) define a diphthong as a vowel with continually changing quality; others (Sweet, 1877; Jones, 1922; Hibbitt, 1948; Heffner, 1949; Trager and Smith, 1955; Romeo, 1968) define a diphthong as a sequence of two vowels or one vowel and one semivowel. There are also phoneticians who treat some diphthongs as consisting of one unit and others as consisting of two units. For example, Pike (1949) suggested that diphthongs such as English /i^{I}, u^{U}, e^{I}, o^{U}/ act as phonetically complex single units (single phoneme), whereas diphthongs such as English /a^{I}, a^{U}, $ɔ^{I}$/ function as sequences of two units (two phonemes).

The above statements concerning diphthongs indicate that phoneticians have long been exploring the dynamic nature of diphthongs in terms of linguistic convention. Some of the differences over treating diphthongs as one or two events arise because the linguistic analysis shows that they are either one or two units in the phonology. In other words, whether diphthongs are to be considered single or multiple events is largely related to problems of phonological status as well as phonetic properties of particular diphthongs. Many other factors are also relevant, for example, historical development, prosodic conditions such as those pertinent to stress and tone, and position of the diphthong in a syllable. I would like to emphasize one particular distinction, which has formed the basis of the traditional classification of diphthongs, namely, whether a diphthong is falling or rising. In many (but not all) cases, this classification may have relevance in determining whether a diphthong involves one or two units.

Diphthongs have often been interpreted as consisting of a syllabic and a nonsyllabic element. This provides a phonological classification into two kinds of diphthongs called falling and rising. Falling diphthongs are those in which the syllabic element is the first component. Rising diphthongs are those in which the syllabic element is the second component.

It has been found (Lipski, 1979) that falling diphthongs behave with greater consistency than rising diphthongs in monophthong-diphthong alternations (both synchronic and diachronic). For example, monophthongization from a falling diphthong yields simple vowels with the same lip rounding feature specification as the second element of the original diphthong. The same the backness feature specification of the first element of the original diphthong remains. Finally, monophthongization yields a vowel of intermediate height specification, representing an aperture toward the center of the movement between the two elements of the original diphthong. Similar correspondences can be found in the diphthongization process as well. Typical examples are /au/~/o/, /ai/~/e/, /oi/~/ɵ/, /ui/~/ɨ/, /eu/~/ø/, /ɛu/~/œ / alternations found in many languages (See Lipski, 1979).

However, no such general pattern for the monophthong-diphthong correspondences involving rising diphthongs occurs. The only common process is the /y/~/ju/ alternation. The diphthongs which may alternate with monophthongs in a relatively regular manner are more likely to be one-event units. The diphthongs which do not alternate with monophthongs in a regular way are more likely to be two-event units. This might be true for the distinction of rising vs. falling diphthongs.

One objective of my study is to examine the phonetic characteristics underlying this type of asymmetry between rising and falling diphthongs as it relates to the above noted monophthong-diphthong alternation. Differences in F2 transition rate, the relation of rate to the F2 range and syllabic position, and the sensitivity of rate to speech tempo for the Chinese diphthongs, which are either falling or rising diphthongs, will be tested and discussed in detail in this study.

However, for a phonetic study, it is convenient to treat diphthongs as consisting of two transcription units no matter whether they are one- or two-event units. For our purposes, the concepts of 'target' and 'movement' in describing diphthongs, as used in the definitions provided by some phoneticians are useful as mediating between static linguistic descriptions and the dynamic patterns observed in diphthongs. Lehiste and Peterson (1961) defined a diphthong as a vocalic syllable nucleus containing two target positions. Ladefoged (1975) defined diphthongs as involving a change in quality within one vowel; as a matter of convenience, they can be described as movements from one vowel to another. Target values corresponding to diphthong components, movement between two targets characterized by transition rate or slope and time constant T, as well as the relationship between target values and movement in terms of rate/Δ F2 ratio, will be measured, compared and discussed in this study.

In addition, since the aim of this study is to explore the relations between phonological transcriptions and dynamic properties of speech, using diphthongs as an example, it is natural that a related type of sound, conventionally called a 'triphthong' will be included. The field of study will also include tauto-syllabic components other than diphthongs or triphthongs, if they can (at least partly) be characterized as an extension of the dynamic movement of F2 in the neighboring vocalic patterns. For instance, sounds examined in the test include the initial stop consonant of a syllable, prevocalic glide, and final dark /l/. Thus, we will examine in this study a broad range of phenomena related to diphthongs and triphthongs. In short, the scope of this study can be stated as the syllables with complex vocalic components conventionally called diphthongs or triphthongs, which can be described, as a matter of convenience, as sounds that involve the movement from one vowel target to another.

1.3. Phonetic Studies of Diphthongs

In investigating the phonetic properties of diphthongs, we are faced with the problem of specifying two things: target value and movement. There have been a number of studies focusing on the target values of diphthong components from a traditional and more static view of speech. Diphthong components in real speech have been compared with the simple vowels used to transcribe them to determine whether the conventional segmental description of a diphthong component is compatible with the simple vowel phonemes in the language.

Several studies (Lehiste and Peterson, 1961, Holbrook and Fairbanks, 1962; Wise, 1965) have been conducted to compare the formant values of diphthong components with those of the corresponding single sounds in English. For example, Wise compared the F1 and F2 values of the final part of the English diphthongs /aI, eI, oI/ with /i, I, j/ and those of /aU, oU/ with /u, U, w/. The capital letters /I, U/ represent the lax vowels corresponding to the tense vowels /i, u/ respectively in English. The results showed that the final part of the first three diphthongs is phonetically in the range of /i/ and /I/. It is not /j/. The final part of the last two diphthongs is usually in the range of /u/ and /U/, occasionally /o/. It is not /w/. In Vietnamese, there are three opening diphthongs distinct in backness. They are usually treated phonemically as /ia, ɯa, ua/ (Haudricourt and Thomas, 1967). The acoustic data (Han, 1968) support the phonetic representations /iɛ, ɯɨ, uɤ/. The phonological forms /ie, ua, uo/ were suggested by Han, because they are more compatible with the acoustic characteristics of the simple vowels than the conventional descriptions /ia, ua, ua/.

It has been claimed that the formant values of a diphthong component shift from those of the simple vowels, but such shifts are not consistent in different languages. Several different cases of shifts have been reported. Diphthong components may be centralized from corresponding simple vowels (in Spanish, Manrique, 1979); Both components of all diphthong components may shift in the same direction in the vowel space (in Dutch, Petursson, 1972; Collier et al, 1982). Dispersion can be found in the vowel space for second components of the diphthongs in languages where only some of the vowels can occur in the second component position of diphthongs (in Estonian, Piir, 1983). In English, the first part of a diphthong shows a bigger shift than the second part (Holbrook and Fairbanks, 1962).

These studies reveal important properties of diphthongs and are useful for phonological accounts of diphthongs in these languages. However, the most important works on diphthongs have in fact studied the acoustic and physiological parameters which can characterize a diphthong as a whole. Lehiste and Peterson (1961) were the first to discuss the notion of formant rate of change in Hz/csec. The rates of change of F2 in diphthongs have been examined under different conditions (Gay, 1968, 1970; Kent and Moll, 1972; Manrique, 1979). These studies treat F2 frequency change and tongue movements as the parameters for characterizing diphthongs, and recognize F2 transition rates as one of the most important acoustic features for diphthongs. Note that this feature is a dynamic property of a diphthong as a whole rather than a property of a phoneme-like segment. In general, these studies focus on the dynamic nature of diphthongs. The time-varying acoustic parameters are then used to characterize diphthongs.

1.4. Structure of the Present Study

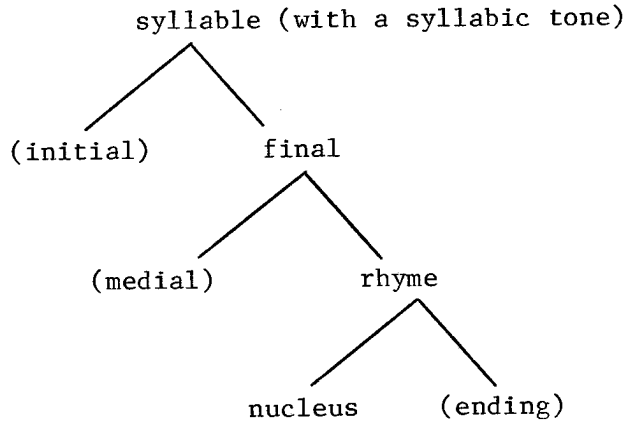
Studies in linguistic, as in all scientific fields, differ in the relative importance which they accord to the description and classification of observable data on one hand and to the development of a formal model on the other hand. On the basis of the model some (ultimately most, if not all) of the observed data become predictable and hence understandable. This study attempts to formulate a model which can predict with fair accuracy the F2 trajectories in syllables with complex vocalic components taking linguistic transcriptions and a specific paralinguistic factor -- speech tempo -- as the basis of prediction.

The materials used in this study are basically syllables containing a diphthong or triphthong in Chinese (Standard Chinese or Beijing Mandarin). In Chapter 2, I will present briefly some aspects of Chinese phonology which are related to diphthongs and triphthongs. The material also includes a selected set of syllables in English. The complex relationship between the linguistic transcriptions and the acoustic properties in diphthongs or triphthongs will be examined. In Chapter 3, I will argue that the complicated extrinsic rules needed in traditional accounts of diphthongal syllables are only required because of a misunderstanding of how a phonological segment sequence is realized in real speech. I will propose a new model--the truncation model-- that can conceptually explain and quantitatively predict the acoustic structure of syllables with complex vocalic components. This model claims that, at speech plan level, all the underlying tauto-syllabic targets corresponding to pre-nucleus element(s) and nucleus element are actually located at the same temporal position, namely, at the syllable initiation. Each phonologically adjacent pair of targets has a transition of specified rate and shape. The acoustic realization of this type of speech plan is achieved by a programmed process--a truncation process between two phonologically adjacent transitions. That is why the static (sequential) correlations between phonological units and acoustic properties look so complex from a traditional viewpoint. In Chapters 4, 5 and 6, the basic truncation model is modified in order to incorporate further complexities such as the details of the rate and shape specifications of the F2 transitions between adjacent phonological targets. These details reflect the most important dynamic features of diphthongal syllables. In these chapters, Gay's theory of a constant F2 transition rate under different tempo conditions (1968)and Kent and Moll's theory of "the further the faster" in F2 rate specifications (1972) will be tested against the data on Chinese diphthongs.

CHAPTER 2. PHONOLOGICAL STATUS OF THE DIPHTHONGS AND TRIPHTHONGS IN CHINESE

Mandarin Chinese is a well-documented language. Although this language has had a very long literary and linguistic tradition and has attracted much attention from scholars around the world, there is no consensus on its phonological system. In this chapter, I will introduce one particular view of the phonological system which will be the basis of the present phonetic study. Some basic aspects of this phonological system will be outlined for the sake of non-sinologist readers. The phonetic observations made in this study are certainly a test for the phonetic forms proposed in the phonological analysis. Some phonetic representations of phonological patterns will be revised based on the acoustic data gathered in this study.

Descriptions of the Chinese phonological system usually consist of two basic components: a traditional analysis of syllable structure and a phonemic analysis (or feature analysis that is heavily based on phonological segment analysis). The analysis of syllable structure is the less divergent part. Most Chinese linguists take for granted that the phonology of Chinese is associated with the following syllable structure:



This hierachical structure analysis is a development of the initial-final-tone structure analysis of Chinese syllables which can be traced back to as early as the sixth century and is still commonly accepted by Chinese phonologists in dealing with dialects. Such a hierachical structure is intuitively appealing, although there is disagreement about the content of each category in the structure for some particular syllable patterns. I will not discuss the initial consonant system in Chinese. Let us focus mainly on vowels in the 'final' part of a syllable (associated with the medial, nucleus, and one of two possible ending types.), which form diphthongs and triphthongs in this language.

There is a great diversity of suggested phoneme inventories for the vowel system in Chinese (Hartman, 1944; Hockett, 1947; Dragonov and Dragonova, 1955; Fu, 1956, Martin, 1957; Xu, 1957; Tōdō, 1963; Cheng, 1973; Hsueh, 1980; You et al, 1980; Xu, 1980; Li and Xu, 1981; Wang et al, 1981; Zhou, 1982, Li, 1984; Pulleyblank, 1986). The divergences are mainly associated with the syllable nucleus, because of different treatments in segmentation and allophonic variations. The difficulties inherent in establishing a phonological system for Chinese lie not merely in segmentation and considerations of parsimony and quality of phonemes, but also in problems of (1) generalization of the syntagmatic distribution conditions, (2) fitting of phonemes into the hierarchical structure of syllables

and (3) formation of natural classes which are sensitive to morphophonemic processes such as r-suffixation widely used in speech. Solving these problems is a different task from a phonetic study of the correlation between the phonological units and phonetic properties. I will not, therefore, judge the different phonological analyses nor propose a new system. In this study I will adopt one phonological system that is commonly accepted and used as the basis of standard romanization of Chinese, the Pinyin system. It should be pointed out that only the phonemic forms in IPA symbols are used in this study. They are different in some cases from the written forms of the Pinyin romanization system because of the pragmatic arrangement of written symbols in the latter. For example, u is used in the Pinyin system for /y/; y is used for /i/ in a syllable without an initial consonant. ao is used for a common phonetic variation of /au/. Table 2.2 is a list of the 35 possible 0-initial syllables in Chinese including the diphthongs and triphthongs with which this study is mainly concerned, written in the transcription I will use. This transcription assumes that there are six vowels in Chinese.

	i	u	y
a	ia	ua	
o		uo	
e	ie		ye
ai		uai	
ei		uei	
au	iau		
ou	iou		
an	ian	uan	yan
en	in	uen	yn
aŋ	iaŋ	uaŋ	
eŋ	iŋ	ueŋ	
oŋ	ioŋ (yuŋ)		

Table 2.2. The 0-initial syllable inventory in Chinese.

Table 2.2 tells us that, in this phonological system, the diphthongs and triphthongs are treated as combinations of two or three vowels. The simple vowels and diphthongs can be followed by a nasal consonant ending. Let us put the simple vowel and nasal ending syllables aside, and just look at the patterns within the panel in the table. Included are 13 diphthong and triphthong patterns. These diphthongs may be medial-nucleus or nucleus-ending combinations. Since nucleus is considered as the syllabic component of a syllable, a medial-nucleus sequence is a rising diphthong and a nucleus-ending sequence is a falling diphthong. The triphthongs here all consist of three component categories, namely, medial, nucleus and vocalic ending. The medials and vocalic endings must be high vowels while the nucleus in diphthongs and triphthongs must be a non-high vowel. We see some gaps in the inventory of the diphthongs and triphthongs. That is, we can find several constraints on possible concatenations of the diphthong or triphthong components:

/y-/ can only occur before /e/; it cannot occur before /ei/ or any other vowel or diphthong.

/i-/ can occur before /a/ and /au/ but it cannot occur before /ai/. It can occur before /ou/ but it cannot occur before /o/. It can occur before /e/ but it cannot occur before /ei/.

Similar constraints apply to /u-/, which can occur before /a/ and /ai/ but cannot occur before /au/ (nor /ou/). It can occur before /ei/ but it cannot occur before /e/.

The diphthong or triphthong formation is not a free combination of vowels. Neither is the distribution of the first vowels consistently determined by the immediately following context. This is also the case for the final vowels in these 0-initial syllables:

/-i/ can occur after /a/ and /e/ but it can not occur after /ia/ and /ie, ye/.

/-u/ can occur after /a/ and /o/ but it cannot occur after /ua/ and /uo/.

One general pattern emerges here, namely that non-adjacent constraints are basic to syllable formation. This is reflected in a general condition on triphthong formation, given in (2.1):

(2.1) Backness disharmony in triphthong formation

$$\begin{bmatrix} V \\ +\text{ high} \\ \alpha\text{ back} \end{bmatrix} \quad \begin{bmatrix} V \\ -\text{high} \end{bmatrix} \quad \begin{bmatrix} V \\ +\text{ high} \\ -\alpha\text{ back} \end{bmatrix}$$

However, this schema permits some additional logically possible but actually non-existing combinations. In order to exlude these, we must specify the features for the second element of a triphthong. The triphthong formation includes the condition that the third element determines the backness of the second one. Thus, (2.1) can be rewritten as (2.2):

(2.2) Triphthong formation

$$\begin{bmatrix} V \\ +\text{ high} \\ \alpha\text{ back} \end{bmatrix} \quad \begin{bmatrix} V \\ -\text{ high} \\ -\alpha\text{ back} \end{bmatrix} \quad \begin{bmatrix} V \\ +\text{ high} \\ -\alpha\text{ back} \end{bmatrix}$$

Thus, we have only /iou/ and /uei/ patterns and the rule (2.2) will rule out /ieu/ and /uoi/. In addition, the /a/ in /iau/ will be in fact /ɑ/ due to the backness in the third component /u/.

The other condition that we need is for the formation of the diphthong patterns. In the diphthongs where there is no low vowel component, the two components must be harmonic in backness.

(2.3) Backness harmony in diphthong formation

$$\begin{bmatrix} V \\ \alpha\text{ back} \end{bmatrix} \quad \begin{bmatrix} V \\ \alpha\text{ back} \end{bmatrix}$$

Therefore, the patterns /io/, /oi/, /ue/ and /eu/ are exluded from the diphthong inventory. The backness of /a/ in /ia, ai, ua, au/ would be in harmony with /i/ and /u/. In addition, we need a special condition for /y/ to be at the first component position to form /ye/ (there is no V+/y/ diphthong in this language).

(2.4) /y/ position constraint

$$\begin{matrix} V & & V \\ \left[\begin{matrix} - \text{ back} \\ + \text{ round} \end{matrix}\right] & & \end{matrix}$$

This study takes the diphthong and triphthong patterns in this phonological system as the basis of its acoustic analysis. The same phoneme-like units as well as the transitions between two members of the same pairs of phoneme-like units will be compared with respect to their acoustic realizations. A model will be proposed, taking the phonological transcriptions of this sort and the speech tempo selections (by means of syllable duration specifications) as the input, to predict realized F2 patterns occurring in speech in Chinese.

CHAPTER 3: A TRUNCATION MODEL FOR ACOUSTIC STRUCTURE OF SYLLABLES WITH COMPLEX VOCALIC COMPONENTS

3.1. Introduction

Many models have been suggested to account for the speech production process and coarticulation, taking linguistic units as the basis. The articulatory syllable model (Kozhevnikov and Chistovich, 1965) and the coproduction model (Fowler, 1980) are perhaps the two most influential ones. There has been a great deal of discussion of them in the phonetic literature. I will not go into the details of these two models. Instead, I will introduce them briefly, with special concern for the definition of the outer bound over which these models apply. Both these models are based on the convention of a consonant-vowel dichotomy. The Kozhevnikov-Chistovich model deals with C_nV syllables where n=1, 2, 3,...; V is [+round]. The Russian syllables /u, tu, stu, ntu, dnu,.../ were examined and the lip rounding of /u/ was found to begin as early as the first consonant in a C_nV began. The authors suggested:

> "In syllables of CV type all the movements of a vowel which are not contradictory to the articulation of the consonant begin with the beginning of the syllable. In other words, it seems possible to formulate a rule that all movements required by a CV syllable are accomplished simultaneously except for those movements which are antagonistic." (Kozhevnikov and Chistovich, 1965. pp 122)

It is obvious that the model covers only a subset of C_nV syllables. It cannot handle the cases where "contradictory articulations" or "antagonistic movements" are involved.

As for the second model, the coproduction model (Fowler, 1980), the situation is quite similar in this regard. Some major ideas in this model can be traced back to much earlier work by Ohman (1966). I will not review the whole model, but only mention one basic point about the notion of coproduction in the model, and that is the idea that vowels and consonants are produced simultaneously in separate "channels". Fowler writes:

> "The production of a consonant or a consonant cluster, then, is imposed on a background of continuous vowel production... The production of an unstressed vowel is superimposed on a trajectory of the shape of a vocal tract from one stressed vowel to another." (Fowler, 1980. pp. 128-130).

As regards this statement of coproduction, Harris (1984) has the following comment:

> "(Fowler's) model does not cover the well known shortening effects of consonants on other consonants (Hawkins, 1973). Perhaps its most serious shortcoming, however, is that it does not deal with competing articulation---the circumstance in which the articulators are constrained during consonant production so that free vowel-to-vowel coarticulation cannot take place." (Harris, 1984, pp. 162)

Fowler's model deals with VCV sequences. However, as Harris indicates, the cases where "competing articulation" occurs cannot be accounted for by the model.

The present study is concerned with syllables having two special properties that seem to be supplementary to those in both the Kozhevnikov-Chistovich and Fowler models.

a) Syllables with complex vocalic components (diphthongal syllables) i.e., CV_n syllables where n=1, 2, 3; V is vowel or glide.

b) Syllables with components involving competing tongue articulations

The materials used in this chapter are from Chinese (Standard Chinese, or Beijing Mandarin) and English. A set of syllables with increasing numbers of components was chosen. We can use the conventional vowel features to represent the tongue movements involved in these syllables as follows, where -> indicates the succession of movements in time.

Chinese:	/ei/'Hey'	/uei/'impressive power'	/tuei/'pile up'
	[-high]->[+high]	[+hi]->[-hi]->[+hi]	[+hi]->[-hi]->[+hi]
		[+back]->[-back]	[-bk]->[+bk]->[-bk]
English:	/el/(L)	/wel/(well)	/dwel/(dwell)
	[-back]->[+back]	[+bk]->[-bk]->[+bk]	[-bk]->[+bk]->[-bk]->[+bk]
		[+hi]->[-hi]->[+hi]	

It is generally observed that the dark /l/ in English is a high back vowel-like sound (Ladefoged, 1975; Keating, 1985). Acoustic data shows that the dark /l/ has very different formant values as compared with clear /l/ (F2=1500 Hz for [l] in [le], F2=870 Hz for [l] in [el] in Bladon and Al-Bamerni, 1976). The formant values for dark /l/ are similar to those for high back vowels.

We can see from the syllable patterns above that the tongue is required to alter both height and backness, sometimes more than once, to produce these complex syllables. We will show that these sets of complex and competing movements are arranged and embedded into a relatively constant syllable duration domain. The structures of these syllables will certainly provide a new perspective for our understanding of many issues conventionally called coarticulation, temporal organization and syllable structure. In this study I will examine the acoustic structure of these complex syllables and test the existing conceptions regarding these syllable patterns; then, I will construct a new model to account for them in terms of acoustic parameters, target specifications and transitions connecting targets. My hypothesis will be presented after a preliminary examination of the acoustic data.

3.2. Procedure

The word lists for Chinese and English are presented above. The Chinese words were randomly scattered in a much longer reading list for a larger study. The list was read by six native male speakers of Beijing Chinese. All the syllables were read at slow, moderate and fast speech tempos, in a carrier sentence 'tʂɤ ʂɿ --- tueima' (This is --- right?) with a declarative intonation. The test words are all syllables with high level tone (tone 1). There are a total of 18 tokens for each word (6 speakers x 3 tempos).

The word list in English was read by 4 native speakers of American English, all phoneticians or phonologists, who were not aware of the purpose of this study at the recording sessions. The words were read once in isolation and once in a carrier sentence 'Say --- again'. In each of these two contexts, the words were read at slow, moderate and fast speech tempos.

The recordings were spectrographically analyzed on a Kay 7800 digital sona-graph with a 0-4000Hz frequency range. An LPC formant analysis was also conducted using the WAVE signal processing program on a PDP-11 computer. Sampling rate was 10,000 Hz. The formant values were calculated and plotted, using 12 coefficients (for some high back vocalic components, 16 coefficients were used instead), a window of 25.6 msec and a step size of 10 msec.

Here the focus will be on the second formant, for many reasons. When patterns of vocalic sequences occur in pairs in a language, e.g., ai/au, ia/ua, ei/ou, iu/ui, iau/uai..., we see very similar F1 but completely different F2 trajectories. That is, the diphthongal patterns can be uniquely specified by the F2 trajectory but not by the F1 or F3 trajectories. Furthermore, the acoustic data for English (Gay, 1968) and Spanish (Manrique, 1979) diphthongs show that the F2 transition rate is one of the basic features of the acoustic patterns of diphthongs.

In our data the F2 values maximally corresponding to the phonological units in a syllable were measured as the realized target values (whether considered properly realized or not). The dynamic property of the F2 trajectory is of central interest in the literature on diphthong studies, though the way to represent F2 trajectory quantitatively differs among investigators. Some investigators use a straight line (slope) representation of F2 transitions (Gay, 1968; Manrique, 1979; Dolan and Mimori, 1986; Toledo, in preparation), others use a curve representation of F2 transition (Rabiner, 1968; Fujisaki, 1980; Yang and Cao, 1982; Lindau, 1985;). It appears that F2 trajectories do not show a constant pattern that can be represented by a unique mathematical formula. It is not even clear if any single non-linear formula would work better than a straight line representation for all F2 transitions found in speech materials. In addition, it may be the case that the exact shape of formant transitions is not important in speech perception. In discussing how F2 transitions are implemented in a speech synthesizer, Holmes (1983) remarks:

> "For reasons concerned with the mathematical process involved (in formant structures), a parabolic rather than linear interpolation was used (in the speech synthesis device), which undoubtedly makes formant tracks look more realistic, but has been found to be subjectively insignificant." (Holmes, 1983).

Based upon the foregoing arguments concerning representations of F2 trajectory, F2 transitions in most of my test materials are represented by straight lines and their rate by slopes (Hz/ms). However, I will also be interested in using an exponential curve to represent F2 trajectories in the discussion of some data (see chapter 6).

3.3 Results and Discussion for Chinese Test Materials

The general patterns of the formant structures of the Chinese syllables /ei, uei, tuei/ can be seen in Figure 3.1. where the LPC formant trackings of these Chinese syllables read by Speaker B1 at a moderate speech tempo are given. Only F2 frequency values are marked in this figure. The capital letters are used to denote the points maximally corresponding to the temporal location of the targets specified for phonological units.

These points are determined largely based on the overall shape of the formant. In /ei/, for example, The target of /e/, E, is the lowest point at the beginning of /ei/; the /i/ target I is assigned at the turning point from a rising F2 transition to an F2 steady state. In /uei/, the /u/ target U is the lowest point at the syllable initiation; The /i/ target I is assigned in the same way as in /ei/. The target for /e/ is supposed to be the turning point E′ from a rising transition initiated at U to another rising transition ending at I. In the majority of /uei/ tokens the above mentioned two rising transitions have quite different rates so that the turning point can be easily determined. In /tuei/, the /t/ target is considered as the highest F2 value at the beginning of the syllable. The /u/ target U′ is supposed to be the valley value in the F2 trajectory. The /e/ target E′′ can be seen at the turning point from U′ to I in Figure 3.1. However, we found a almost straight line F2 trajectory from U′ to I in most tokens. In these cases, no turning point can be found to correspond to the /e/ target.

Figure 3.2 is a synthesis of the means of measured items, including the mean F2 values and temporal positions of the targets as well as the mean of calculated F2 transition rate between each pair of temporally adjacent targets, pooled across 6 speakers at 3 different speech tempos. The standard deviations of these means will be provided in the ongoing discussions. Some dotted lines are marked which will not be discussed until later in this chapter.

Before we look at the measured F2 values, let us recall the assumption we made in the previous chapter that targets corresponding to the phonological units are specified at the speech plan level (plan vs. executor as discussed in Fowler, 1980). The target value for a given phonological unit is assumed to be the same in different environments. Thus, the target value for /e/ in Chinese is the same in /ei/, /uei/ and /tuei/, and the realized value in the simplest syllable, /ei/ (there is no simple vowel syllable [e] in Chinese) can be taken as the representative target value.

Let us first look at the F2 values for /e/ in the two syllables, namely, E in /ei/ (1785 Hz, SD=217) and E′ in /uei/ (2035 Hz, SD=111). The paired t Tests show that E′ in /uei/ is significantly higher in F2 than E in /ei/ (df=17, t=5.631, $p<0.001$). Considering that E in /ei/ is a realized /e/ target and its value is more likely a representative /e/ target value, and considering also that E′ is an ending point of a rising u->e transition and its F2 value has risen higher than that of E, we may treat E′ as an overshot /e/ target.

Let us now compare two /u/ values, U in /uei/ (962 Hz, SD=194) and U′ in /tuei/ (1304 Hz, SD=204). The paired t Tests show that U′ is significantly higher than U ($t=9.527$, df=17, $p<0.001$). If we regard U in /uei/ as a realized /u/ target and its F2 value as more likely to be a representative /u/ target value, we may call U′, which is an ending point of a falling t->u transition, an undershot target.

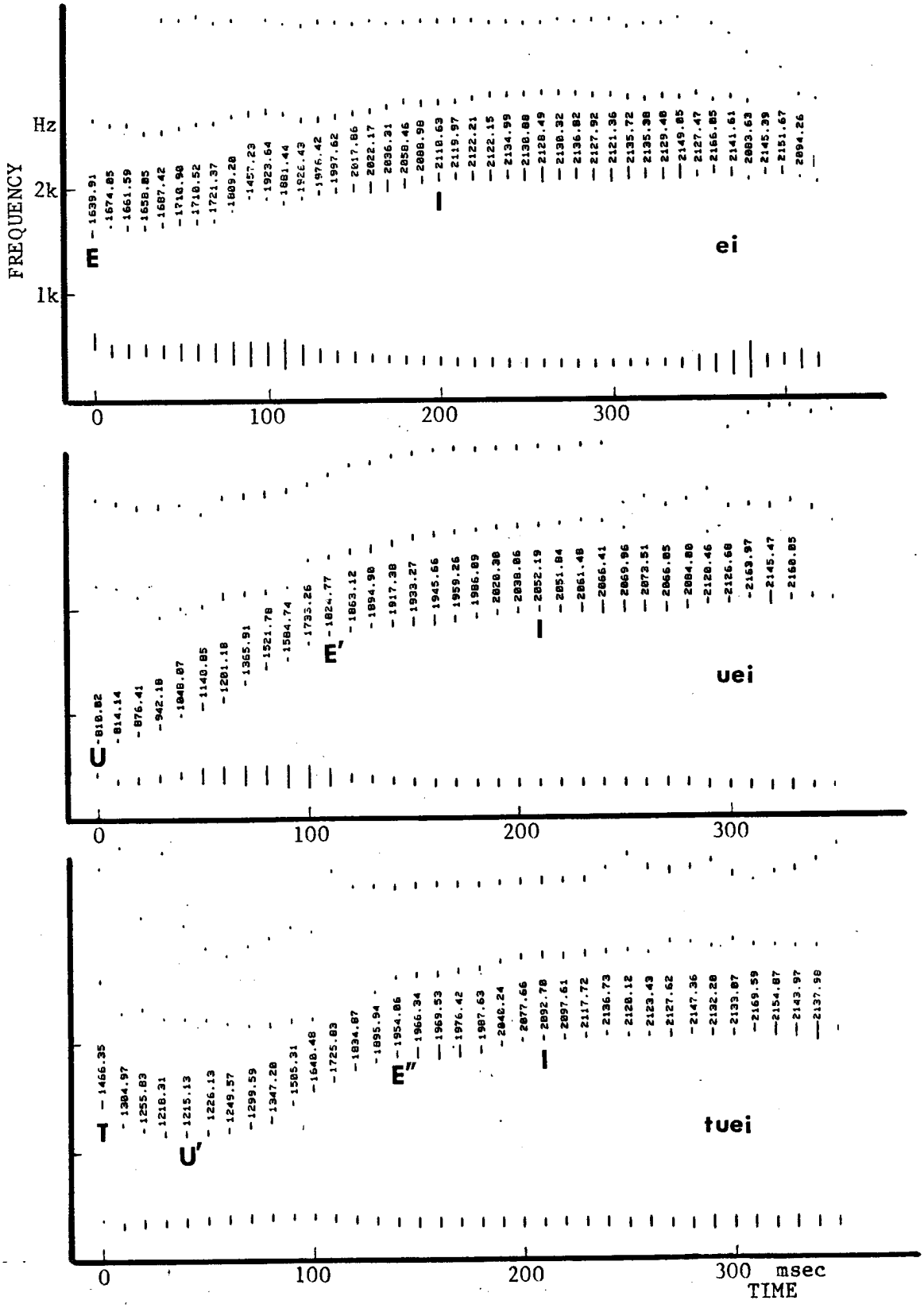


Figure 3.1.LPC formant trackings of /ei/, /uei/ and /tuei/ in Chinese, read by Speaker B1 at a moderate Speech tempo.

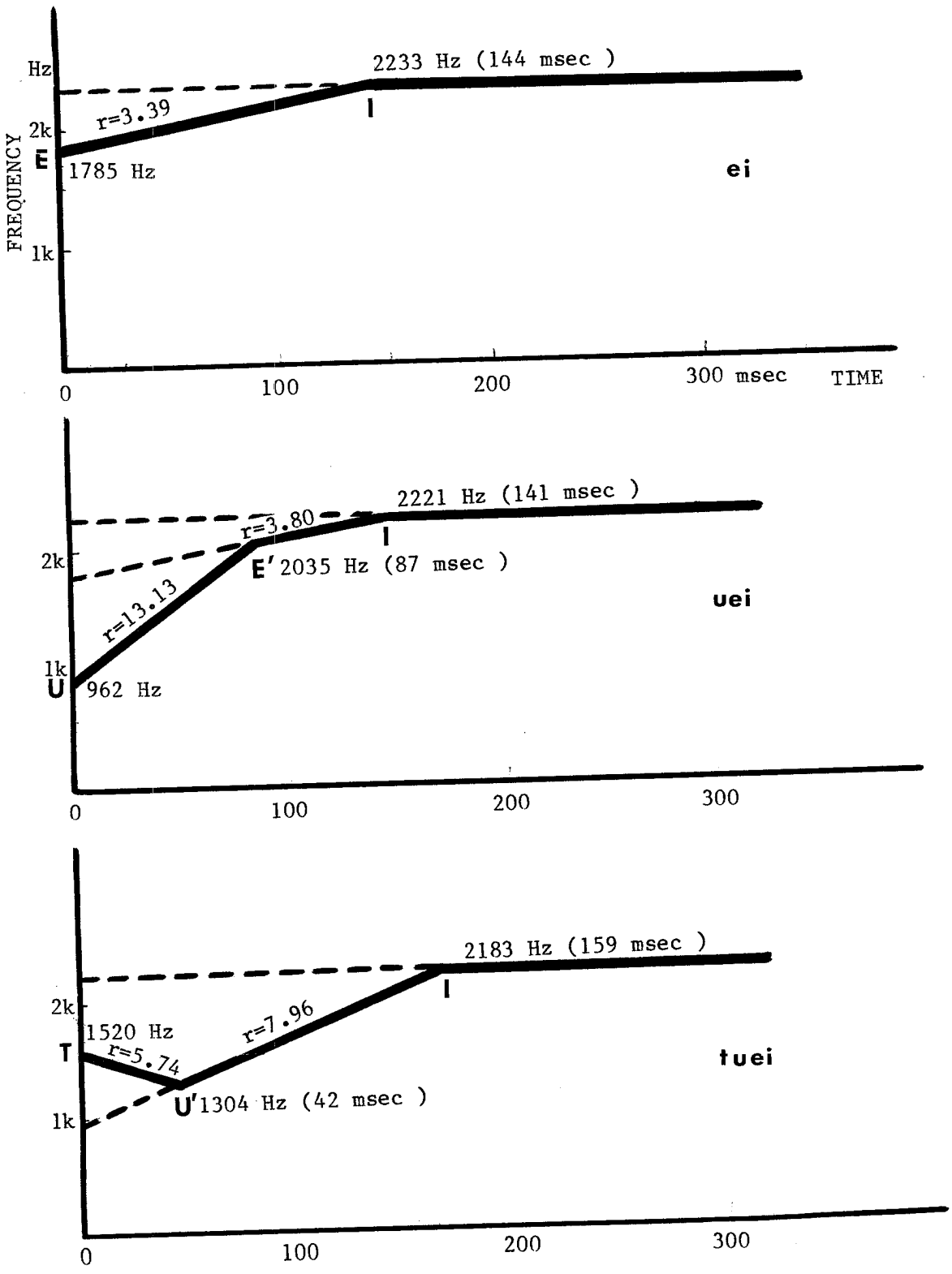


Figure 3.2 Mean F2 values and F2 transition rates in /ei/, /uei/ and /tueɪ/ in Chinese, pooled across 6 speakers x 3 tempos.

The /i/ values in /ei/ (2303 Hz, SD=192), /uei/ (2221 Hz, SD=153) and /tuei/ (2183 Hz, SD=129) have been examined as well, but no significant difference was found among them (t=0.607, t=1.763, t=1.419).

Furthermore, there is no significant difference in transition rate (t=0.600) between the e->i transition in /ei/ (3.39 Hz/msec, SD=1.85) and that in /uei/ (3.80 Hz/msec, SD=1.80)B.

As mentioned above, in two-thirds of the /tuei/ tokens, the u->e and e->i transitions straightened into one single transition. The resulting rate (7.96 Hz/msec, SD=3.39) is slower (t=5.563, df=17, $p<0.001$) than that of the u->e transition in /uei/ (13.13 Hz/msec, SD=4.49). This change is optional and must take place at the phonetic interpolation stage which is later, rather than the phonological level; otherwise, for reasons which we will elaborate on later (Kent and Moll's hypothesis of "the further the faster"), we would expect its transition rate to be even faster than the u->e transition.

The F2 patterns illustrated in Figure 3.2 raise some very interesting questions. These questions are:

Why is one target overshot (e.g. /e/ in /uei/) while the neighboring targets are not? Why is this the case even in fast speech, where targets are supposed to be undershot for lack of time?

Why is one target undershot (e.g., /u/ in /tuei/) while another tauto-syllabic target is overshot (/e/) or maintains a long steady state (e.g., 160 msec steady state of /i/ in /tuei/)? Why does the undershooting occur even at a very slow speech tempo, where targets are supposed to be properly reached since there is plenty of time to do so?

Let us consider how we can model the F2 patterns in syllables such as those illustrated in Fig. 3.2. and answer the above raised questions within the existing research framework. The standard linguistic theory takes for granted that speech is a concatenation of targets which are ordered in a sequence corresponding to the sequence of phoneme-like units at the phonological level. One type of model possible within this framework assumes that the steady state duration of each target is fixed. Given that the positions of two relevant targets are determined by the phonological ordering and the steady state duration of the two targets, the rate and duration of a transition between them can be predicted. This model accounts for the F2 pattern in /ei/ quite well. Assuming that the syllable final target /i/ has a fixed duration of 180 msec and the /e/ target has no steady state, the e->i transition is automatically realized when we put the /i/ target duration at the end of the syllable. However, this model is unable to predict the temporal position of /e/ in the triphthong /uei/ where there are two targets which have no steady state duration, namely, /u/ and /e/, though the /i/ steady state seems to be intact. Let us improve this model by assuming that the final /i/ steady state is fixed so that the two first targets would be located evenly in the remainder of the syllable so as to maximize the temporal separation of the targets in the syllable. In other words, the /u/ target would be located at the syllable onset and the /e/ target would be located midway between the /u/ target and the /i/ onset. Then the two transitions are realized automatically. The measured durations for the u->e and e->i transitions in /uei/ are 87 and 54 msec. Even if we consider these durations roughly equal, some other facts, such as the /e/ target overshoot remain unexplained by this type of model.

Another type of model might assume that the durations of transitions, as well as target values, determine the acoustic realization patterns. In /ei/, for example, if the duration of the e->i transition is fixed as being 150 msec, the remainder of the syllable is automatically realized as a steady state of /i/. However, the e->i transition duration is much shorter in /uei/ than it is in /ei/. Hence, rules which vary the transition duration will be required. However, this model fails to explain why the e->i transition duration should be shortened while the final steady state for /i/ remains relatively constant. Moreover, the /e/ overshooting in /uei/ cannot be accounted for in this model without extrinsic rules.

A different model might assume that transition rates determine the acoustic realization patterns. The fact that the e->i transitions have a constant rate in /ei/ and /uei/ seems to support this model. Given fixed rates for the t->u, u->e, e->i transitions of 5.7, 13.1 and 3.4 Hz/msec, respectively, it would require typical transition durations of 96, 65, and 118 msec, respectively, to travel from one target value to another (assuming approximate target values of 2200 Hz for /i/, 1800 Hz for /e/, 950 Hz for /u/ and 1500 Hz for /t/). Thus, if a /uei/ syllable has a duration of 183 msec or longer, we might expect the target values to be achieved since it only requires 65 msec for the u->e transition and 118 msec for the e->i transition. Similarly, a /tuei/ syllable of 280 msec. (96 + 65 + 118 = 279) or longer would allow plenty of time for properly realized targets. Unfortunately, the actual acoustic patterns are not as simple as this model predicts. Even in very long syllables at slow speech tempo (with syllable durations ranging from 300 to 500 msec), some targets deviate greatly from their standard values, while others may sustain a steady state.

Obviously, then, none of these assumptions taken alone will answer the criterion of observational adequacy. A serious model of what is going on must incorporate some blend of these assumptions, e.g., specifying steady state durations and transition rates will necessarily mean that transition durations vary. The resulting account would then be open to criticism on the grounds that it would be unnecessarily complex, and lacking in explanatory power.

The weakness of the models discussed above is due, in my view, to a misunderstanding of the nature of sequential concatenation of components within a syllable in speech. I propose, instead, an alternative model for accounting for the same data.

The basic hypotheses for this model are as follows:

(3.1) Although the tauto-syllabic components are sequentially ordered at the phonological level, their targets are temporally concentrated at the syllable initiation (usually, the vowel onset) in the speech plan.

(3.2) In the speech plan, the connection between each pair of phonologically adjacent targets has a specified rate. The rate is largely based on the difference between the two target values involved in the transition. The rate may be slightly affected by the phonological ordering of the two targets, the complexity of the syllable, language-specific characteristics and speech tempo (See chapter 5).

(3.3) To produce a syllable, all planned transitions begin at the same temporal position -- the syllable initiation -- each with the specified rate. The observable acoustic realization pattern results from a programmed process.

That is, the phonologically preceding transition truncates the on-going phonologically following transition.

All these hypotheses interact and I will treat them as three parts of a single general hypothesis or model which I will refer to as the 'truncation model'. The metaphor of an "all-out torch relay race" may help to illustrate how these three parts integrate into a general process. Let us imagine that all the athletes in a torch race line up at different points on the start line, and then start running at the same time, each in a direction calculated so that the paths of the athletes will cross at different places. The torch is held by the first athlete and is given to the second athlete at the moment when the two on-coming athletes' paths cross. The second athlete then gives the torch away only when he meets the on-coming third athlete, and so on. The underlying transitions in a complex syllable are analogous to the paths of the athletes; the realized F2 pattern is analogous to the route followed by the torch as it is handed from athlete to athlete.

Let us go back to the questions raised for /e/ overshooting and /u/ undershooting in examining the results from Chinese syllables /ei/, /uei/ and /tuei/. The reason why /e/ is overshot in /uei/ can be explained easily by the truncation model. According to Hypothesis (3.1), the targets T, U, E and I are all temporally located at the syllable initiation in the speech plan. Both u->e and e->i transitions are rising transitions, with specified rates as assumed in Hypothesis (3.2). The intersection (the point where the u->e transition truncates the e->u transition, as implied by Hypothesis (3.3)) must be higher than both starting points of transition, namely, E and U targets. Since the /u/ overshooting cannot be accounted for by the concepts of any existing theory, this fact is a particularly favorable piece of evidence for the truncation model. I would like to present more materials showing how the u->e transition in /uei/ goes beyond the point corresponding to the /e/ target in /ei/. Figure 3.3 presents the LPC formant tracings in /uei/ syllables read by our six speakers at a moderate speech tempo. The /e/ target values in /ei/ are provided for the /uei/ token of the same speaker. For the F2 trajectories in these /uei/ tokens, no turning point can be found corresponding to the /e/ target as realized in /ei/.

As for the question concerning the /u/ undershooting in /tuei/, the truncation model can also provide an explanation. According to this model, the undershot U' in /tuei/ results from the truncation process between a falling transition (t->u) and a rising transition (u->e). Both begin with the same temporal position --- the syllable initiation. Any possible intersection between a rising transition (increasing in the frequency dimension) and a falling transition (decreasing in the frequency dimension) which have the same temporal starting point must be somewhere in the frequency dimension between the two starting frequency values. Since the onset target value of the second transition (u->e) is the offset target of the first transition (t->u), the truncation at the intersection, which is somewhere between the two onset targets of the transitions, entails the undershooting of the offset target for the first transition and the disappearing of the onset target for the second transition (u->e).

As we can see, the overshooting and undershooting of targets can be easily explained by the truncation model without introducing extrinsic rules. Furthermore, the overshooting and undershooting can not only be conceptually explained but also quantitatively predicted by this model.

Assuming that any two phonologically adjacent transitions can be represented as

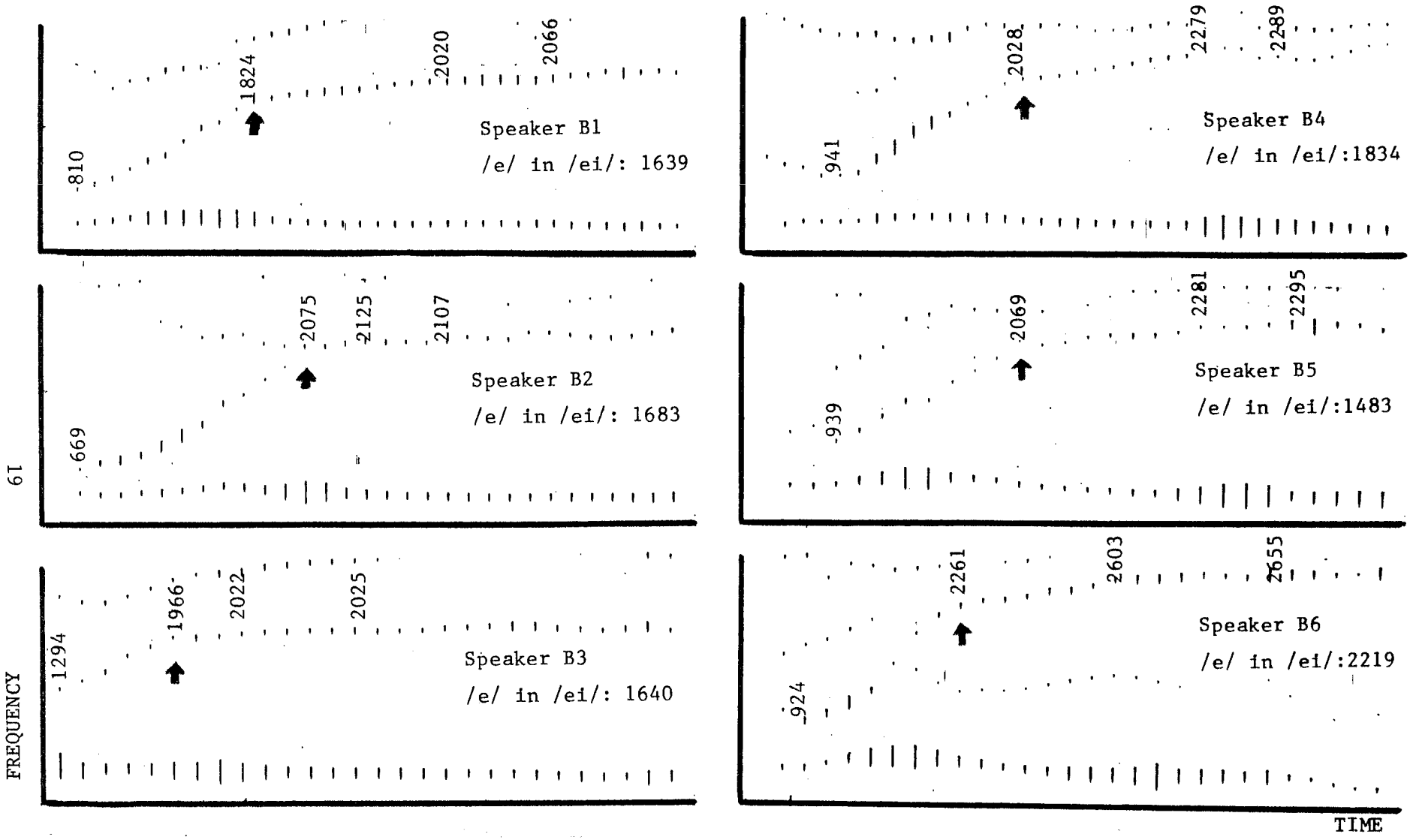


Figure 3.3. LPC formant plots of /uei/ in Chinese, read by six speakers at a moderate speech tempo. Small arrows indicate the overshot /e/ values.

$$(3.4)\quad f(t) = f(i) + rt$$

and

$$(3.5)\quad f'(t) = f'(i) + r't$$

where f(i) is the onset target of a transition, r is the transition rate and t is time. Recall that the undershot or overshot target is the intersection of transitions. The temporal position of the intersection, thus, is

$$(3.6)\quad t = [f'(i)-f(i)] \,/\, (r-r')$$

and the F2 value of the intersection is

$$(3.7)\quad F(t) = f(i) + r\,[f'(i)-f(i)] \,/\, (r-r')$$

The F2 values of the overshot /e/ target and undershot /u/ target calculated from (3.7) are shown in Table 3.1, compared with the mean LPC data. The most significant feature of this type of prediction is that the target value independently measured in the simpler syllable is used as the underlying unrealized target value to predict the F2 pattern of a more complex syllable.

Table 3.1. Predictions of E overshooting and U undershooting

		prediction	LPC data	SD	prediction error
E′ in /uei/	F2	2071 Hz	2035 Hz	111	36 Hz (3% of overall ΔF2)
	time	84.5 ms.	87 ms.	25	2.5 ms.(1% of syl. duration)
U′ in /tuei/	F2	1286 Hz	1304 Hz	204	28 Hz (2% of overall ΔF2)
	time	40.7 ms.	42 ms.	20	1.3 ms.(.5% of syl duration)

Note that the prediction errors for the F2 value and temporal position of these overshot and undershot targets are only a small percentage (0.5% - 3%) of the range of F2 variations. Thus, the prediction can be considered to be accurate. As we can see, the truncation model provides an elegant way of accounting for the overshooting and undershooting phenomena in syllables with complex vocalic components in Chinese.

3.4. Results from English Test Materials

The formant patterns in English syllables /el, wel, dwel/ can be seen in the examples given in Figure 3.4. These tokens were read by Speaker BH in a carrier sentence, at a moderate speech tempo. For the F2 trajectory in /el/, it is obvious that the highest point at the syllable initiation and the lowest point at the end of F2 value at the final part of the syllable are the targets for the two phonological components, /e/ and /l/, respectively. The transition between these two targets is very much like a straight line. In /wel/, it is easy to determine the targets /w/ and /l/. The F2 value maximally corresponding to the component /e/, E′, is the peak value of the F2 trajectory. The same treatment is used for the component /e/ in /dwel/, E". The F2 value maximally corresponding to the component /w/ in /dwel/, W′, is that at the valley of the F2 trajectory.

The mean F2 values and their mean temporal positions as well as the mean F2 transition rates are provided in Figure 3.5. Paired T Tests show that E′ in /wel/ (1434 Hz, SD=187) is significantly lower than E in /el/ (1782 Hz, SD=230; df=23, t=7.016, p>0.001.) Considering that the w->e transition is a rising transition and the F2 value maximally corresponding to the component /e/ is lower than the assumed realized target in /el/, we can say that E′ is an undershot target.

E" in /dwel/ (1366 Hz, SD=189) is significantly lower than E in /el/ (1782 Hz, SD=230; t=8.316, df=23, p<0.001.) Therefore, E" is also an undershot target.

E" in /dwel/ is significantly lower than E′ in /wel/ (t=5.564, df=23, p<0.001). That is, E" is undershot more than E′.

W′ in /dwel/ (856 Hz, SD=141) is significantly higher than W in /wel/ (693 Hz, SD=149; t=6.480, df=23, p<0.001). Since the /w/ target is the ending target of a falling d->w transition in /dwel/ and is higher than the realized /w/ target value in /wel/, W′ is an undershot target too.

The undershooting of targets in these English words can be easily accounted for by the truncation model. Since all tauto-syllabic targets are located at the syllable initiation, the model will predict that one rising transition and one falling transition as neighbors (in the sense of phonological ordering) lead to target undershooting (for the preceding one of the two transitions), while two rising transitions or two falling transitions lead to target overshooting. A rising w->e transition and a falling e->l transition result in /e/ undershooting in /wel/ and /dwel/. A falling d->w transition and a rising w->e transition result in /w/ undershooting. We can see that the data in English support the explanation by the truncation model.

Let us now examine the F2 transition rates in these English syllables. The e->l transition rates in /el/ (r=4.66 Hz/msec, SD=2.06), /wel/ (4.04 Hz/msec, SD=1.47) and /dwel/ (3.81 Hz/msec, SD=1.47) are not significantly different in general (t=2.073 for /el/ vs. /wel/; t=0.985 for /wel/ vs. /dwel/), except for a weakly significant difference between that in /el/ and that in /dwel/ (t=0.2612, df=23, p<0.05, the rate is slower in /dwel/ than in /el/). The w->e transition rate in /dwel/ (5.72 Hz/msec, SD=1.68) is significantly slower than that in /wel/ (7.81 Hz/msec, SD=2.46; t=5.206, df=23, p<0.001.) These results indicate that transitions are slower in more complex syllables than in simpler syllables.

Like what we saw for Chinese syllables, the undershot targets in these English words can be quantitatively predicted by the truncation hypothesis. Again, the

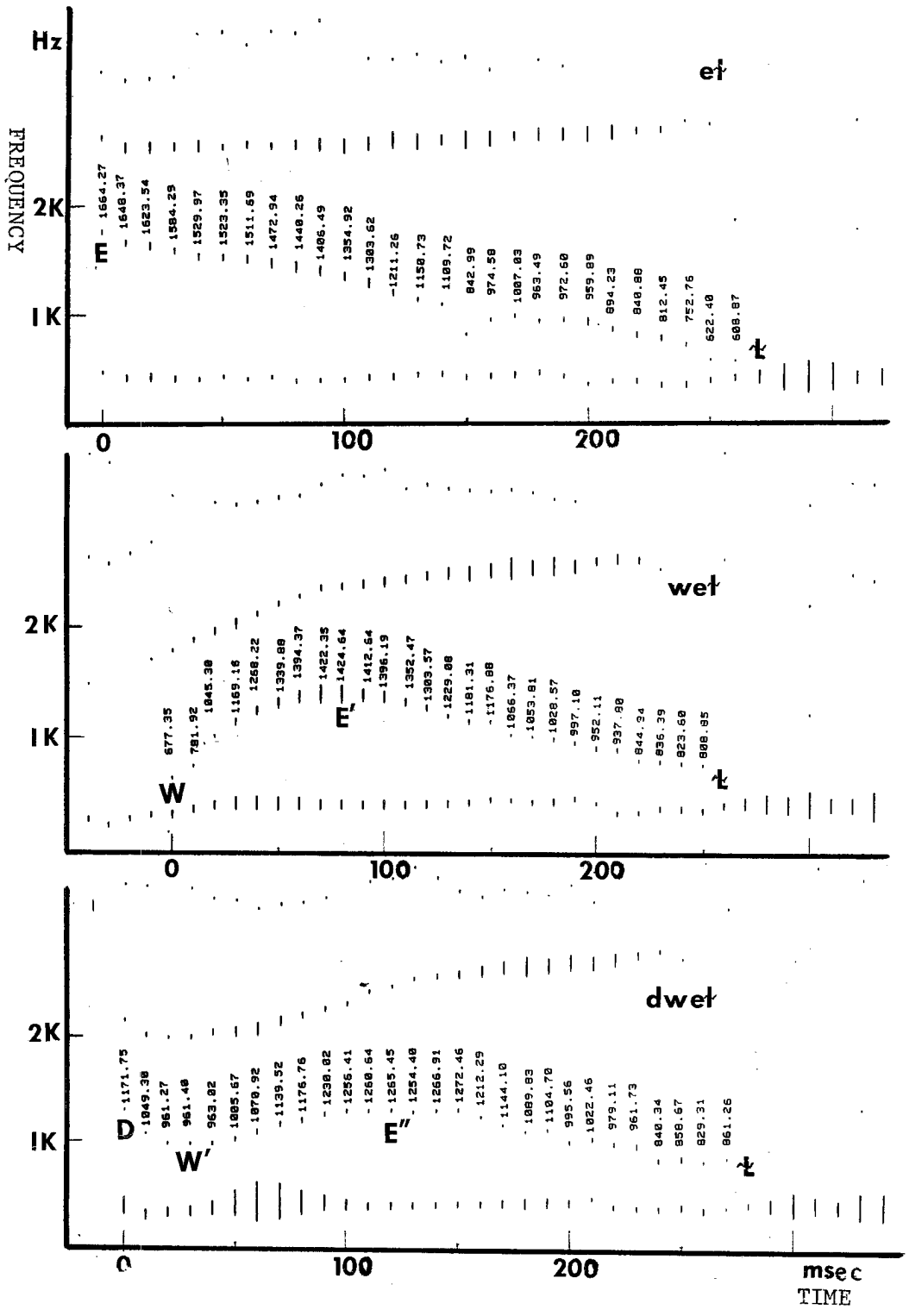


Figure 3.4.LPC formant trackings of /eɬ/, /wɛɬ/ and /dwɛɬ/ in English, read by Speaker BH at a moderate speech tempo.

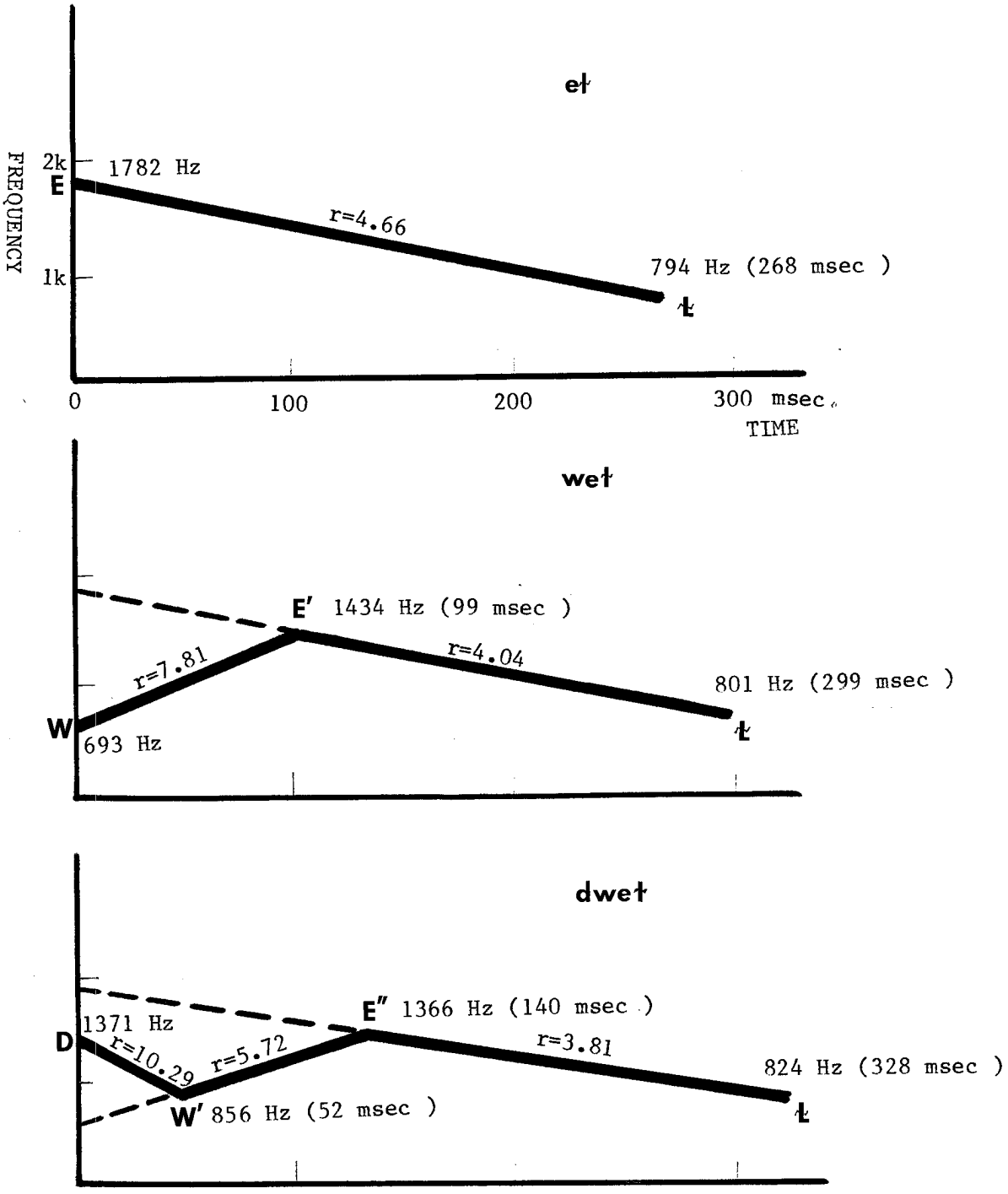


Figure 3.5. Mean F2 values and F2 transition rates in /el/, /wel/ and /dwel/ in English, pooled acro s 4 speakers x 3 tempos x 2 contexts.

crucial point here is to use the independently measured target values in simpler syllables as the underlying target values in the prediction of the F2 patterns in more complex syllables. The calculated E and W undershot values are listed in Table 3.2, where the LPC data are also provided for comparison.

Table 3.2. Predictions of E and W undershooting in English syllables /wel, dwel/ with the data in /el/ as reference

		Prediction	LPC data(SD)		Error of prediction
E′ in /wel/	F2	1375 Hz	1434 Hz	187	59 Hz(5% of overall ΔF2)
	time	87.3 ms.	99 ms.	34	11.7 ms.(4% of syl. dur.)
E" in /dwel/	F2	1293 Hz	1366 Hz	189	73 Hz(6% of overall ΔF2)
	time	105 ms.	140 ms.	52	35 ms. (10% of syl. dur.)
W′ in /dwel/	F2	916 Hz	856 Hz	141	60 Hz(5% of overall ΔF2)
	time	42.4 ms.	52 ms.	16	9.6 ms.(3% of syl. dur.)

The prediction errors for most values are only about 5% of the variation range involved in the syllables and also much smaller than the standard deviations of the mean LPC values. Since the prediction is good, we can see that the truncation model is supported by the data in both Chinese and English.

I have presented the basic aspects of a new model for accounting for syllable structures. It has been shown that, at this stage, the model can predict the F2 patterns in these CV_n syllables with components involving competing tongue movements. The undershooting and overshooting phenomena can be accounted for in a simple way without extrinsic rules being required to reset the target values in syllable internal context. Instead, the prediction is made based solely on the underlying target value independently measured in simpler syllables. Alternative models that can be constructed based on the traditional view of speech as a concatenation of underlying targets distributed in a sequence fail to account for the data in an economical and explanatory fashion. The model covers speech facts supplementary to those reported in Kozhevnikov and Chistovich (1965) and Fowler (1980). That is, the truncation model concerns the acoustic realization of some antagonistic articulations and the C-V dichotomy used in their models is redundant in accounting for the structure involved here.

However, before the truncation model can have adequate predictive power, it needs to be tested against more extensive data. This will reveal that more components must be incorporated into the model. For example, we must account for the factors determining or affecting transition rates. We must also deal with all types of steady state or quasi-steady state in F2 trajectories. The precise temporal position of the syllable initiation must be defined in various syllable types including those with different initial consonants. Issues such as the factors influencing F2 transition rates, the temporal nature of syllable initiation and the possible F2 steady state at the beginning part of a transition, will be discussed in detail in later chapters.

CHAPTER 4. TRUNCATION MODEL FOR THE SYLLABLES /ai/, /uai/, /au/, /iau/, /ou/ AND /iou/

4.1. Introduction

The preceding chapter presented the truncation model for complex syllable structures, based on the analysis of the F2 patterns in /ei, uei, tuei/ in Chinese and /el, wel, dwel/ in English. As we have seen, many models are capable of accounting for the F2 pattern in the diphthong /ei/. Since their predictions do not differ, we cannot test whether one theory is more suitable than the others simply by examining /ei/ itself. However, it could be seen by examining /uei/ and /tuei/ that the truncation theory is the most satisfactory theory for the complex F2 patterns in both diphthongs and triphthongs. The reason is that, in these cases, two or more F2 transitions interact but the targets in one of these transitions can be determined by reference to the corresponding fully realized form in the simpler syllable, namely, the diphthong /ei/. We can therefore test the prediction of the model using separately measured data.

Besides /ei/ and /uei/, examined in the previous chapter, we also have the dipthong/triphthong pairs ou/iou, au/iau and ai/uai in Chinese. A general model should be able to account for all these patterns. Taking one triphthong, /iou/, as an example, the truncation model must make the following assumptions concerning the F2 pattern: the targets for /i/, /o/ and /u/ are all located at the syllable initiation. The i->o and o->u transitions originate from the same position, namely, the syllable initiation. The rates of the two F2 transitions are specified. The o->u transition is phonologically preceded by the i->o transition, and is thus truncated by the i->o transition. A similar analysis can be made for /iau, uai/ to predict as well the F2 trajectories in these two triphthongs.

The present chapter is devoted to testing this type of analysis for /iou, iau, uai/ as related to their diphthong counterparts /ou, au, ai/. To predict the F2 patterns in these triphthongs, the underlying target values and the standard transition rates must be specified by referring to the executed values in the relevant diphthongs. The procedure is similar to that in the previous chapter. The same speakers were used. The word list included diphthong and triphthong patterns which can be grouped into three pairs, in which the members of each pair differed minimally in the presence versus absence of a 'medial'.

/ou/ 'Europe'	/ai/ 'sadness'	/au/ 'boil'
/iou/ 'superior'	/uai/ 'askew'	/iau/ 'waist'

In the following subsections, I will present the measured data and discuss the F2 pattern pair by pair. I will first specify the F2 target values and their temporal positions in the diphthongs based on the LPC formant tracings. The special phenomenon found in the data, the /a/ steady state at the initial part of the F2 trajectory, will be discussed in detail. Finally, the F2 patterns in triphthongs are compared with those in their diphthong counterparts in order to test the truncation process between two transitions involved in triphthong production. A notion of interpolation of transition shape will be incorporated into the truncation hypothesis presented in Section 3.3 of the preceding chapter. Since there are many similar properties among these diphthong/triphthong pairs, the discussion in this chapter will deal mainly with the case of the ai/uai pair.

Only a brief report will be given for the au/iau and ou/iou pairs in order to avoid repetitiveness.

4.2. /ai/ and /uai/

The formant patterns in /ai/ and /uai/ can be seen in the LPC plots in Figure 4.1. These are the examples read by our six speakers B1-6 at a moderate speech tempo. The F2 trajectory in /ai/ is superimposed on the formant structure in /uai/ aligned in phase so that the two F2 trajectories match maximally (determined by eye judgement). In other words, when the F2 trajectory for /ai/ is superimposed on the formant pattern of /uai/, it will not necessary start exactly from the syllable initiation of /uai/. However, the temporal difference between the onset of the two F2 trajectories (in /ai/ and /uai/) does not excede 20 msec. The arrow indicates the earliest temporal point where the F2 of /ai/ and F2 of /uai/ coincide maximally (determined by LPC values) so that in the remainder of the syllable the two F2 trajectories can be considered to overlap.

Let us put the formant pattern in /uai/ aside for a moment and look at the F2 trajectory in /ai/ first. The F2 in /ai/ is quite different from that in the diphthong /ei/ which we examined in Chapter 3. In the case of /ei/, as shown in Figure 3.1, it was clear that the lowest point at the beginning of the diphthong and the point at the end of the rising transition corresponded to the two phonological units /e/ and /i/ respectively. The e->i transition was rather straight there. The whole F2 trajectory in /ei/ could therefore be divided into two parts -- the e->i transition and an /i/ steady state. In /ai/, however, the F2 trajectory consists clearly of three portions. The first portion is a steady state of /a/ (e.g. in Figure 4.1 B1, B4), or an obvious rising portion from a position similar to standard /a/ value (e.g. B2, B6), or somewhere between these two cases, a slightly rising contour (e.g. B3, B5). The next portion is a fast rising F2 transition, reaching /i/ value. The final portion is a possible /i/ steady state. We will define the four points which demarcate these portions as A_0, A_1, I_0 and I_1, and refer to these three portions as A_0A_1, A_1I_0, I_0I_1. Figure 4.2 illustrates the mean values and temporal positions of the turning points of the F2 trajectories in /ai/ and /uai/. These means are pooled over 18 tokens (6 speakers x 3 tempos). Suppose that the syllable final point (at the end of a quasi-steady state of /i/) is influenced by the following cross-syllable context, a point which is not at issue here, we still have three points, A_0, A_1 and I_0 to be matched to the two phonological units in a diphthong /ai/. I_0 is no doubt the /i/ target. However, questions may be raised over interpreting data with regard to /a/. Where is the /a/ target? A_0 or A_1? What is the status of a steady state or quasi-steady state of /a/ at the phonetic or speech plan level? How can an F2 steady state be incorporated into a prediction model of syllable structure?

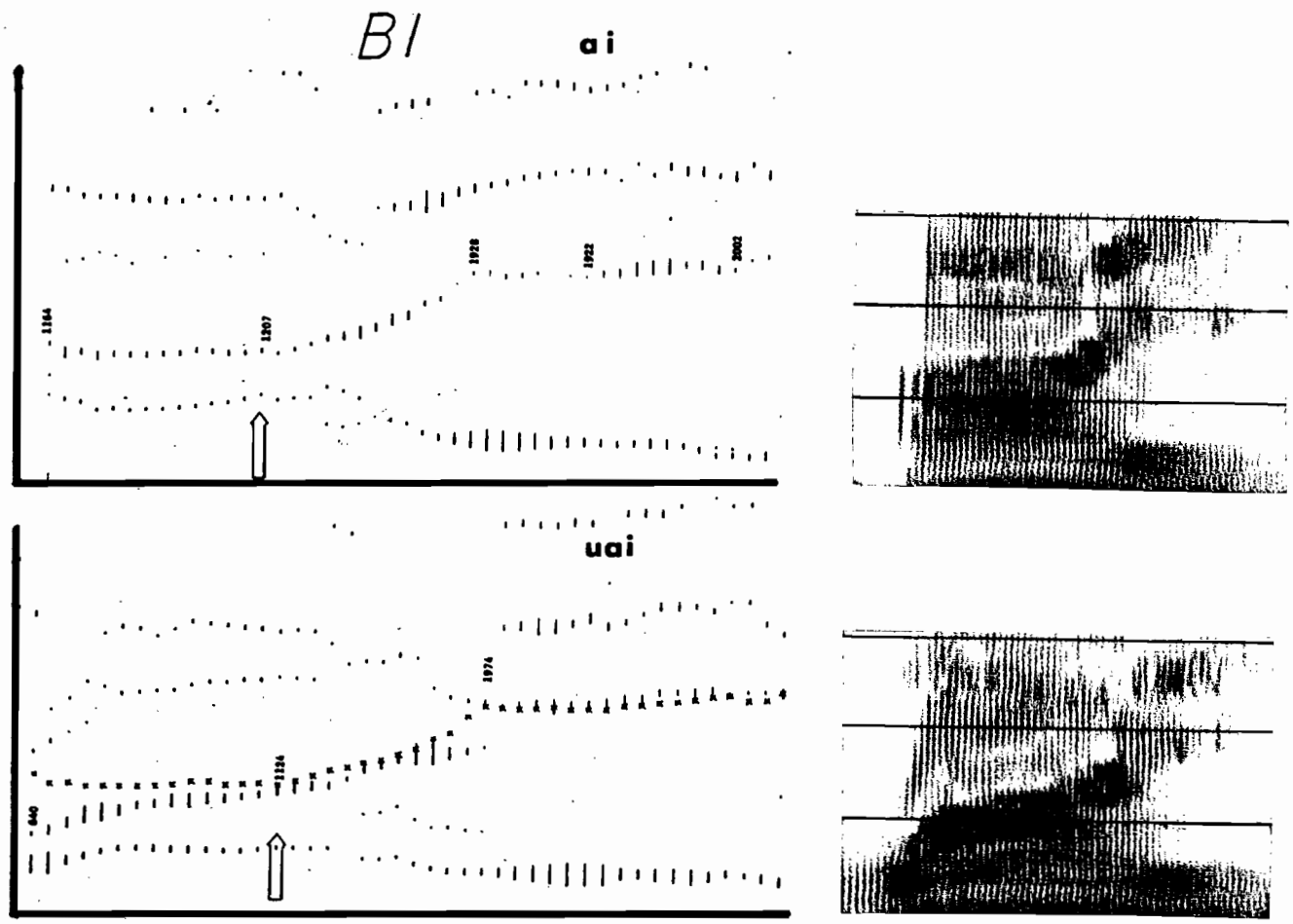
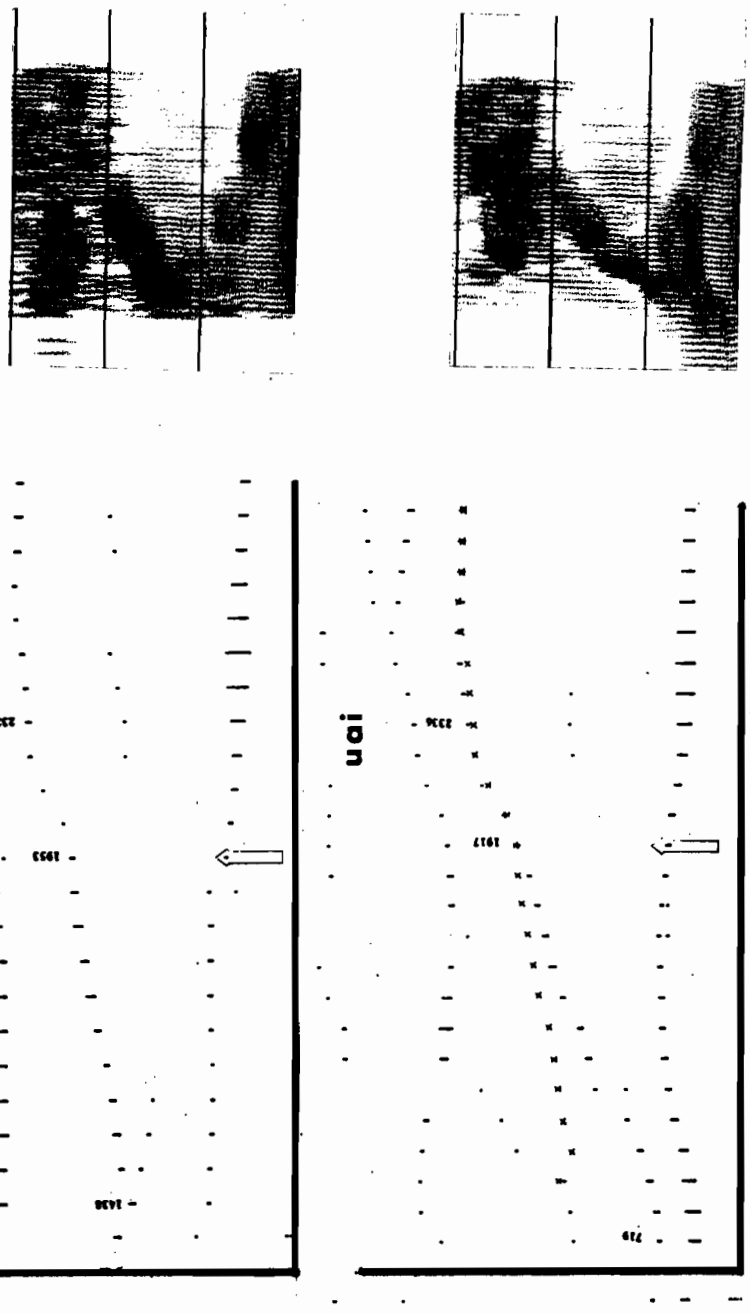
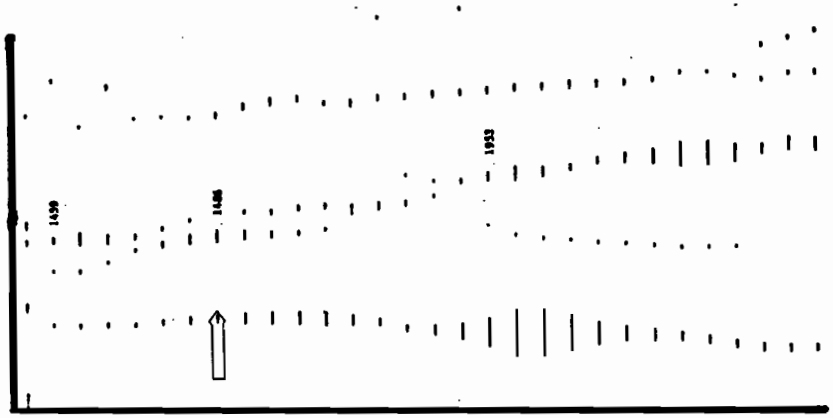
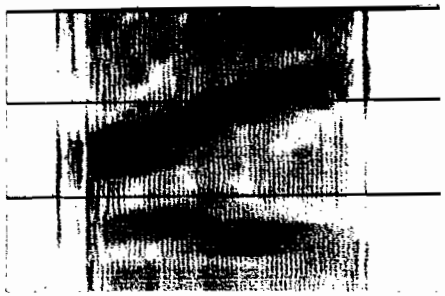


Figure 4.1. LPC formant tracks and spectrograms of /ai/ and /uai/ read by six speakers at a moderate speech tempo. In the LPC graphs, some F2 values are provided for the discussion. The F2 trajectories of the /ai/ are also traced (with x) onto the formant pattern of the /uai/ read by the same speaker. The empty arrow indicates the earliest point where the two F2 trajectories coincide maximally so that in the remainder of the syllable the two F2 trajectories can be considered to overlap.

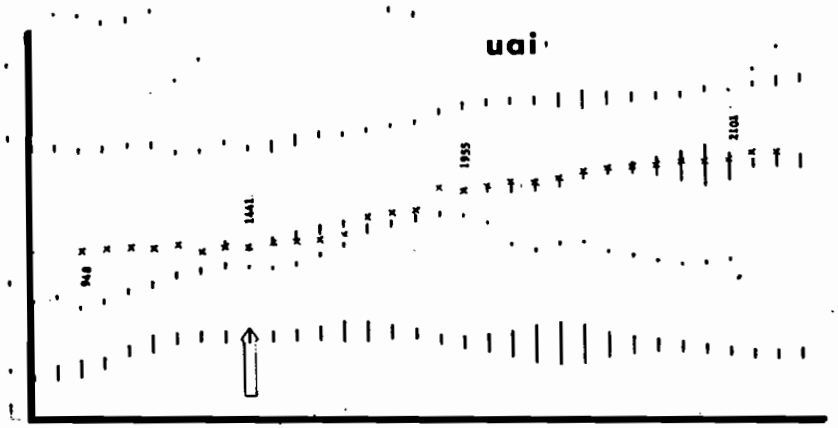

B2
ai
uai

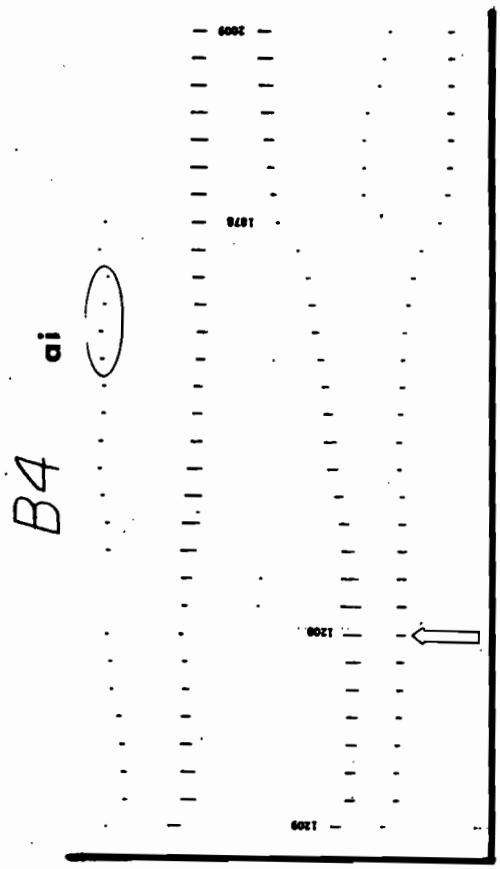
B3 ai

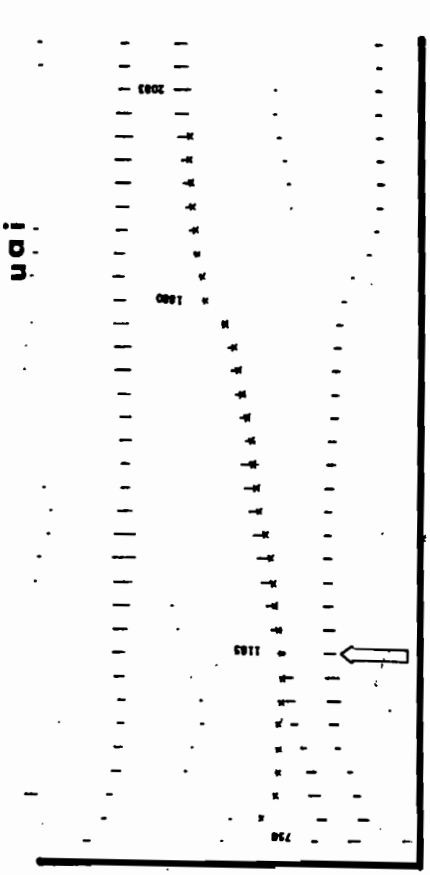


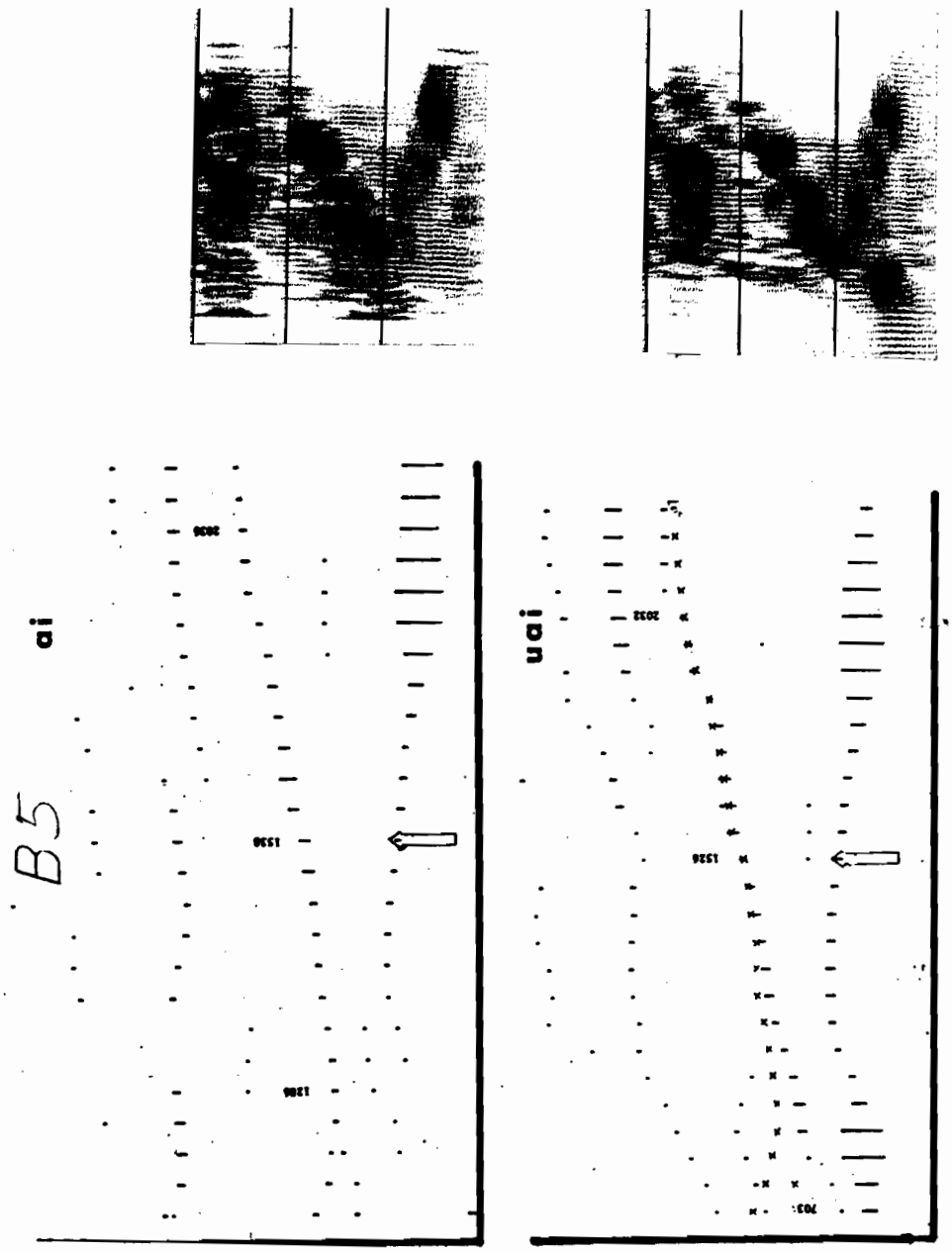


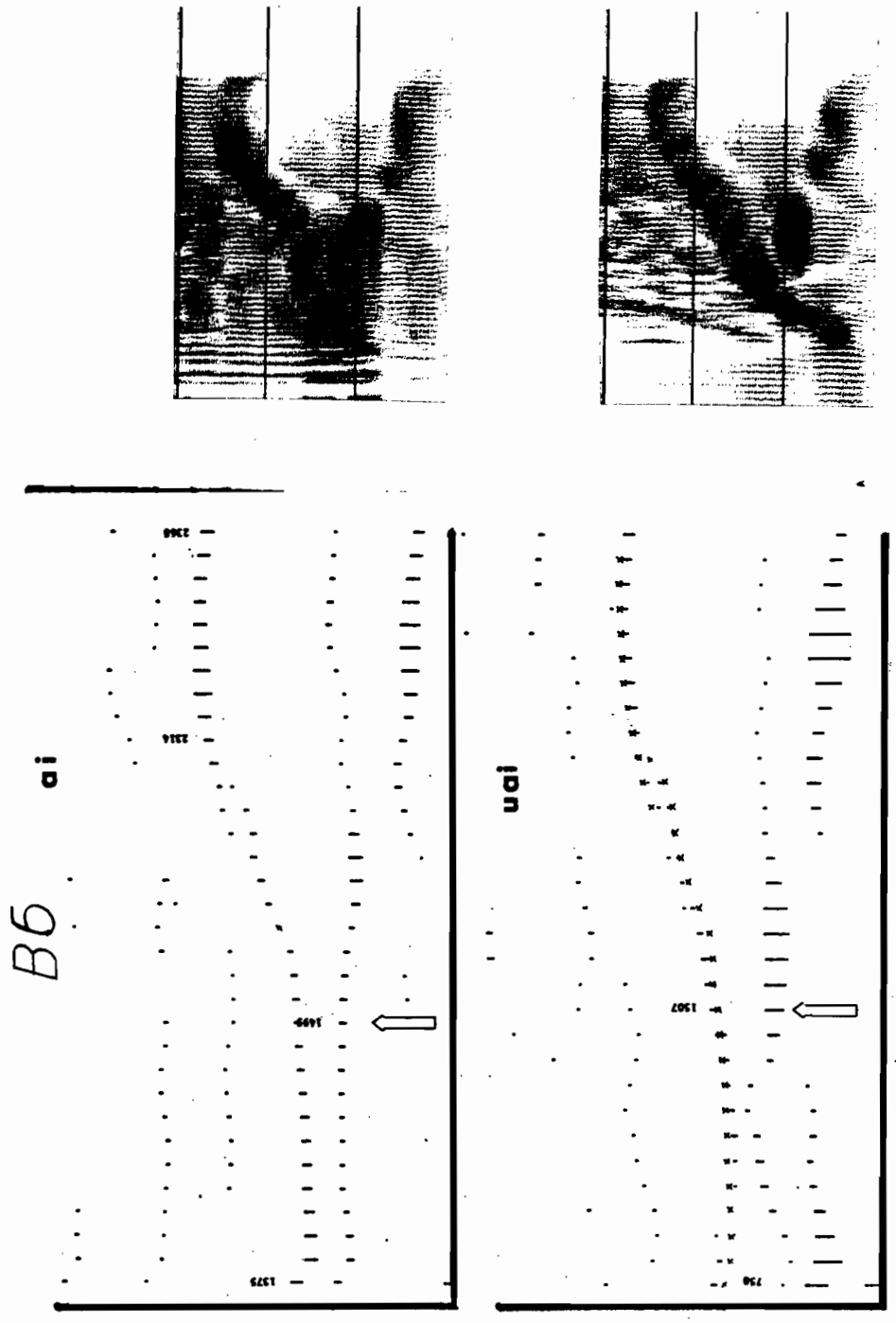
uai




B4
ai
2009
1878
1209
1209


ua i
2003
1600
1165
750


B5
ai
uai


B6
ai
uai
2360
2316
1899
1373
1507
750

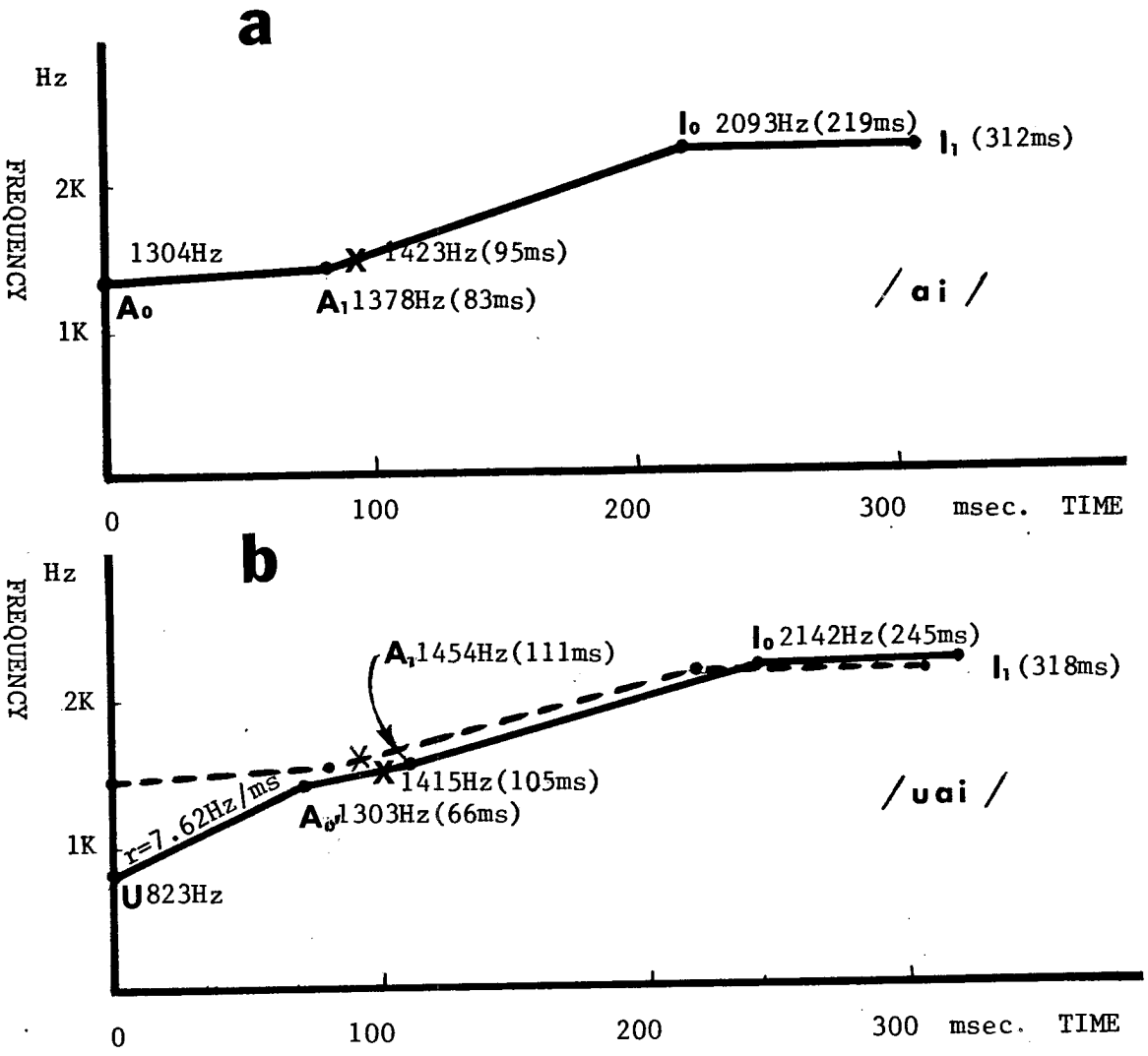


Figure 4.2. Mean measured values in /ai/ (a) and /uai/ (b), pooled across 6 speakers x 3 speech tempos. The letter X denotes the temporal point where the F2 trajectories in the two syllables are considered to begin to overlap (when shifted in phase to maximize overlap). The dotted line on (b) represents the F2 trajectory of /ai/ unadjusted for phase.

In general, there are several types of steady states found in speech materials. One is the F2 steady state corresponding to simple vowels such as that in the syllable /i/. It is assumed in this study that the target of a simple vowel is located at the syllable initiation. Since it is not followed by any other tauto-syllabic target, the simple vowel target is realized as a steady state formant from the syllable initiation to the end of the syllable. If the syllable is followed by another syllable, the steady state will be truncated by a possible anticipatory cross-syllable transition.

Related to this type of steady state is the F2 steady state in the final part of the syllable, such as the /i/ steady state in the diphthong /ei/ (See Figure 3.1). This /i/ steady state results from the simple vowel target /i/ being realized throughout the syllable and truncated by an e->i transition. As presented above, we also find a final steady state or quasi-steady state in /ai/ (and in /au/ too). These final steady states are assumed to be actually the rest of the simple vowel steady states after truncation by the phonologically preceding diphthongal transitions, whatever the transition shape is. The presence or absence of this type of final steady state is determined by the F2 transition rates involved in the syllables as well as the syllable duration, and thus can be predicted by the truncation model.

A third type of steady state is that occurring at the initial part of the syllable, such as /a/ steady state in /ai/. The realizations of an initial and a final steady state are not symmetrical along the time axis. The initial steady state is not as automaticaly realized as the final steady state. At least some basic aspects of the syllable components, such as the temporal realization of the tauto-syllabic target following that steady state and the transition rate between two targets, are influenced by an initial steady state, but none of the component targets would be significantly affected by a final steady state. Furthermore, the occurrence of the /a/ steady state in an /ai/ syllable seems optional rather than obligatory. The duration of an /a/ steady state ranges from barely noticeable length to approximately 150 msec. In /ai/ syllables, it is also possible that the F2 trajectories show a slower rising transition followed by a faster rising transition. This indicates that the /a/ steady state is not stable. There are intermediate states between a straight a->i transition and an /a/ steady state followed by an a->i transition. These are tokens with a slowly rising F2 between A_0 and A_1. This type of ambiguity of steady state, quasi-steady state and transition must also be accounted for in a predictive model.

There are several different ways in which we might consider treating this sort of /a/ steady state at the first part of a syllable in terms of target specification. We might specify the /a/ target as being temporally located at the end of the steady state A_0A_1 (namely, at the onset of the a->i transition), thereby keeping the basic assumption that all targets in speech are timeless points and the real concern is the location of each target and the transition rate between two targets. However, since the /a/ steady state duration varies from 0 to 150 msec, the target at the end of the steady state would be temporally indeterminate. The conceptual problem underlying this difficulty is that one can hardly imagine a temporally mobile target corresponding to the first phonological unit of the syllable. In addition, in /uai/, as shown in the lower part of Fig. 4.1. B5, there would be two realized targets for /a/, which are located at different temporal points. One is at the end of the u->a transition, as an ending target; the other is at the beginning of the a->i transition, as an onset target.

It is by no means certain that this treatment can be accommodated in a predictive model without sacrificing simplicity.

Another way to treat the /a/ steady state in /ai/ is to assume that the steady state is actually intrinsic to the /a/ target. Thus, we have two types of target with regard to temporal nature: one type has 0-time, and the other is elastic, with varying steady state duration. This treatment is compatible with recent research in speech synthesis by rule, which has shown that the targets themelves need to be considered as complex objects exhibiting composition, scope and internal cohesion (Allen, 1984). The conceptual weakness of this approach is that we would have an indeterminate aspect (i.e., the varying steady state duration) as a part of the nature of the target. Furthermore, the situation will be complicated if, besides the transitions between targets, a target itself can be partly (as well as completely) truncated.

A third possible treatment would be to assume that the most variable entity along the time axis of a speech parameter is the transition. We can still keep the unified notion of a target as a timeless element. Transitions are interpolated in various ways. A transition between two given syllable components does not always follow a straight line, because of possible articulatory constraints. Phonologically neighboring transitions may interact in a rather complex way, leading to the observed acoustic patterns. Very often, a turning point occurs in an a->i or a->u transition, thus making the transition look as if it is bent. A possible objection to this treatment of an /a/ steady state as part of the bent a->i transition is that, in many other cases, a complex form is the basic form and a simpler form can be derived from the basic form rather than vice versa.

However, there is a piece of evidence in favor of the approach with the /a/ target at the syllable initiation (namely, the beginning of the /a/ steady state) rather than at the end of the /a/ steady state. We can find some disyllabic words the second syllables of which have an initial /a/ steady state. An example /pʰi˨˩ # au˨˩˦/ 'fur-lined jacket' is given in Figure 4.3. The first syllable has a rising tone (T2) and the second syllable a low concave tone (T3). # is a syllable boundary marker. The spectrogram shows that the F2 steady state of the simple vowel /i/ in the first syllable is truncated by an anticipatory i->a transition, which directs its ending target toward the syllable boundary rather than the end of the /a/ steady state in the second syllable /au/.

The fourth possibility would be the approach which recognizes the /a/ target as located at the syllable initiation. However, the /i/ target is located at a later temporal point. Therefore, The F2 trajectory goes steadily from the /a/ value at the syllable initiation until it meets the underlying /i/ target which is at a certain interval from the syllable initiation. Then, a rising transition occurs toward the realized /i/ target, the temporal position of which is determined by both the /i/ target value and the transition rate. Sometimes, a straightening process takes place so that a straight F2 transition from the /a/ target at the syllable initiation to the realized /i/ target at I_0 occurs instead of an /a/ steady state followed by a rising transition A_1I_0. This type of straightening process has been already seen in Chapter 3. The u->e and following e->i transition are straightened to become one single slower transition in /tuei/.

I will adopt the fourth approach in this study, with full awareness that this approach is only one of the possible interpretations of the data. More specifically, this study treats A_0 as the location of the /a/ target and A_1, namely, the end of the /a/ steady state, as the location of underlying /i/

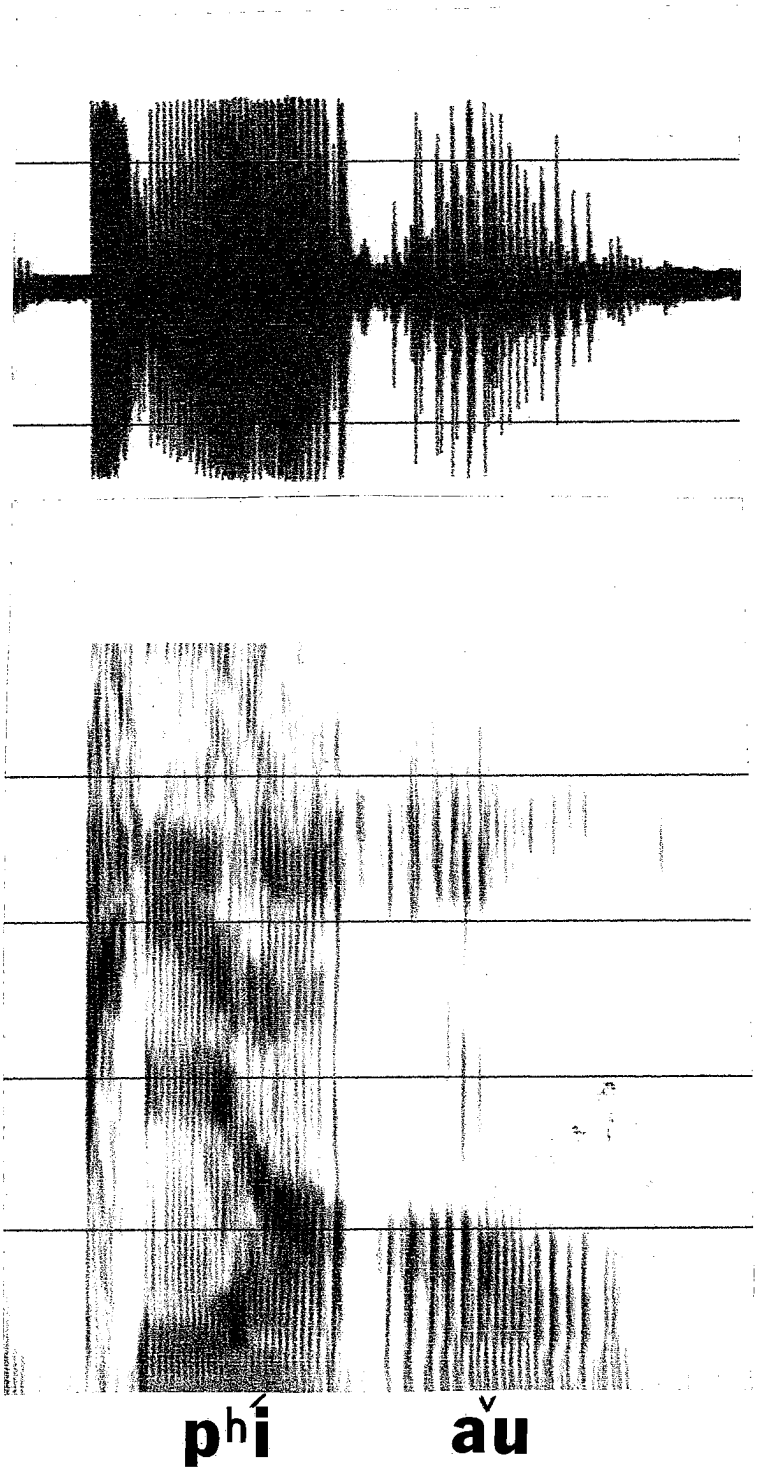


Figure 4.3. Spectrogram and wave form of a disyllabic word /$p^{h\acute{}}$i # a$\check{}$u/ 'fur-lined jacket', read by speaker B1. / is a rising tone marker. v is a low concave tone marker. An anticipatory i->a F2 transition goes toward the /a/ target at the initiation of /au/.

target. Because the /i/ target is not at the syllable initiation, the hypothesis of all the tauto-syllabic targets located at syllable initiation stated in Chapter 3 needs to be constrained. Since /ai/ is a combination of a nucleus element (which is syllabic) and an ending element (which is non-syllabic), we can restrict the hypothesis by stating that the tauto-syllabic targets correspoding to pre-nucleus and nucleus elements are all located at the syllable initiation, but those corresponding to post-nucleus elements are not. This would not conflict with the analysis in Chapter 3, because the final element /i/ in /ei/, /uei/ and /tuei/ in Chinese and /l/ in /el/, /wel/ and /dwel/ in English are considered to be syllabic elements (argument will be provided later based on already observed acoustic patterns and further analysis of the data).

In this approach, the F2 trajectory in /ai/ can be approximated by a derivation as follows. First, let us place the target /a/ (mean value is 1303.83 Hz, SD=126.27) at the syllable initiation. Then, we must specify the temporal position for underlying /i/ target (mean value is 2093.67 Hz, SD=145.50). and the transition rate between these two targets. Because of the so called straightening process involved, the only consistent measurements of transition rate and duration are those of the transition between realized /a/ and /i/ targets (A_0I_0). I will refer to these measurements simply as the rate and duration for the a->i transition. Since the measured rates and durations are correlated with the syllable duration (r=-0.7110 for the transition rate and r=0.8666 for the transition duration) and since the A_0I_0 duration is about 70% of the syllable duration, we can derive the a->i transition rate by a graphic procedure of putting the realized /i/ target I_0 at the temporal point of 70% of syllable duration and connecting A_0 at syllable initiation and I_0. The resulting pattern (also with a final /i/ steady state) is one possible F2 pattern of /ai/, namely, the one without an /a/ steady state because of the straightening process. However, the straightening process is not obligatory. Since the /a/ quasi-steady state duration is also correlated with the syllable duration (r=0.7419), and since the mean duration of this portion is 27% of the mean syllable duration, we can have a possible /a/ steady state and some possible F2 patterns of /ai/ with a probable steady state or rising quasi-steady state as shown in Figure 4.4. Two types of variations are logically possible: one is to change the slopes of the quasi-steady state and the following faster rising portion (Figure 4.4.a); the intermediate states between a straight-forward rising a->i transition and a steady state /a/ followed by a fast rising transition can be accounted for by this type of variation. Another possible type of variation is in the duration of these two portions (Figure 4.4.b); thus, the a->i transition with different steady state durations could be explained by this type of variation. Further studies with more data may be needed to see which type of variations is dominant, or, whether some possible types of variation occur sulmutaneously. An u->a transition is also marked in Figure 4.4, but will be discussed later in this section.

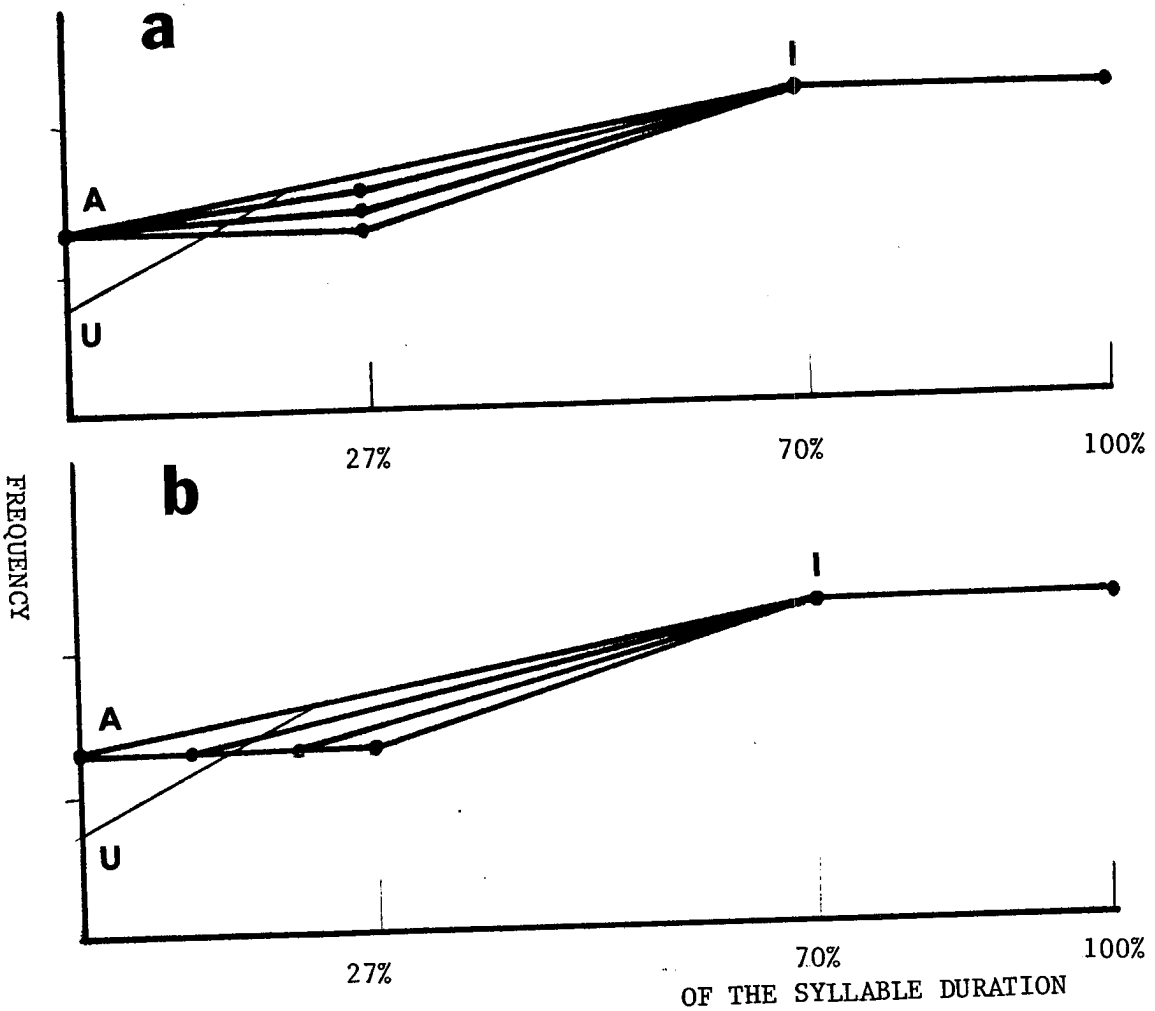


Figure 4.4. Scheme of two possible realization patterns of F2 in the syllable /ai/.

The F2 patterns of the two extreme cases of /ai/, namely, the straight and gradual rising transition from the syllable beginning and the transition with an /a/ steady state followed by a fast rising transition, are very much like the F2 patterns of the genuine diphthong /ɛi/ (with gradual rising transition) and the pseudo diphthong /aj/ (with a steady state followed by a rapid F2 transition) in Dutch (Collier et al, 1982, Fig. 1, 2, 5). It was reported by that study on Dutch that the activity of the muscles responsible for tongue fronting (genioglossus) and backing (styloglossus) indicates less horizontal tongue movement for the genuine than for the pseudo diphthongs. The genioglossus is moderately active throughout /ɛi/ while the styloglossus exerts almost no backward pull. However, early peaks of styloglossus activity and late peaks of genioglossus activity were found in /aj/, indicating fronting of the tongue from a backed position. It was thus suggested that the genuine and pseudo Dutch diphthongs differ in the dynamic property of tongue advancement between the first and second elements. Though the difference between the two patterns is phonologically distinctive in Dutch but not in Chinese, it is possible that the phenomenon in Chinese can be ascribed to the same articulatory mechanism. The two variations in F2 patterns of /ai/ may be due to the free variations in the dynamic property of tongue advancement between the two components /a/ and /i/. A gradual and slow tongue fronting movement results in a straightforward a->i F2 transition from the beginning of the syllable while a backing movement followed by a relatively fast fronting movement leads to an /a/ F2 steady state or quasi-steady state followed by a fast rising F2 portion.

Let us return to the ai/uai pair. With the F2 trajectory in /ai/ fixed, the F2 in /uai/ can be predicted by the truncation model. That is, the /a/ steady state or the a->i transition, whatever its overall shape is, is truncated by a preceding u->a transition. The examples in Figure 4.1 illustrate this type of truncation process. We can see in /uai/ tokens that their F2 trajectories and the traced F2 (marked by x) of the corresponding /ai/ tokens (of the same speaker and at the same tempo) only differ noticeably in the first 1/4, 1/3 or 1/2 of the duration. The arrow designates the earliest point where the F2 trajectory of /ai/ and that of /uai/ coincide maximally. Starting from that coincident point, The two F2 trajectories almost overlap in the remainder of the syllable. The mean F2 values of this maximally coincident point are provided in Figure 4.2 (marked by X). The two X's are not at the same temporal position in Figure 4.2 because of the slight temporal expansion (by 0-30 msec) in the first part of the syllable /uai/ due to the movement being complicated by an additional u->a transition. That is why the second half of the F2 trajectory in /uai/ (solid curve) lies slightly behind in phase position from that of the F2 trajectory (dashed curve) in /ai/. The mean rate of the u->a transitions is also provided in this figure. Two turning points related to /a/, A_0, and A_1, can be found in the /uai/ F2 trajectory. The point X is different from both these two A points in F2 value and in temporal position. That is, the coincident point of the F2 in /ai/ and that in /uai/ is not located at the intersection between a straight line u->a transition and a straight line /a/ quasi-steady state or a->i transition. Instead, the two F2 trajectories seem to meet slightly later in time and higher in frequency value. This tendency can be seen throughout all the ai/uai pairs in Figure 4.1. Taking B1 as an example, the u->a transition goes up relatively fast toward the /a/ F2 value until about 80 msec, then changes rate and moves almost parallel with the F2 trajectory of /ai/ for some 60 msec. This phenomenon is very much like the 'buffer' period in a relay race where the two athletes keep company in order to pass the baton smoothly. For a more accurate truncation model, therefore, we may need a local smoothing process for the period where the truncation between two phonologically adjacent transitions takes place.

As a summary of the data shown in Figure 4.1-2, the fact that the F2 in /uai/ differs from that in /ai/ merely in the first 1/4 or 1/2 of the duration and the remaining portions of both syllables are nearly identical is a strong piece of evidence for the truncation process between the u->a and the /a/ steady state or the a->i transitions. The intersection between the two portions can be determined after a particular a->i transition shape (with or without an /a/ steady state) is interpolated. The truncation process takes place regardless of the length of the /a/ steady state or the overall a->i transition shape. A /uai/ F2 trajectory can be better approximated by adding a local smoothing process near the point where the two F2 trajectories truncate each other.

4.3. /au/ and /iau/

Many properties of F2 patterns in the au/iau pair are found to be similar or analogous to the ai/uai pair examined above. Therefore, I will only report the data of /au/ and /iau/ briefly.

The LPC plots in Figure 4.5 give us a general idea of the F2 patterns in /au/ and /iau/. The examples are the tokens read by our six speakers at a moderate speech tempo. The F2 trajectory in /au/ is traced onto the formant structure in /iau/ adjusted in phase so that the two F2 trajectories in /au/ and /iau/ match maximally. The empty arrow indicates the earliest temporal point where the F2 of /au/ and /iau/ coincide maximally so that in the remainder of the syllable the two F2 trajectories almost overlap. The means of F2 values and temporal position are shown in Figure 4.6.

The following summarizes the properties of F2 patterns in the au/iau pair. The corresponding properties of the ai/uai pair were found important for the prediction of the F2 of the ai/uai pair in the preceding section.

a) THREE TYPES OF F2 CONTOUR

Very much as in /ai/, the F2 in /au/ may exhibit a steady state of /a/ (Figure 4.5 B6, B4), a quasi-steady state (B2, B5) or a straight forward falling transition (B1, B3).

b) THREE PORTIONS IN F2 TRAJECTORY

Similar to A_0A_1, A_1I_0 and I_0I_1 in /ai/, three portions can be found in /au/, namely, A_0A_1, A_1U_0, U_0U_1. U_0 is no doubt the target position of /u/. Using the same line of argument as for A_0 in /ai/, we may assume A_0 to be the target position for /a/ in /au/. Also, the underlying /u/ target is at the end of the /a/ steady state and the realized /u/ target is at U_0. The straight falling transition from /a/ at syllable initiation to the realized /u/ target at U_0 results from a straightening process.

c) (QUASI-)STEADY STATE DURATION CORRELATES WITH SYLLABLE DURATION

Again, as in /ai/, the steady or quasi-steady state duration correlated with the syllable duration (r=0.6583). The mean duration of the steady state (66.67 msec, SD=62.78) is about 22% of the syllable duration (308.33 msec, SD=103.19). Thus, we may specify a probable duration of steady state /a/ in the a->u transition interpolation.

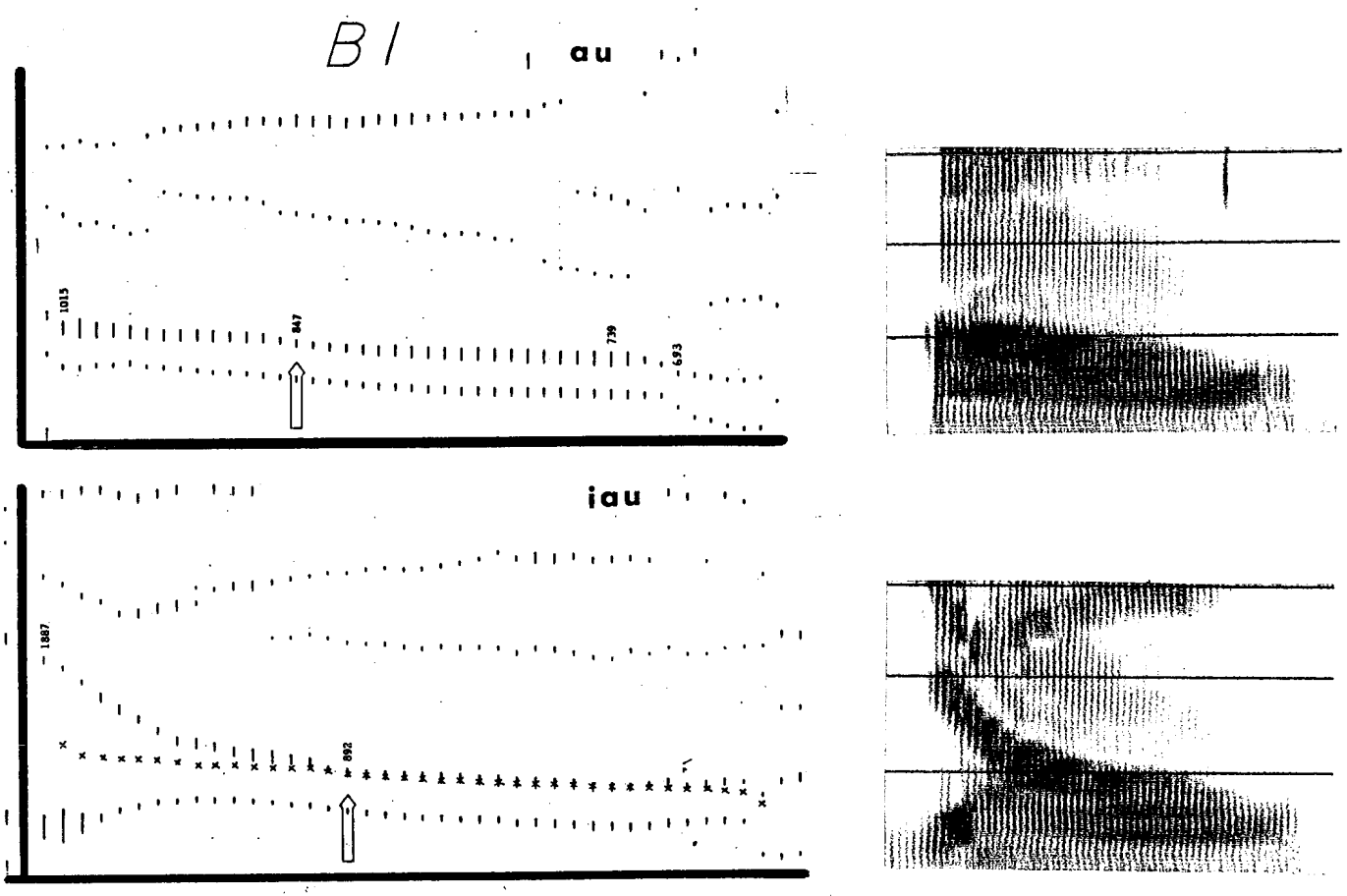
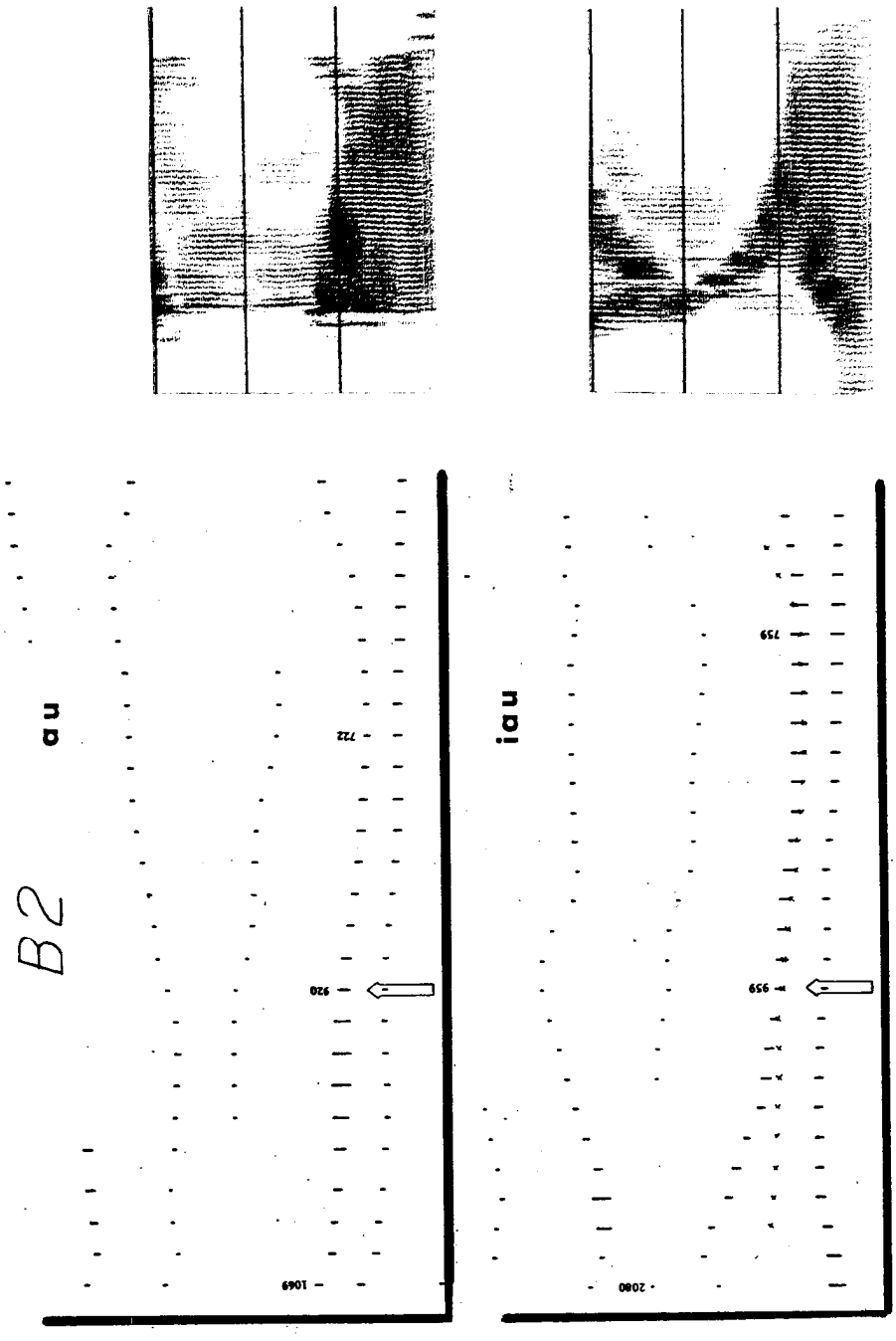
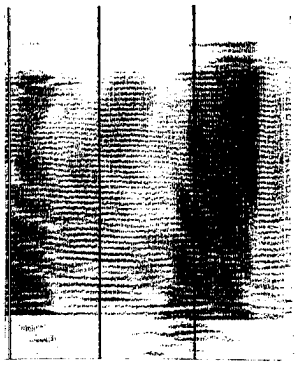
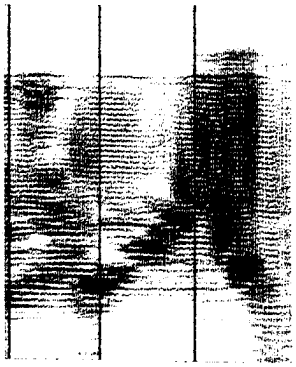
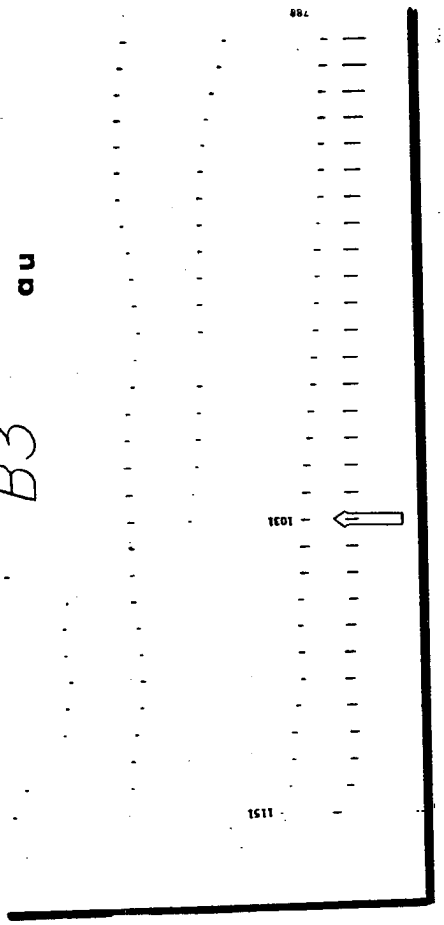


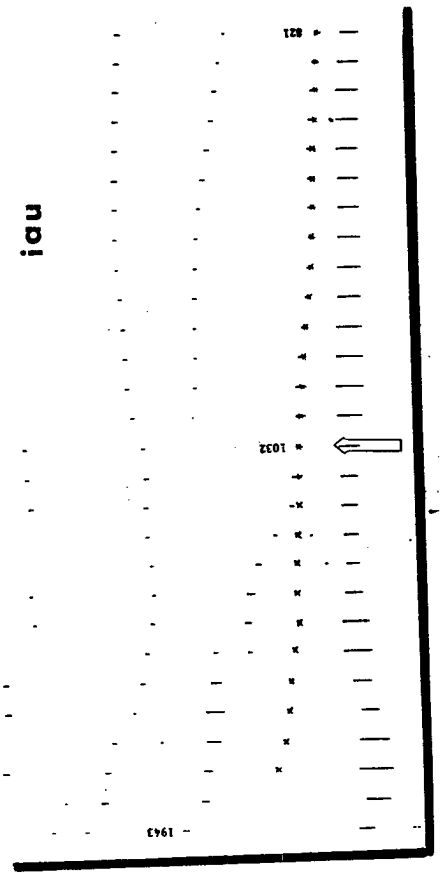
Figure 4.5. LPC formant tracks and spectrograms of /au/ and /iau/ read by six speakers at a moderate speech tempo. In the LPC graphs, some F2 values are provided for the discussion. The F2 trajectories of the /au/ are also traced (with x) onto the formant pattern of the /iau/ read by the same speaker. The empty arrow indicates the earliest point where the two F2 trajectories coincide maximally so that in the remainder of the syllable the two F2 trajectories can be considered to overlap.

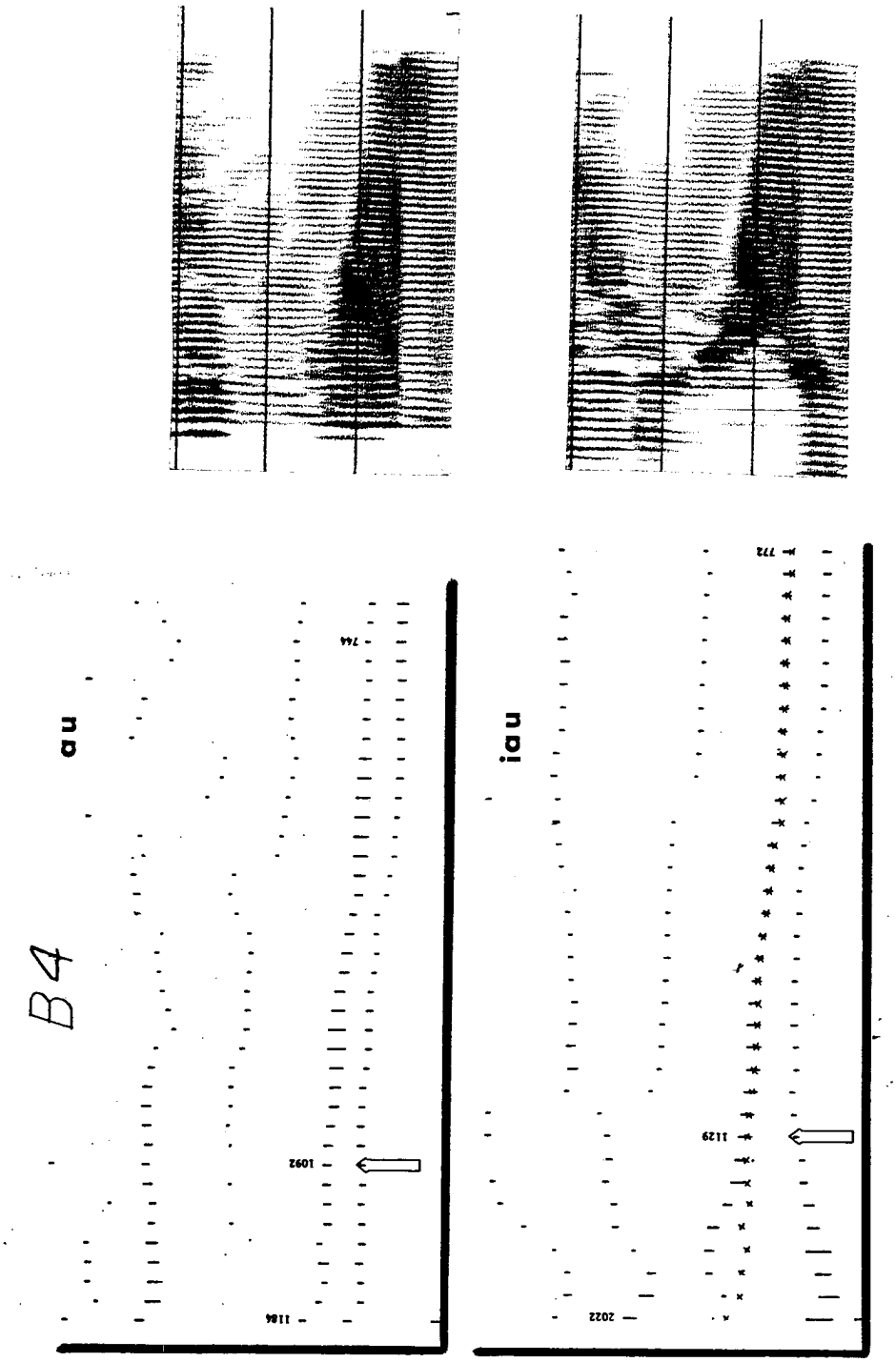

B2
au
722
920
1069
iau
759
959
2080



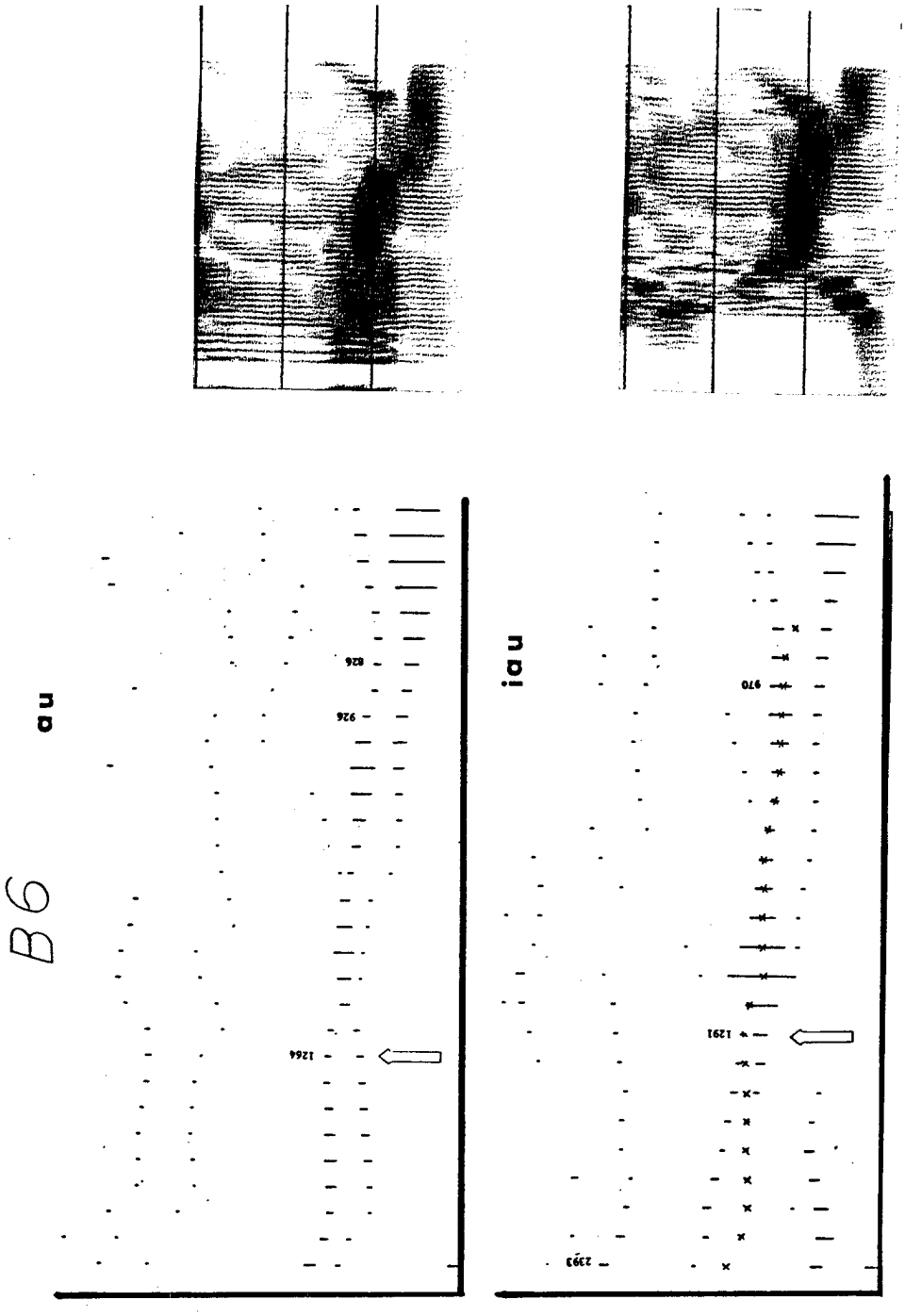



B3
au
788
1031
1151


iau
821
1032
1943


B4
au
744
1092
1186
iau
772
1129
2022

B5
au
712
932
1284
iau
720
1008
2136


B6
au
826
926
1266
iau
970
1291
2393

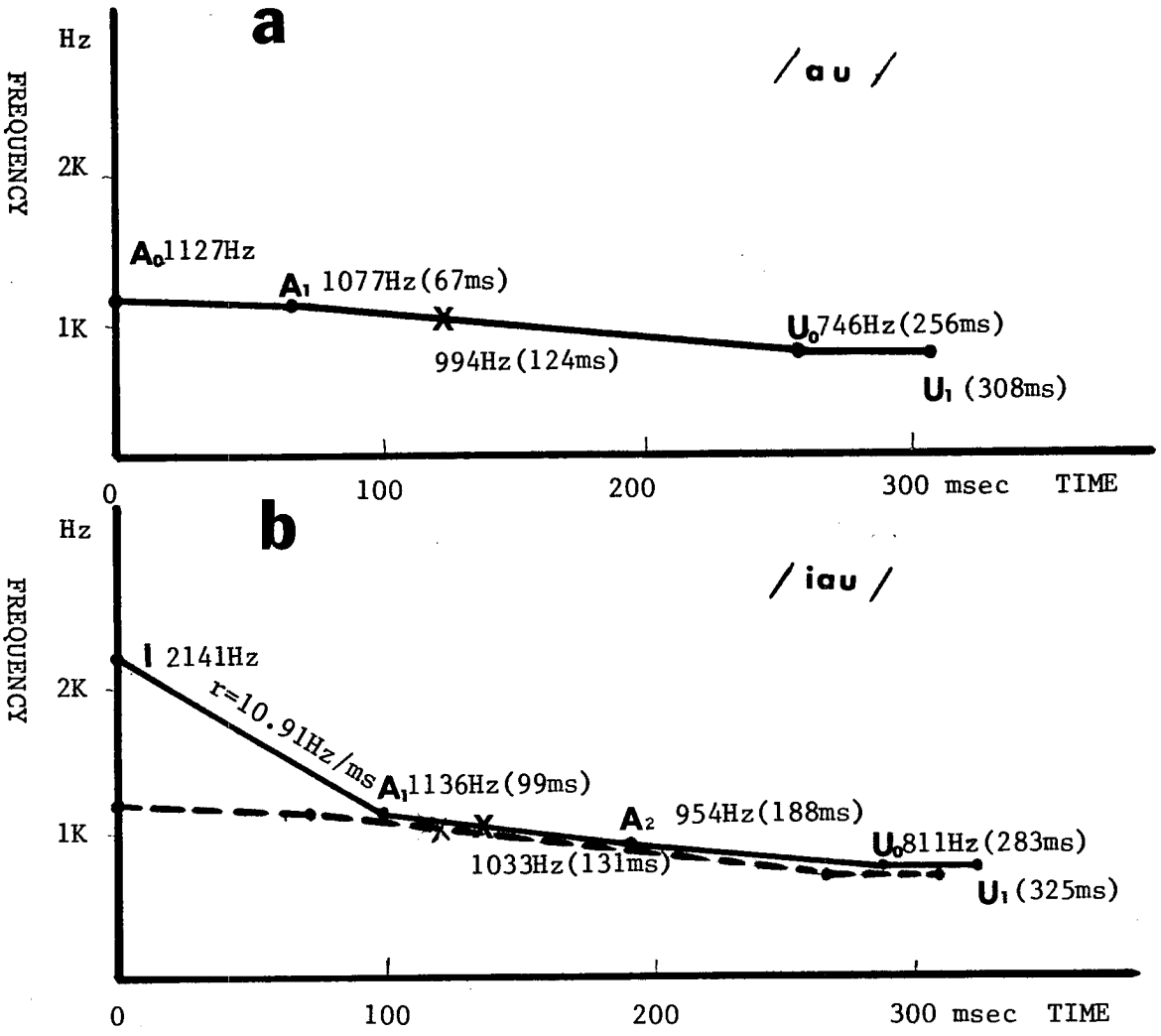


Figure 4.6. Mean measured values in /au/ (a) and /iau/ (b), pooled across 6 speakers x 3 speech tempos. The letter X denotes the temporal point where the F2 trajectories in the two syllables are considered to begin to overlap (when shifted in phase to maximize overlap). The dotted line on (b) represents the F2 trajectory of /au/ unadjusted for phase.

d) /u/ TEMPORAL POSITION CORRELATES WITH SYLLABLE DURATION

The temporal position of /u/ target, namely, the duration of A_0U_0 correlates with the syllable duration (r=0.9079). The mean temporal value for U_0 (256.67 msec, SD=101.52) is about 85% of the syllable duration. This enables us to derive the rate of the straight a->u transition which goes from the syllable initiation to the realized /u/ position.

e) TRUNCATION

Figure 4.5 shows that the F2 in /iau/ and the F2 tracing of /au/ differ merely in the first 1/4 or 1/2 of syllable duration, the remaining portions of the syllable being almost identical. This is to say that the /iau/ results from /au/ truncated by an i->a transition, regardless of the shape of the a->u transition.

f) LOCAL SMOOTHING NEAR THE INTERSECTION

Figure 4.5 also shows that a local smoothing takes place at the intersection of the two phonologically adjacent transitions, that is, the temporal point where the two transitions truncate each other. B1 is the clearest example of this sort of smoothing.

g) SLIGHT TEMPORAL EXPANSION OF SYLLABLE PATTERN

In both Figure 4.1 and 4.5, the traced /ai/ and /au/ F2 trajectories are 10 or 20 msec delayed against the background of /uai/ or /iau/, taking syllable initiation or the /a/ target position as a zero reference point for time. In Figure 4.2, the second half of the F2 trajectory in /uai/ lies slightly behind in phase position from that of the F2 trajectories (dashed tracings) in /ai/. The same situation can be found in Figure 4.6 for the au/iau pair. These facts show that the triphthong patterns are slightly expanded in time. Notice that the remainder of the syllable appears the same after the truncation process. We may think that this slight temporal expansion occurs at the early transitions, namely, the u->a in /uai/ and i->a in /iau/. Since the extent of the expansion due to the more complex transitions involved in vocalic components is only 10-30 msec, we may ignore it for the moment, until we need a very accurate prediction.

Using all the above properties of /au/ and /iau/ syllables, we can derive some possible F2 patterns (see Figure 4.7) using the same approach as for /ai/ and /uai/ in the preceding section. There are some possible types of variation in F2 patterns of /au/ with a probable steady state or quasi-steady state. One is to change the slopes of the quasi-steady state and the following relatively fast falling portion (Figure 4.7a). Another type is to change the duration of these two portions (Figure 4.7b).

In short, the comparison between the F2 patterns in /au/ and /iau/ shows clearly a truncation process in the production of a triphthong. As the data indicated, the au/iau pair confirms again that the interpolation between two targets in a diphthong needs to be specified before the truncation process.

4.4. /ou/ and /iou/

This pair is different from the ai/uai and au/iau pairs in that it has no clear steady state or quasi-steady state that can be measured from the F2 trajectory. Figure 4.8 shows the F2 trajctories in the tokens of /ou/ and /iou/, read by our six speakers at a moderate tempo. Let us first examine the general F2 shape in

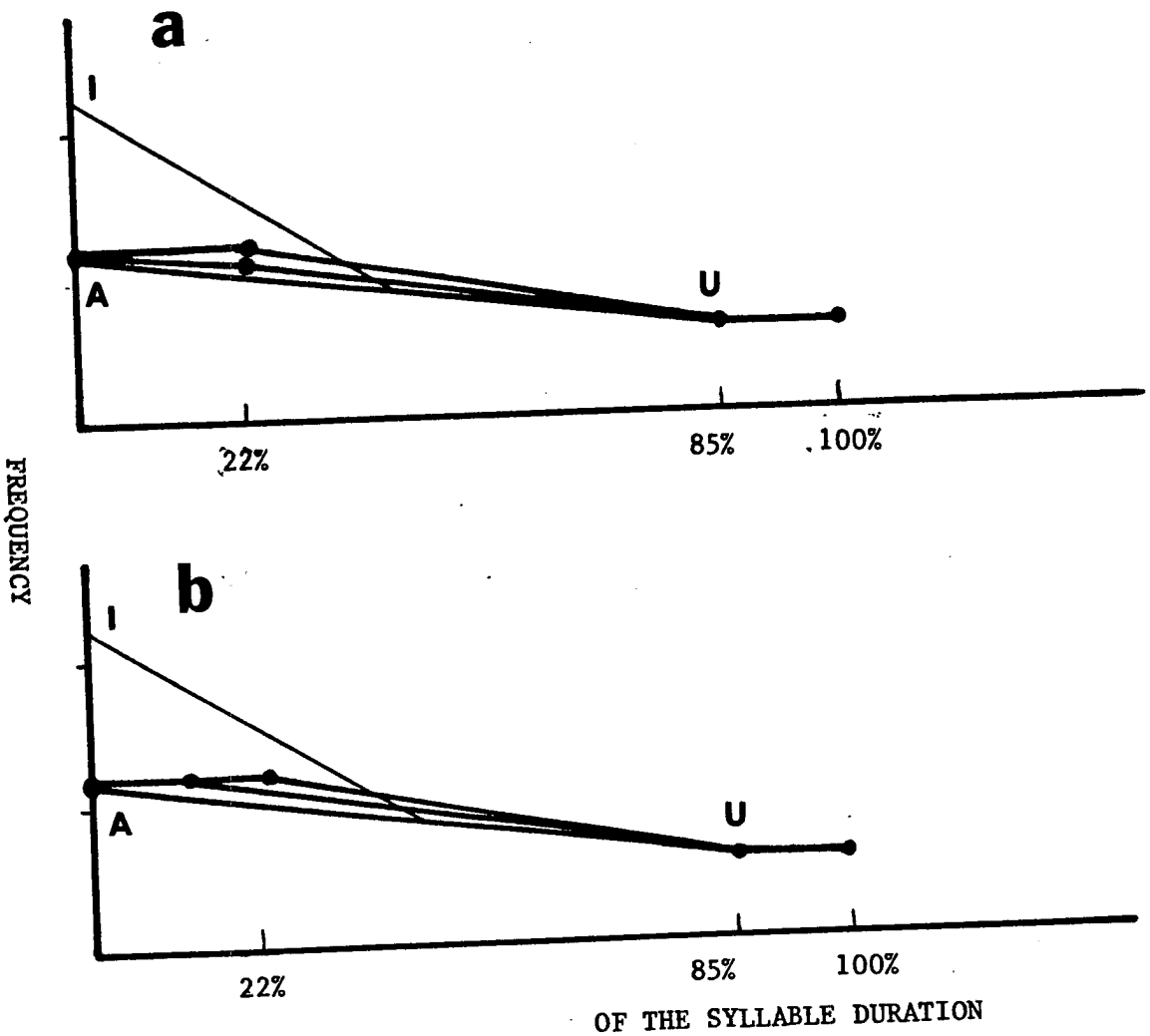


Figure 4.7. Scheme of two possible realization patterns of F2 in the syllable /au/.

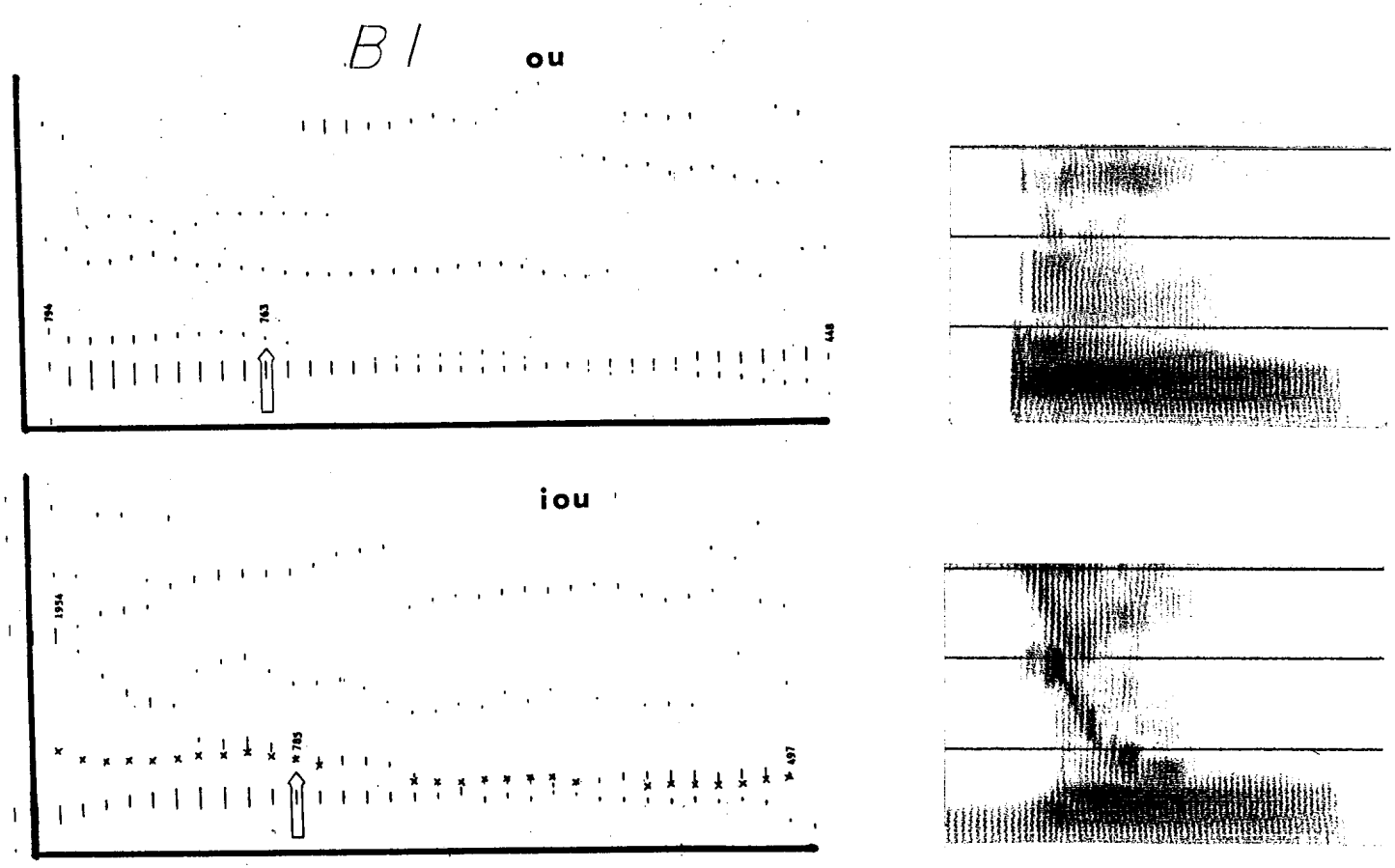
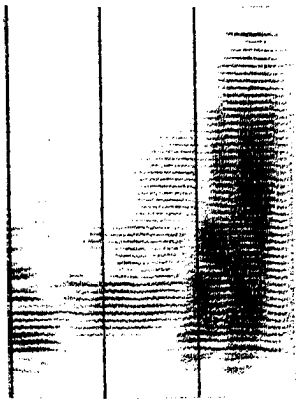
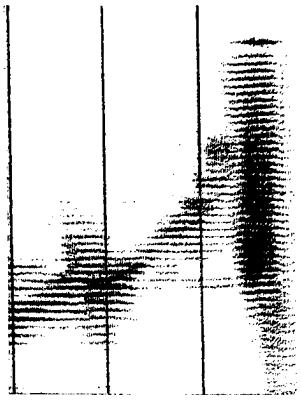
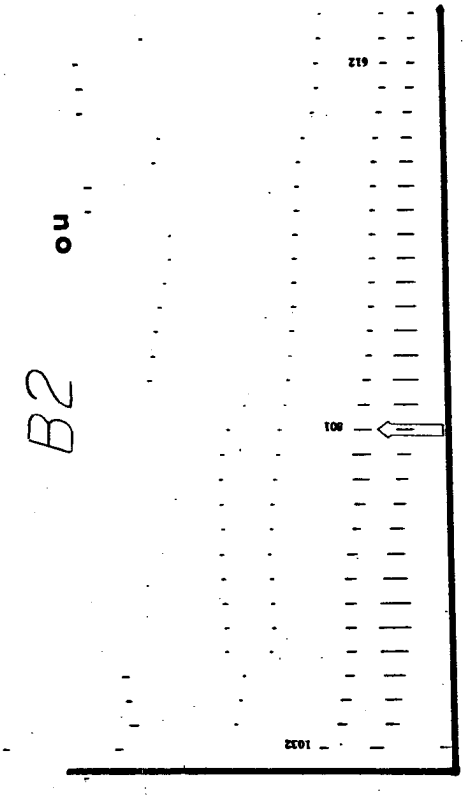
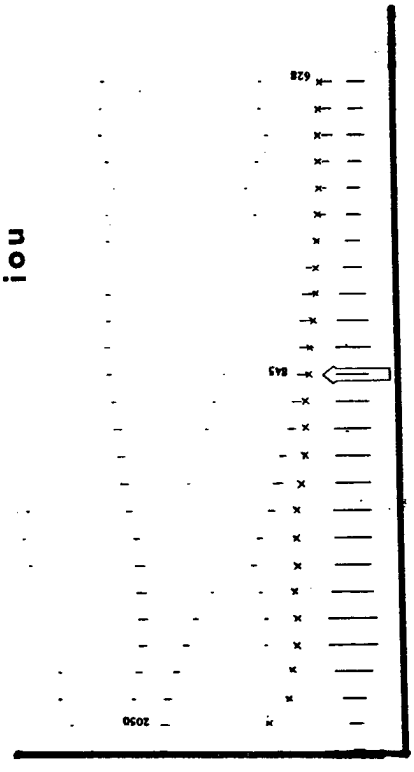


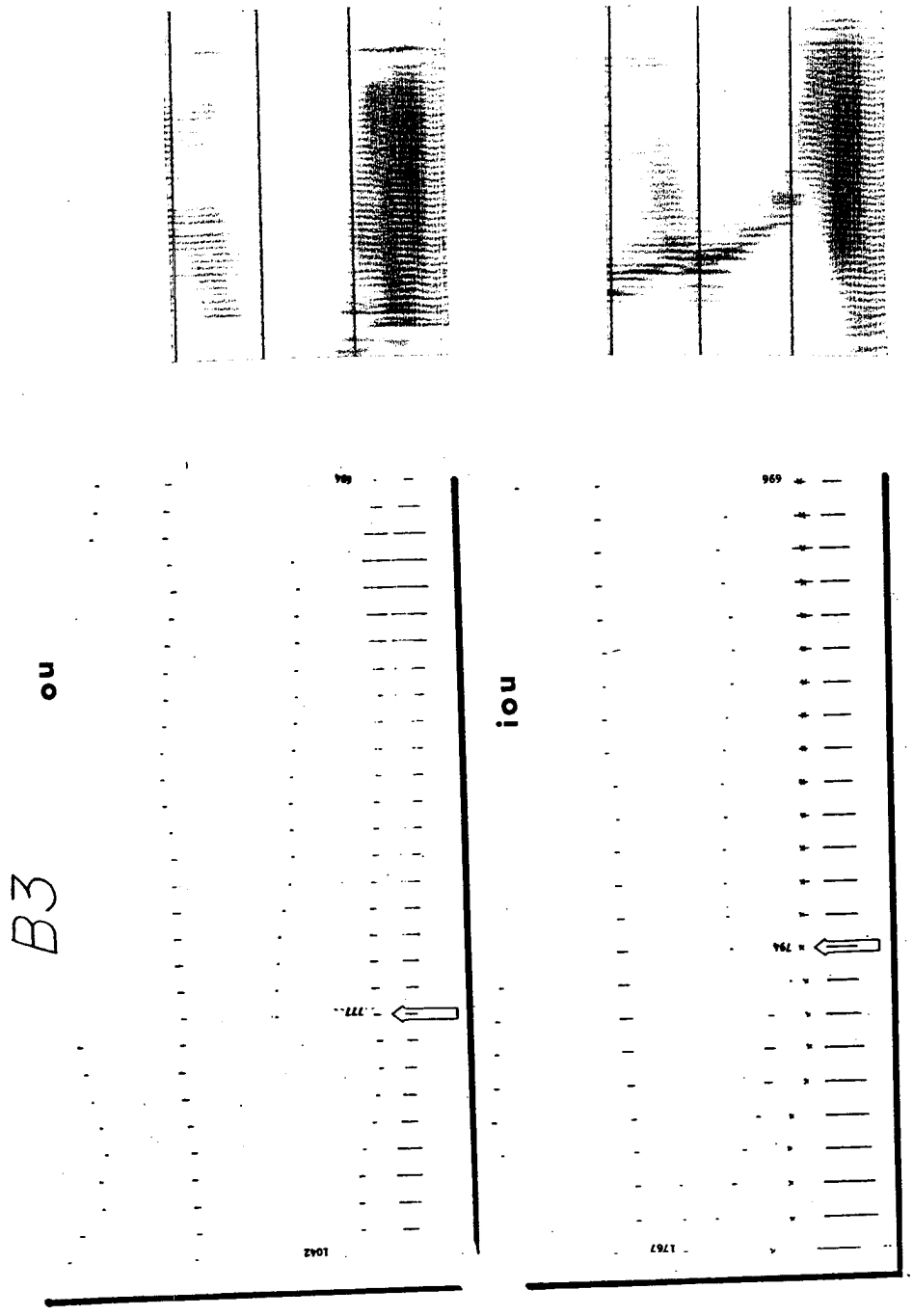
Figure 4.8. LPC formant tracks and spectrograms of /ou/ and /iou/ read by six speakers at a moderate speech tempo. In the LPC graphs, some F2 values are provided for the discussion. The F2 trajectories of the /ou/ are also traced (with x) onto the formant pattern of the /iou/ read by the same speaker. The empty arrow indicates the earliest point where the two F2 trajectories coincide maximally so that in the remainder of the syllable the two F2 trajectories can be considered to overlap.





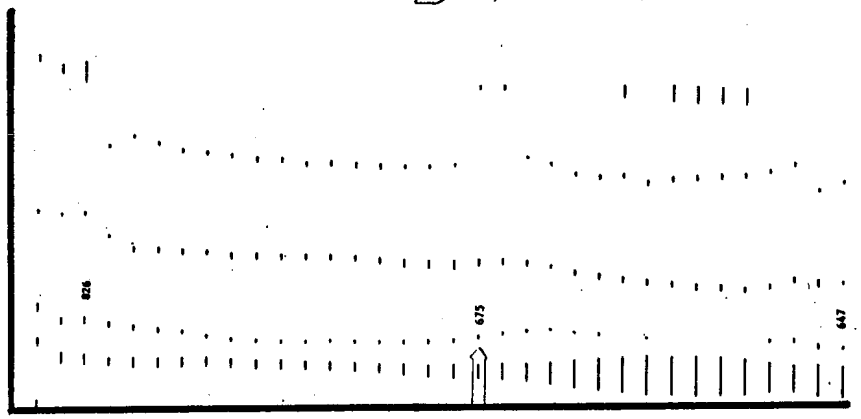

B2
ou
612
801
1032

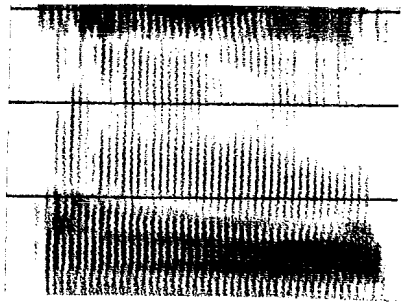

iou
628
843
2050


B3
ou
iou

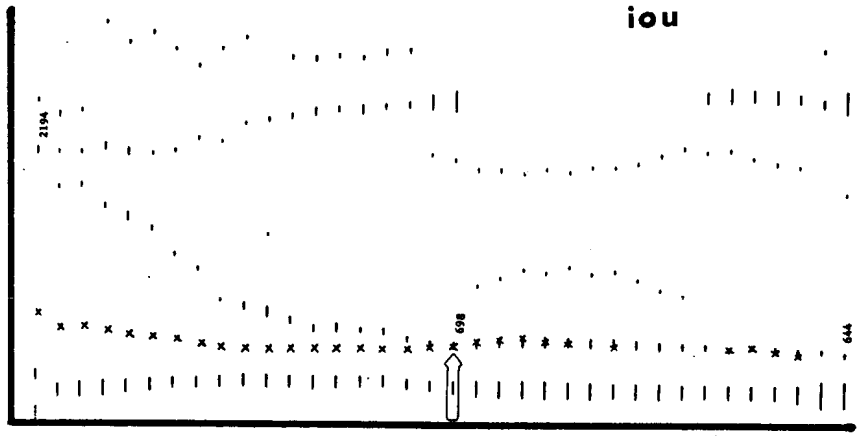
B4

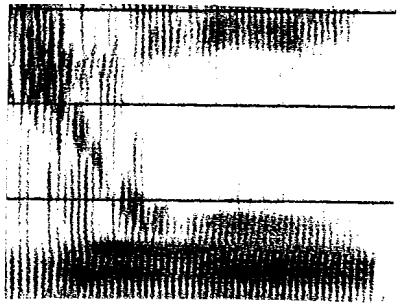
ou

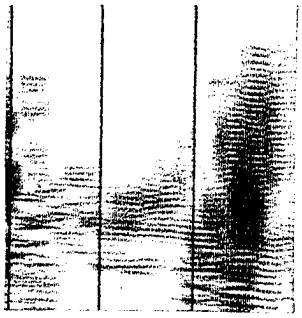


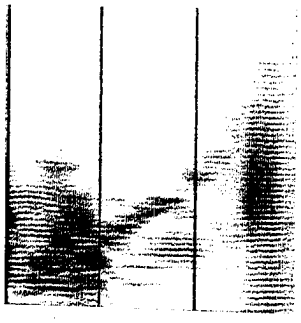


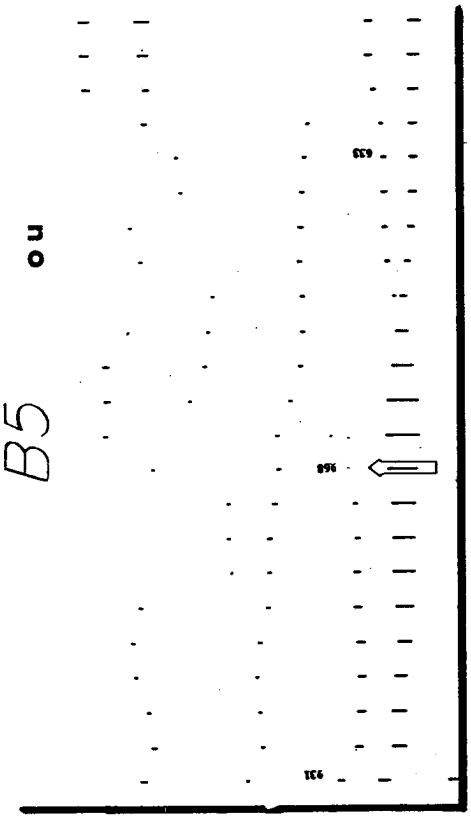
iou

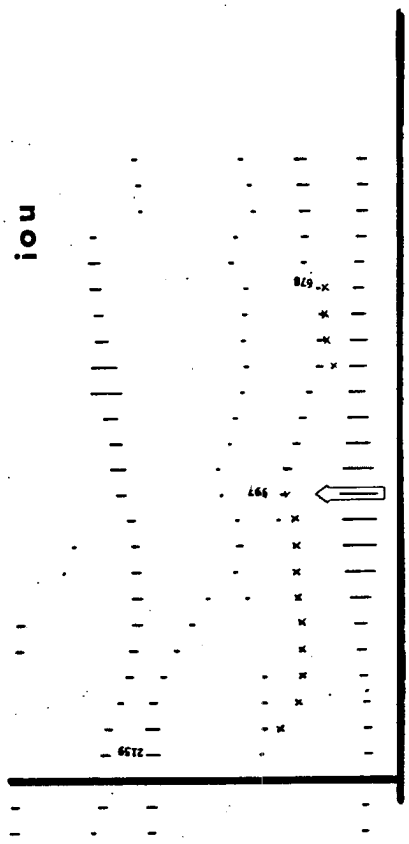


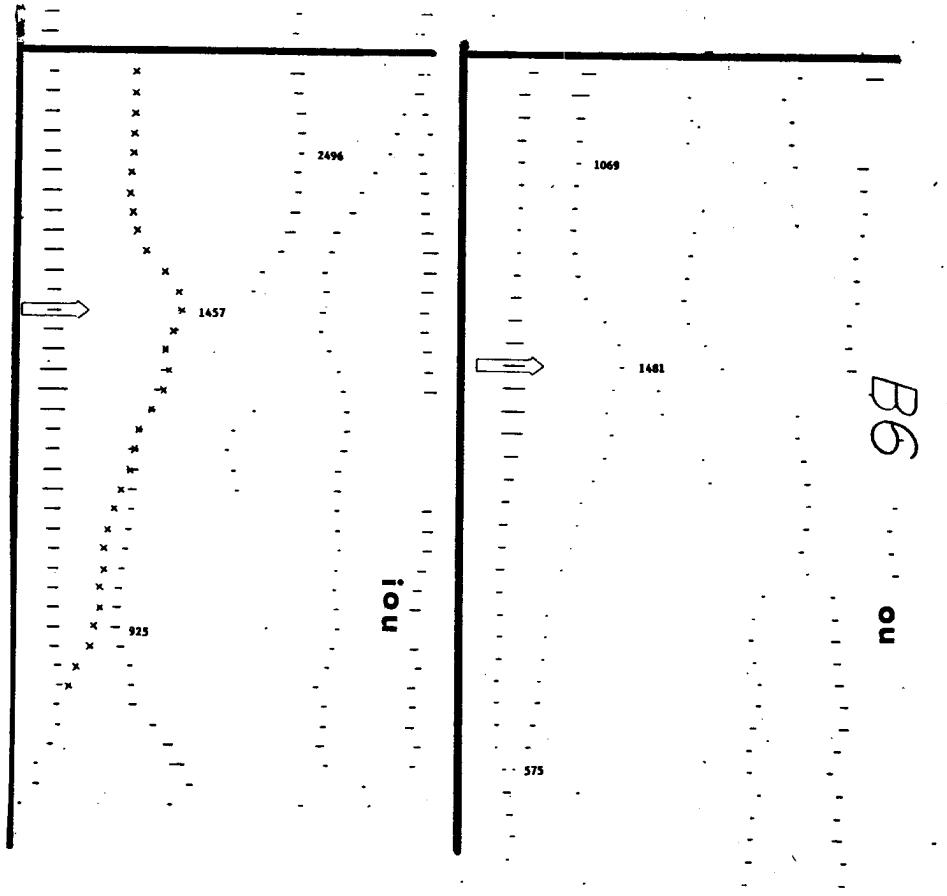


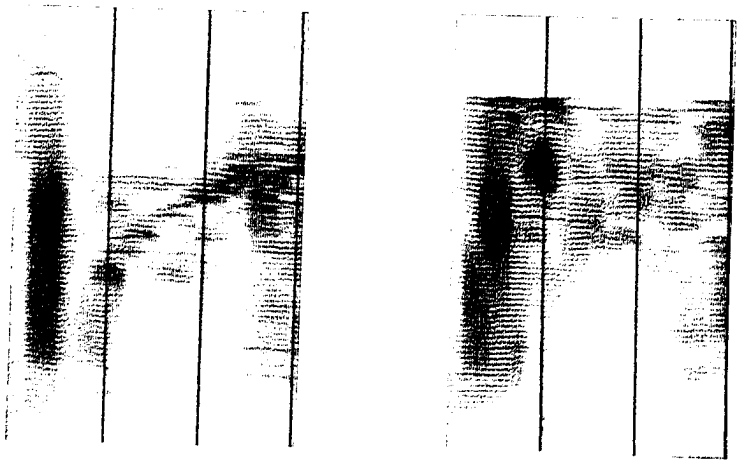





B5
ou
633
968
931


iou
678
997
2139


2496
1457
925
no!
1069
1481
575
no
B6



the diphthong /ou/. It appears that there is no steady state at either the initial part or final part of this diphthong. In fact, the LPC hasn't worked well in some tokens (e.g. in B6 in Figure 4.8; the spectrogram shows the formant trajectories more clearly), and since the total frequency variation range in F2 trajectory is small (approximately 400 Hz) in /ou/, it is technically difficult to distinguish a quasi-steady state from a 'transition' in the conventional sense. The highest F2 point in /ou/ can be assumed to be at (or very close to) the beginning of the syllable, while the lowest F2 point is likely to be located at the end of the syllables though some rising anticipatory transition to the following cross-syllable boundary context can be observed. It is both reasonable and convenient to treat the realized o->u transition in /ou/ as originating at the syllable onset and ending at the syllable offset. In this approach, the o->u transition rate can be roughly derived by dividing the F2 range by the syllable duration.

The truncation process in the production of this pair can be detected by comparing the F2 trajectories in /ou/ and /iou/ of corresponding tokens (those read by the same speaker at the same tempo). It can be clearly observed that the two F2 trajectories almost overlap in the second half of the syllables. The only major difference is the i->o transition occurring at the first 1/3 or 1/2 of the syllable, sometimes with a local smoothing near the point where the i->o transition and the o->u transition intersect (see B4 in Figure 4.8 for the most clear case).

The mean F2 values and mean temporal positions of the targets or turning points, as well as the mean i->o transition rate, are provided in Figure 4.9. With these data as the standard values of targets and transition rates, the F2 patterns in /ou/ and /iou/ for syllables of different lengths can be approximately predicted by the truncation model.

4.5. Summary

We have examined four diphthong/triphthong pairs in Chinese, ei/uei, ai/uai, au/iau and ou/iou. A truncation hypothesis has been proposed to account for the observed F2 patterns in these syllables in terms of target (timeless, with specific frequency value) and transition interpolation (connection of the two phonologically adjacent targets, with specifified rate). This hypothesis can be intuitively understood as follows. If a language allows, in a syllable without final consonant clusters, a sequence of segment-like units such as ABCD (in which D may be a post-nucleus element), it must also allow a sequence BCD and a sequence CD, but does not necessarily allow the sequence AB, ABC or BC. The present study presents data showing that the CD sequence does not compress itself in order to make room for an additional initial component B to form a BCD sequence within a relatively constant syllable duration domain. The BCD sequence in turn does not compress itself in order to make room for an additional initial component A to form an ABCD sequence. Rather, to form the sequence BCD, CD will sustain its full realized form as a two-unit syllable and only the first part (of the syllable) will be truncated by a B->C dynamic transition. The resulting BCD sequence keeps its form as a three-unit syllable but its first part can be truncated by an A->B transition to form an ABCD syllable. In short, adding a component to the initial part of a syllable is not realized by compressing the original components in the syllable, but by truncating the initial part of the syllable to allow a dynamic connection between the new component and the first component of the original syllable while the remainder of the syllable remains

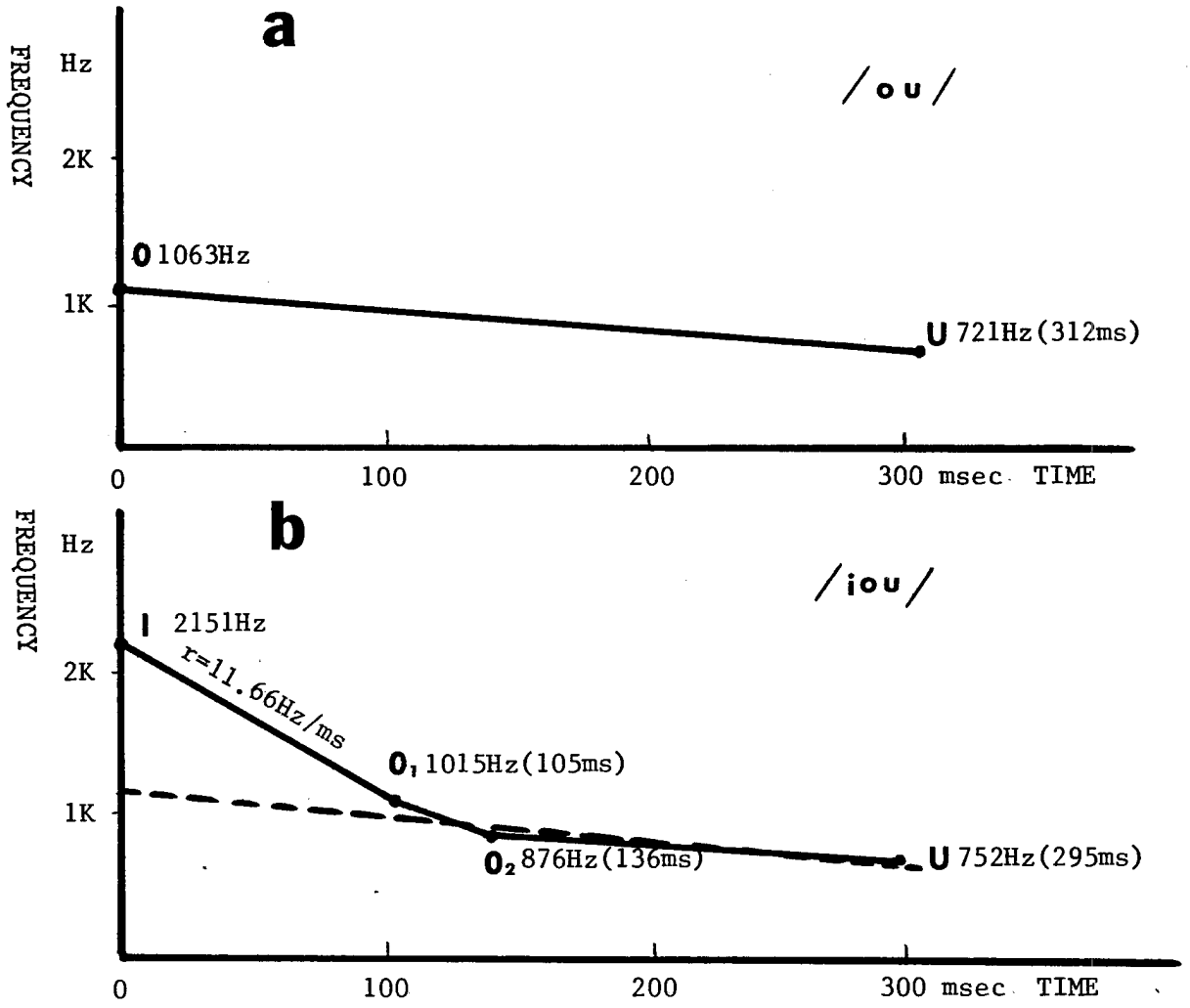


Figure 4.9. Mean measured values in /ou/ (a) and /iou/ (b), pooled across 6 speakers x 3 speech tempos. The letter X denotes the temporal point where the F2 trajectories in the two syllables are considered to begin to overlap (when shifted in phase to maximize overlap). The dotted line on (b) represents the F2 trajectory of /ou/ unadjusted for phase.

intact. With this pattern of complex syllable formation in mind we would expect that some non-marginal components in a syllable be shifted from their standard identities because of the truncation process. The phonetic values of these non-marginal components of a complex syllable can only be predicted by referring to their fully realized forms in relevant simpler syllables.

CHAPTER 5. F2 TRANSITION RATE (1)

5.1 Introduction

The truncation model suggested in the previous chapters delineates a general principle that governs the acoustic realization of syllables with complex vocalic components. However, the real F2 trajectory of a given syllable depends on many linguistic, paralinguistic and non-linguistic factors in a particular speech event. The basic linguistic factors for determining the observable F2 pattern of a syllable with complex vocalic components may include the underlying target values of the components and the rate for connecting phonologically adjacent targets. The F2 target values can be considered as being correlated with the vowel feature of backness in the classic acoustic-articulatory vowel space, and thus can be approximately derived from the known linguistic specifications of the sounds. However, we do not understand yet whether the F2 transition rate between two targets can be specified according to a general principle based on the linguistic features of the sound. This chapter attempts to find out the possible factors underlying the F2 rate specifications. On the one hand I will look for the cause of diversity in F2 transition rate among different diphthongs (or different pairs of phonologically adjacent targets); on the other hand, I will also test factors such as speech tempo, which yield possible variations of F2 transition rate for a given diphthong or a given pair of targets. More specifically, I will review in section 5.2 the proposal of Kent and Moll (1972) that the further the distance of articulatory movement, the faster the rate is. In the correponding acoustic domain, the correlation between the difference between two target values and the transition rate will be examined. In section 5.3, the effect of speech tempo on F2 transition rate will be investigated.

The analysis procedure is the same as that in the previous chapters. In addition to the four diphthong-triphthong pairs ei/uei, ai/uai, au/iau and ou/iou, the data in this chapter also include the following five diphthongs:

ia 'duck'
ua 'frog'
uo 'nest'
ie 'Je- in the word Jesus'
ye 'make appointment'

5.2. The Correlation between the F2 Range and the Rate of Transitions

One possible factor determining the different F2 transition rates among diphthongs or triphthongs is the articulatory or acoustic distance between the two targets that are joined by the transitions. It has been suggested that the rate of the tongue body movement is determined in a large part by the magnitude of the tongue movement, based on cinefluorographic films of tongue-point displacements during selected vowel gestures (Kent and Moll, 1972). More specifically, the further the tongue body has to move in executing a vowel gesture, the greater is the articulatory velocity. As a general notion, velocity is directly proportional to the distance traveled and is inversely proportional to the time taken. Kent and Moll's hypothesis implies that, while the distance which the tongue needs to travel varies, the time of the movement does not change proportionally. This is compatible with the fact that, though the distance of articulatory movement varies with different sounds in speech, the time of the

movement is greatly constrained by the inter- and intra-syllabic timing associated with the higher level temporal organization of a particular speech event.

As far as vowels are concerned, equal articulatory movements produce roughly equal acoustic changes. Accordingly, the acoustic consequence of the tongue movements would roughly follow the same law as Kent and Moll's principle that the more divergent the two target values are, the faster the transition rate is. If this hypothesis can be substantiated, the transition rate which is mainly associated with the tongue movement could be derived from the difference between the two F2 targets being connected, which are in turn derivable if the vowel features are specified for them.

The acoustic consequences of Kent and Moll's "the further the faster" hypothesis was tested against the Chinese diphthong data. Here, we assume that the acoustic consequences of other articulators such as lip gesture have only minor effects on the acoustic transition rate and would not obscure the general pattern of the dynamic tongue articulations. The correlation between F2 range (Δ F2) and the transition rate was calculated using a total of 162 tokens of the nine Chinese diphthongs. The rates of $A_0 \rightarrow I_0$ and $A_0 \rightarrow U_0$ are counted as the transition rates in the diphthongs /ai/ and /au/ respectively. The correlation coefficient is 0.6627 (correlation squared is 0.4392, df=160. two-tailed T-Test for linear relation has T=11.194, $p<0.001$; Equation is $y=0.005798 x - 0.017413$). The F2 transition rates are plotted against the Δ F2 in Figure 5.1. It is clear that the F2 range and the transition rate are only moderately correlated.

The moderate correlation between the F2 range and the transition rate indicates that the "the further the faster" hypothesis is tenable only as a loose principle for F2 transitions for all the Chinese diphthongs. It may be affected or constrained by some other factors. By examining the F2 transition rate, the F2 range and the ratio between the two measures for each diphthong (in 18 tokens) as listed in Table 5.1, we can find a factor -- the phonological order of the diphthong components -- which greatly lowers the correlation. The transition rates and the rate/ Δ F2 ratio are considerably different when the two targets are in opposite phonological order, though Δ F2 remains quite similar for both orders.

Taking the case of the /ai/ and /ia/ as an example, we can see from Table 5.1. that the two diphthongs share a similar F2 range (875 and 870), but the transition rate is much faster in /ia/ (7.07) than in /ai/ (3.93). The ratio between the Δ F2 and the rate is much higher for /ia/ (0.0081) than for /ai/ (0.0045). This is also the case for the pairs au/ua and ei/ie. However, the pair ou/uo does not show clear effects of component ordering. Since only the transition rates (but not the F2 ranges) are different due to the different phonological ordering, the correlation between the F2 range and the transition rate is expected to be low. That is probably a strong reason why these two measures are only moderately correlated in Chinese.

Assuming that the reason why the F2 transition rate is only moderately correlated with the F2 range is because of clear effects from the phonological order, another question will be addressed. The question is whether one of the two phonological orders abides by the principle of "the further the faster" while the other phonological order always differs from the first phonological order by a constant, or whether the second order can be derived from the first by some other

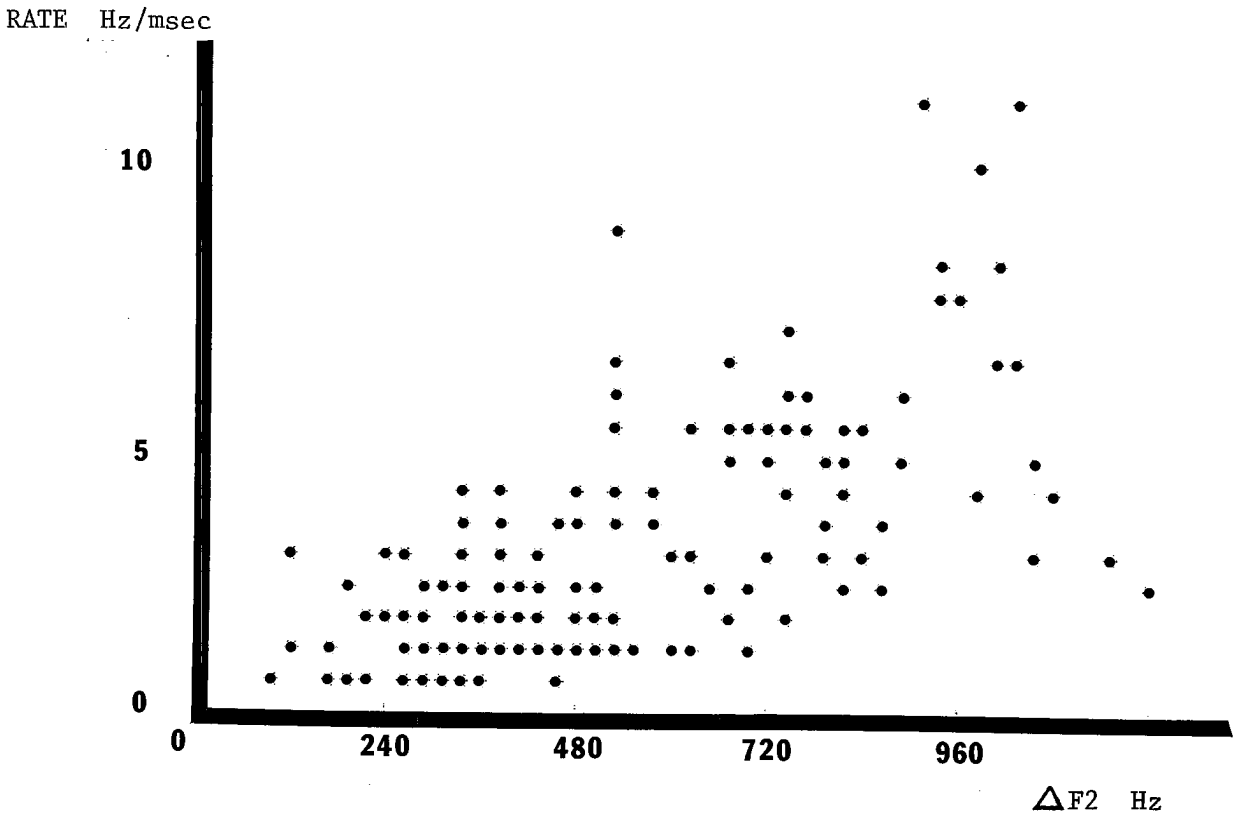


Figure 5.1. Correlation between Δ F2 and transition rate.

Table 5.1. Mean F2 range (Δ F2), transition duration, transition rate and ratio between the rate and the Δ F2 for each diphthong in Chinese (n=18). Standard Deviations are provided in parenthesis.

diphthong	Δ F2(Hz)	Trans.dur(ms.)	Trans. rate	rate/ Δ F2 ratio
ai	875.06(164.84)	219.44(76.86)	3.93(1.32)	0.0045
ia	869.50(144.96)	130.56(34.55)	7.07(2.19)	0.0081
au	395.89(95.19)	256.67(101.52)	1.62(0.47)	0.0040
ua	370.78(124.07)	109.98(22.38)	3.57(1.37)	0.0096
ei	448.50(188.68)	143.89(52.59)	3.39(1.86)	0.0076
ie	507.11(200.90)	213.33(116.47)	2.09(0.59)	0.0041
ou	341.60(134.60)	316.11(93.50)	1.18(0.58)	0.0035
uo	406.50(168.96)	332.22(109.41)	1.37(0.75)	0.0034
ye	469.44(178.18)	217.67(127.24)	2.23(1.66)	0.0048

rule. To answer this question, let us first compare the mean rate/ Δ F2 ratio for the nine diphthongs with those for individual diphthong patterns of different phonological orders. The mean ratio for the nine diphthongs in Chinese is 0.0058. The ai/ia pair of diphthongs have the ratios 0.0045 and 0.0081 respectively. Both are considerably different from the mean ratio for the nine diphthongs. Note that the transition from a higher vowel has greater rate/ Δ F2 ratio than that toward a high vowel. The deviations from the mean are even greater in the case of the au/ua pair (0.0040 and 0.0096). As in the ai/ia pair presented above, the ua/au case also shows that the transition from a high vowel yields a higher ratio than that toward a high vowel. Great deviations can also be found in the ei/ie pair (0.0076 and 0.0041). Here, the transition toward a high vowel yields a higher ratio than that from a high vowel. However, no difference can be found in rate/ Δ F2 ratio in the ou/uo pair. The deviations in the ei/ie pair are in the opposite direction in relation to the position of the high vowel of the diphthong, compared with the ai/ia and the au/ua pairs. A general pattern emerges that the pairs involving a low vowel /a/ are different from those without it. The presence or absence of a low vowel has a more extensive effect which I will discuss in the next section.

Since most of the diphthongs (ai, au, ie, ye, ou, uo, which we will call Subset 1) have a rate/ Δ F2 ratio of about 0.004, it is convenient to treate this ratio as the basic one. The supplementary set of diphthongs (Subset 2) includes ia, ua, ei, which have the phonological components in opposite order to some members of Subset 1 (ai, au, ie). The diphthongs in Subset 2 have the rate/ Δ F2 ratio of about 0.008, which is double that of Subset 1. We can thus roughly derive the F2 transition rate of each diphthong by multiplying the F2 range by this standard rate/Δ F2 ratio, and doubling it if the diphthong is in Subset 2. Since the total

number of Chinese diphthongs is small, a table look-up procedure will be much easier than a derivation procedure for use in speech engineering.

The fact that the three diphthongs /ia, ua, ei/ in Subset 2 have a rate/Δ F2 ratio double that of the other diphthongs raises interesting questions about the phonological status of these three diphthongs. The F2 transition rates are considerably faster in these three diphthongs than in their counterparts with reversed component order. It has long been recognized that a difference in formant transition rates from an initial element serves as a basic acoustic cue for distinguishing a glide from a vowel for that initial element. In a perceptual study by Liberman et al (1956), increases in the formant transition duration with the same onset and offset values (consequently, decreases of the formant transition rate) caused judgment changes from /we/ to /ue/ and from /je/ to /ie/. Lehiste and Peterson (1961) reported that, in Finnish, a sequence of vowel plus vowel may contrast with a sequence of semivowel plus vowel in such word pairs as /iäinen/ 'eternal' vs. /jäinen/ 'icy'. In the analysis of a number of contrasts of this type the formant frequencies associated with the targets remained constant, but the change in the rate of formant movement produced a meaningful difference. Another example given by Lehiste and Peterson is a comparison between similar sounds in two languages, such as the diphthongs [ui] and [iu] in Estonian compared with /wi/ and /ju/ in English. The slopes of the transitions and the durations of the target positions were the only differences, since the formant positions were identical. A conclusion was therefore made by the authors that since the target position remains constant, the slope of the transition is a criterion for defining the difference between diphthongs and glides.

Let us go back to the Chinese diphthong data. Note first the large difference in F2 transition rate between the two diphthongs /ia, ua/ (put aside /ei/ for a moment) and their counterparts with reversed component order /ai, au/. Following the above argument with regard to the distinction between glide and vowel, it is reasonable to think of the difference in F2 transition rate between /ia, ua/ and /ai, au/ as reflecting the difference between a glide-vowel combination and a vowel-vowel combination. Therefore, the diphthongs /ia, ua/ are phonetically [ja, wa]. We need a rule acounting for the phonological process:

$$(5.1) \qquad \underset{[+high]}{V} \quad \text{----> [-syllabic] / } \_\_\_\_\_ \quad \underset{[+low]}{V}$$

Besides /ua/([wa]) and /ia/([ja]), the diphthong /ei/ in Chinese also has a much faster F2 transition rate than its counterpart with reversed component order, /ie/. The acoustic pattern of /ei/ is rather strange considering its phonological status. It has been shown in Chapter 3 (as well as in He, 1985) that /ei/ has an e->i transition followed by a (sometimes quite long) steady state of /i/. This pattern raises doubt concerning the conventional treatment of /ei/ as having a nucleus vowel /e/ and an offglide /i/. We can find a similar situation for /e^{I}/ in English as reported in Lehiste and Peterson (1961). There is no steady state for the first element of /e^{I}/, but a slow glide appears toward the target position. Often, the first part of /e^{I}/ has been called the "full vowel" and the second element the glide or semivowel. In the dialect of American English used in their study, it is actually the second element that has a steady state and the first element that is phonetically a glide.

The problem of /ei/ discussed above necessitates a reconsideration of the phonological status of its components. Since /ei/ is grouped together with /ia, ua/ regarding the rate/ Δ F2 ratio, we consider that /ei/ may behave accordingly.

That is, based on the acoustic data, the first element in /ei/ is also phonetically a glide-type sound, e-glide, and the second element is a full vowel /i/. This treatment is compatible with Lehiste and Peterson's description of English $/e^{I}_{I}/$.

Note that there is no considerable difference in F2 transition rate between /ou/ and /uo/. Therefore, two diphthongs with similar movements in vowel height, /ei/ and /ou/, do not behave in a similar way. A parallel difference in acoustic pattern between $/e^{I}_{I}/$ and $/o^{U}_{U}/$ in English was also found by Lehiste and Peterson. Unlike $/e^{I}_{I}/$ which has long glide as the first element, $/o^{U}_{U}/$ involves a short steady state as the first element and a long glide as the second element. Therefore, the initial glide occurs only in a diphthong of the /ei/ type but not in one of the /ou/ type. The rule we need to yield the phonetic representation of a glide should be restricted to /ei/:

$$(5.2) \qquad \begin{matrix} V \\ [-back] \end{matrix} \quad \text{------>} \quad [-\text{syllabic}] \quad / \quad \text{———} \quad \begin{matrix} V \\ [+high] \end{matrix}$$

Since /i/ in /ei/, /uei/, /tuei/ is the nucleus, all the elements in these syllables have their target at syllable initiation, according to the hypothesis revised in Chapter 4. regarding the temporal location of target corresponding to pre-nucleus and nucleus element. English /l/ in the words /el/, /wel/, /dwel/, as well as in many other words, also has a syllabic nature, because of its overall influence on the rhyme part of a syllable. Therefore, the target of /l/ is located at the syllable initiation.

A further question might be whether the effect of the ai/ja type of difference in diphthong components on acoustic transition pattern is due to a universal principle or is just a language-specific phenomenon. The F2 transition rates of the pairs of diphthongs in opposite phonological order were investigated in an acoustic study for Spanish (Manrique, 1979). It was reported in that study that the F2 transition rates in /ai/ and /ia/ are 8.5 and 7.0 respectively at a moderate speech tempo (the data show almost the same transition rates at fast speech tempo as those at the moderate speech tempo for all the diphthongs); the rates in /au/ and /ua/ are 5.85 and 6.47 respectively; the rates in /ie/ and /ei/ are identical (3.70); the rates in /uo/ and /ou/ are 3.05 and 2.30 respectively. It is obvious that the phonological order does not have any significant influence on transition rate in Spanish. This indicates that the stronger effect of phonological order in Chinese on F2 transitions is due to the presence of the consonant elements in the first component position of one type of diphthongs and thus is a language-specific phenomenon.

In addition to the general patterning of the order effect, the transition rate is quite different for the similar phonological elements in the two languages. Taking an extreme case, we find the rate in Spanish /au/ (6.47 for moderate speech and 6.75 for fast speech) is almost four times as much as that in Chinese /au/ (1.62 for $A_0 \rightarrow U_0$ transition and 1.99 for $A_1 \rightarrow U_0$ transition as the mean of three tempos). These facts related to F2 transition rate support the conclusion made by Lindau (1985) based on a cross-language comparison of the transition durations in the 'same' diphthongs. She claims that the timing of the diphthongal transition is not constant for the 'same' diphthong in different languages, and that the timing relationships can be predicted from a general principle of duration varying with the distance (in target values) for some diphthongs, but not all. As far as the transition rates in Chinese diphthongs are concerned, language-specific factors such as the phonological rules which yield [j, w, ɐ̯],

must be considered before we can determine the transition rate based on the F2 range in diphthongs.

Let us now briefly examine the relation between the transition rate and the transition duration (see Figure 5.2). The correlation between these two measures is -0.6192 (df=160, T=-9.974, $p<0.001$). Since in most of the diphthongs the transition duration is almost the same as the syllable duration, the correlation will be low if the syllable duration varies in accord with other factors such as speech tempo (which I will investigate in the next section) or higher level linguistic organization, such as word length, prosodic structure and so on, and, in addition, if Δ F2 does not vary in accord with these same factors to a similar extent because the Δ F2 is mainly related to the nature of minimal phonological units. In addition, the correlation has been found to be very low between the F2 range and the transition duration in the nine diphthongs in Chinese. The r is only -0.0659 (df=160, T=-0.835). This indicates that the timing in transition is independent from the difference in target values.

In short, the F2 transition rates in different diphthongs in Chinese are found to be only moderately correlated with the F2 range. Kent and Moll's hypothesis of 'the further the faster' is at work as a loose principle for all the Chinese diphthongs. Phonological 'order' (combined with other considerations such as the position of the high vowel component and the vowel feature [low]) has been found to be a strong factor affecting F2 transition rate. The diphthongs /ia, ua, ei/ have much faster F2 transition rates than the diphthongs with the 'same' components in the reversed order. Since considerable difference in formant transition rate is an acoustic feature distinguishing a glide from a vowel, they are phonetically [ja, wa, e̯i]. Leaving these glide-vowel combinations out, Kent and Moll's hypothesis works quite well for Chinese diphthongs. The difference between the above-mentioned two sets of diphthongs demonstrates a characteristic of diphthongs in Chinese, as opposed to Spanish where no significant difference in transition rate can be found for the same pair of phonological elements in the inverse order. Furthermore, the F2 transition rates for most diphthongs (/ai, au, ie, ye, ou, uo/) can be roughly predicted from the F2 range directly by a basic rate/ Δ F2 ratio (0.004). The remainder of the diphthong patterns ([ja, wa, e̯i]) need the doubled basic ratio (0.008).

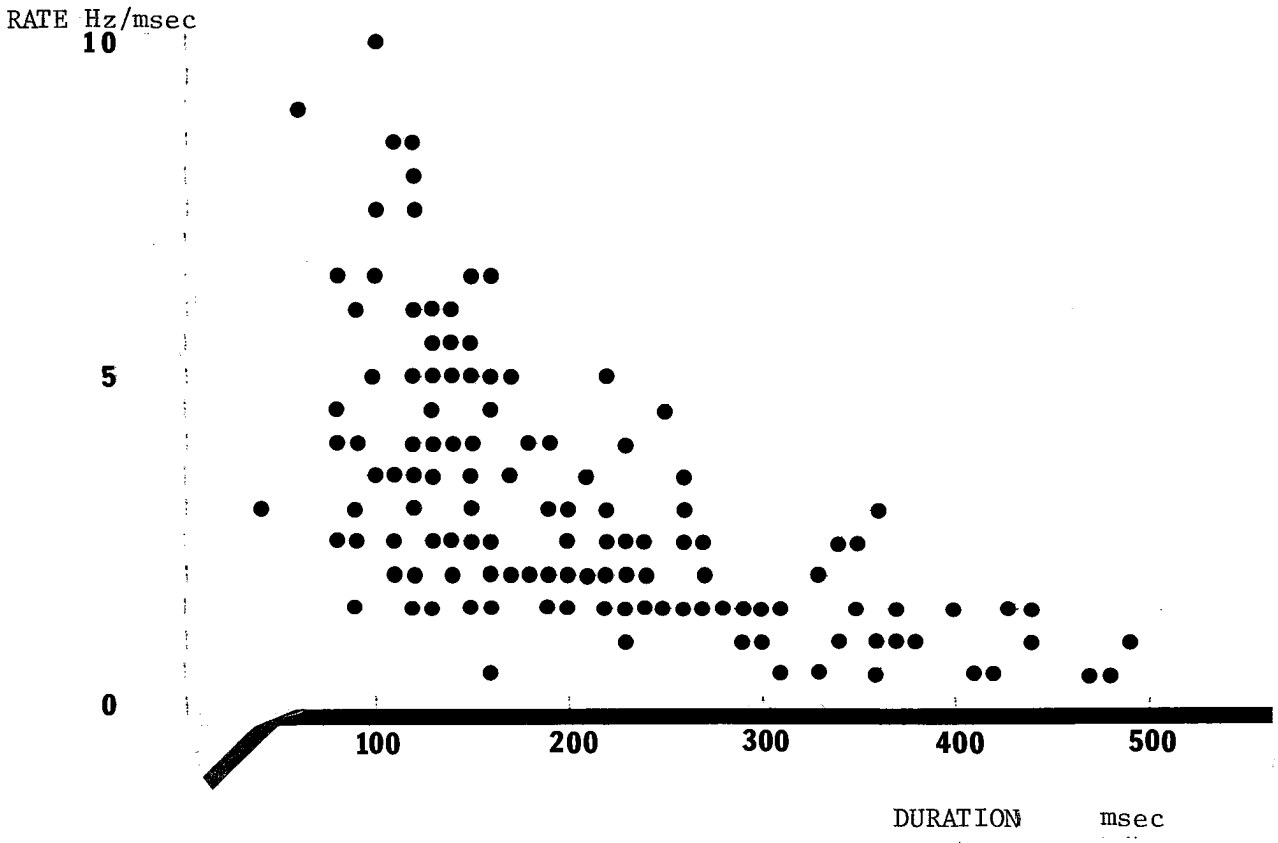


Figure 5.2. Correlation between F2 transition duration and rate.

5.3. Effect of speech tempo on F2 transition rate

We have just examined factors such as F2 range and phonological order in the specification of transition rate in different diphthongs. Now, we are concerned with factors affecting the rate variations of a particular diphthong. Several investigators have reported on the issue of the possible effects of speech tempo on F2 transition rate. It has been claimed that there are two features governing diphthong formant movement in American English: onset frequency position and the second formant rate of change; each English diphthong has a constant F2 transition rate across three different tempos (Gay, 1968; Kent and Moll, 1972). Similarly, it was found that the Spanish diphthongs show little variation in F2 transition rate when speech becomes fast. In contrast, speech tempo is reported to significantly affect the F2 transitions in English diphthongs (Dolan and Mimori, 1986). More specifically, the F2 transition rate decreases as speech tempo speeds up, as a result of changes in transition onset and offset frequencies. It is also reported by the same authors that Japanese showed less tempo-dependent variability in F2 transition than English, supporting the idea that the structure of temporal reorganization of speech for different tempos is language-specific.

The present section aims to find out whether Chinese diphthongs are tempo-independent like those in Spanish, Japanese (and English reported by Gay, Kent and Moll) or tempo-dependent like those in English reported by Dolan and Mimori. This distinction is of great importance for a predictive model for acoustic patterns involving complex transitions. If F2 transition is tempo-independent, then it can be derived from the linguistic factors alone and the F2 trajectory for speech at different tempos will vary only in steady state duration. If F2 transition rate is tempo-dependent, then the rate must be derived each time from syllable durations at particular tempos as a para-linguistic factor.

The effect of tempo on F2 transition rate will be tested by examining the correlation between the F2 rate and the syllable duration of each diphthong or triphthong pattern, which are the measures of speech tempo. The reason for using syllable duration to represent the different speech tempos is that the three speech tempos at which the speakers read the test words were chosen arbitrarily, and very often, the tokens were read at similar speed though they were asked to be read at two different categories of tempo, such as fast and moderate, or slow and moderate. In addition, great variety can be found among speakers for a particular category of tempo. Thus, the syllable duration may reflect the tempo better than the three category labels.

The correlation between F2 transition rates and the syllable durations in the 18 tokens for each transition pattern was calculated. The correlation coefficients for all the transitions involved in 13 Chinese diphthongs and triphthongs are listed in Table 5.2, according to the rank order of the correlations. The rate of the transitions from A_0 at the syllable initiation is used in calculating the correlation for the syllables /ai/ and /au/.

No general pattern emerges in this table, since for some of the diphthongs variation in transition rate correlates with variation in syllable duration and for others it does not. For example, the first five transitions show a moderate correlation with syllable duration (r is beyond 0.6) while the final three transition rates show very low correlations with syllable duration (r is below 0.2). The other nine transitions reveal only intermediate correlations (r is between 0.4-0.6 for most of them).

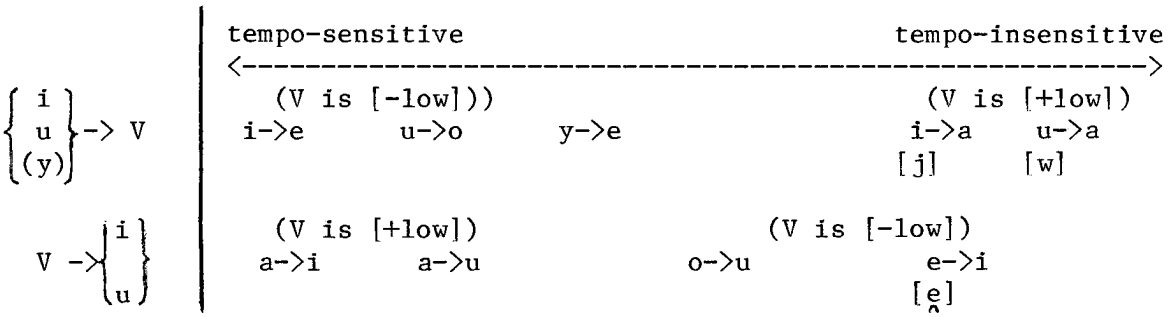
Table 5.2. Rank order of transition rates according to the correlation between the syllable duration and the F2 transition rate. The rate/ ΔF2 ratios for the diphthongs are given as well.

order	transition	syllable	syl. duration (SD)	rate (SD)	r	rate/ΔF2
1	i->e	ie	359.44 (106.63)	2.09 (0.59)	-0.7164	0.0041
2	a->i	ai	312.22 (115.58)	3.93 (1.32)	-0.7110	0.0045
3	a->u	iau	325.56 (103.94)	2.08 (0.87)	-0.6775	
4	u->o	uo	332.22 (109.41)	1.37 (0.75)	-0.6444	0.0034
5	a->u	au	308.33 (103.19)	1.62 (0.47)	-0.6439	0.0040
6	u->e	uei	319.44 (118.25)	13.13 (4.49)	-0.5697	
7	o->u	ou	316.11 (93.50)	1.18 (0.58)	-0.5373	0.0035
8	i->a	iau	325.56 (103.94)	10.91 (3.45)	-0.5216	
9	a->i	uai	310.56 (113.84)	5.11 (1.73)	-0.4802	
10	y->e	ye	355.00 (93.26)	2.23 (1.66)	-0.4510	0.0048
11	u->a	uai	310.56 (113.84)	7.62 (3.10)	-0.4073	
12	e->i	ei	340.00 (117.92)	3.39 (1.86)	-0.4025	0.0076
13	o->u	iou	295.56 (85.08)	1.94 (0.98)	-0.3045	
14	i->o	iou	295.56 (85.08)	11.66 (3.89)	-0.2284	
15	i->a	ia	290.56 (100.26)	7.07 (2.19)	-0.1948	0.0081
16	e->i	uei	319.44 (118.25)	3.80 (1.83)	0.1501	
17	u->a	ua	275.56 (87.46)	3.57 (1.37)	0.0022	0.0096

Examining the elements in the diphthongs and their relation to variation in syllable duration, we may find some patterns conditioning such a rank order of correlations. The a->i and a->u transitions in diphthongs are among the transitions with the highest correlations, while the i->a and u->a transitions are among the transitions with the lowest correlations. It appears from this fact that the 'phonological ordering' of the two targets affects the sensitivity of rate to tempo. This is also true in the case of the i->e and e->i transitions but to a lesser extent for the case of the u->o and o->u transitions. It is

interesting that, again, /ia, ua, ei/ stand out as different from other diphthongs. This supports the analysis of these three diphthongs as really being [ja, wa, ei] phonetically. When we consider vowel features such as [high] for the onset and offset targets of transitions, we find two different tendencies of tempo-sensitivity. For the transitions involving a low vowel /a/, the transitions from the high element /i/ or /u/ tend to be tempo-insensitive while the transitions toward a high element tend to be tempo-sensitive. In contrast, for the transitions where there is no low vowel involved, the transitions from the high vowel tend to be tempo-sensitive while the transitions toward the high vowel tend to be tempo-insensitive. The pairs of diphthongs with a low vowel element do not behave in a parallel way to the pairs without a low vowel element. Note the resemblance between the rank order of the tempo-sensitivity and the rank order of the rate/ ^ F2 ratio. It is likely that the faster rate in diphthongs with a low vowel as the second component correlates with a high rate/ ^ F2 ratio and tempo-insensitivity, while the slower rate in diphthongs with a low vowel as the first component correlates with a low rate/ ^ F2 ratio and tempo-sensitivity.

The general pattern of the conditions for tempo-sensitivity in diphthongs can be illustrated by the following diagram



For the transitions in triphthongs, the correlations show a less regular pattern. The a->u transition in /iau/ is much more tempo-sensitive than the u->a transition in /uai/, supporting the above analysis for the transitions in diphthongs. However, the a->i transition in /uai/ and the i->a transition in /iau/ are similar with regard to the correlation with syllable duration. Considering the fact that the transitions in triphthongs are affected by their mutual interactions, which result in more variations or cause uncertainty in the measurement of the rate, and considering also the absence of strong counter-evidence against the above analysis for diphthong data, we may assume that the transitions in triphthongs behave in a similar way as in the corresponding diphthongs, noting, however, that there is still some unexplained variation.

The results of this study of Chinese diphthongs have indicated that speech tempo does affect the F2 transition rate. Some transitions are more tempo-sensitive than others. A preliminary analysis of conditions of different tempo-sensitivity suggests that both the 'phonological ordering' (associated with the presence or absence of [j, w, e̞]) and the vowel feature [low] are relevant. Unlike the data reported for Spanish and Japanese and those for English reported by Gay (1968) and Kent and Moll (1972), the Chinese data demonstrate diphthong-specific characteristics of tempo effect on transition rate. Our findings thus reflect a language-specific way of timing reorganization in speech production for different tempos.

5.4. F2 transition rate in a predictive model

We have examined the following factors affecting the F2 transition rate: F2 range and speech tempo, phonological order, the presence or absence of the vowel feature [low]. Besides these factors, the complexity of a syllable may also slow down the transition rate. For example, the transition rates involved in the English word /dwel/ and the u->e transition in the Chinese word /tuei/ are slower than the corresponding transition rates in diphthongs as shown in Chapter 3. In addition, the phonological positions of the two targets vis-a-vis syllable boundary also affect the F2 transition rates. For example, the a->i and a->u transitions in Chinese have faster rates across a syllable boundary than in syllable internal positions (see next chapter). All these facts tell us that the F2 transition rates in diphthongs are highly variable and may not lend themselves to precise predictions in terms of simple cause-effect relations.

Though the observable F2 transition rate in diphthongal syllables results from complex interactions among the above discussed factors, an approximate prediction from the phonological elements that are present is still possible. A theoretical derivation may entail the following steps. First, the F2 target values can be determined from the phonological elements in a diphthong. Secondly, a basic rate/Δ F2 ratio (0.004) can be used to derive the F2 transition rate for most of the pairs of targets. A subset of the pairs of targets ([ja, wa, e͡i]) needs a doubled basic ratio (0.008). Such a derivation is meaningful in practice only if the F2 transition of a given pair of targets is a tempo-insensitive transition. In fact, a table-look up procedure will be much simpler than such a derivation method.

The next steps will be important for both theoretical and practical purposes. The syllable duration must be specified based on the intersyllabic timing, which in turn is determined by higher level linguistic structures as well as paralinguistic aspects such as speech tempo. With the syllable duration set up, the most tempo-sensitive (or actually syllable-duration sensitive) transition rate can be calculated by the formula (5.1):

$$(5.1)\quad r = \Delta F2 \;/\; D_s$$

where D_s is syllable duration. The most tempo-insensitive transition rates can be considered as standardized or fixed, irrespective of the speech tempos. For practical purposes, the rates for the phonetic j->a, w->a, j->o, w->e͡, and e͡->i transitions in Chinese diphthongs and triphthongs can be standardized and the rest of the diphthongs can be predicted by formula (5.1).

With all the targets at the syllable initiation and with the rates of all the transitions specified in the way presented above, one can simply apply a truncation principle to yield an approximate F2 trajectory.

For a more accurate prediction, we must determine the probable transition rates for the transitions which are moderately tempo-sensitive or moderately tempo-insensitive. Even for the tempo-sensitive transitions such as a->i and a->u, the rates are only moderately correlated with the syllable durations (r=-0.7110 and r=0.6775 respectively). Furthermore, the transition durations are not equal to the syllable durations in these two diphthongs. One possible way to derive the transition rates is to derive the transition durations from the syllable durations by a constant proportion, as presented in section 4.2 and 4.3 in the previous chapter. For a better prediction, we also must consider syllable

complexity and other factors affecting transition rate. Before the truncation process takes place, the particular transition shape (presence or absence of a steady state at the initial part of a transition) must be chosen. A smoothing process may occur at the intersection between two phonologically adjacent transitions. The variations in target values at the syllable marginal position have to be considered as well, though the target undershooting and overshooting at the syllable internal position can be largely predicted by the truncation model.

Taking into account that some of these factors are optional and the acoustic pattern results from different combinations of the weights of many factors, we could not expect that a given phonological transcription at a given tempo yields exclusively one single acoustic pattern by a set of strict one-to-one cause-effect relations. A set of slightly different patterns will be accepted as the possible acoustic realizations of a given diphthong at a given tempo, using the approach suggested in this study. However, the predictable patterns will be constrained and the truncation model will certainly rule out some impossible patterns. For example, the F2 trajectory in /tuei/ in Chinese would never be the consecutive concatenation of the full range t->u, u->e and e->i transitions even though the syllable duration is long enough to incorporate the three transition durations.

In summary, the F2 transition rates have been found to be moderately correlated with the F2 range of the transitions. This indicates that the acoustic consequence of Kent and Moll's hypothesis of 'the further the faster' works only as a loose principle for all the Chinese diphthongs. One factor that might decrease the correlation between the transition rate and the F2 range is the phonological ordering of the two components of the diphthongs. A further examination of individual diphthong patterns reveals that the rate/ Δ F2 ratios for most of the diphthongs with one component order (except ou and ou) are about 0.004 while the ratios for the rest of the diphthongs with opposite component order (consisting of [j], [w] or [e̯]) are doubled (0.008) When the vowel feature [low] is involved in the diphthong, the transition from a high vowel has doubled rate/ Δ F2 ratio of that for the transition toward a high vowel. This is not the case for the diphthongs where there is no low vowel.

This study has also investigated the possible variations in F2 transition rate for a given diphthong. It has shown that the speech tempo affects the F2 transition rate in diphthongal syllables. There is in fact a range of tempo-sensitivity for F2 transition rate from the most tempo-sensitive F2 transitions to the relatively tempo-insensitive ones. It appears that, in some diphthong pairs with reverse order of the components, the F2 transition in one member is tempo-sensitive while that in the other is relatively tempo-insensitive. The conditioning of the tempo-sensitivity is related to both the phonological order and some vowel features. More specifically, when a low vowel is involved in the diphthong components, the transition toward a high vowel tends to be tempo-sensitive while the transition from a high vowel (a glide) tempo-insensitive; when there is no low vowel component in the diphthong, the transition from a high vowel tends to be tempo-sensitive, while the transition toward a high vowel tempo-insensitive. Both the test with rate/ΔF2 ratio and with the correlation between rate and syllable duration show a division between two groupes, /ia, ua, ei/ ([ja, wa, e̯i]) and /ai, au, ou, uo, ie, ye/.

For the first approximation of a prediction based on the truncation model suggested in Chapter 3, the F2 transition rates of some tempo-sensitive

transition can be derived by r= Δ F2 / D_s (D_s is syllable duration), while the rates for the most tempo-insensitive transitions can be derived from the F2 range and the rate/ Δ F2 ratio. Then the realized F2 trajectory in the syllable can be approximately predicted by the truncation process.

CHAPTER 6. F2 TRANSITION RATE (2)

6.1. Introduction

As we have discussed in Chapters 3-5, the F2 transition rate plays a central role in the acoustic realization of different diphthongs or triphthongs at different speech tempos. However, the specification of F2 transition rates depends on (1) how the onset and offset points (on frequency and time dimensions) are defined and (2) how the transition course mathematically is represented.

I have only presented one simple type of measurement of F2 transition rate in order to introduce the truncation model for the acoustic realization of diphthongal syllables. As regards the determination of the onset and offset of an F2 transition, the following approaches are used in this study:

(a) Syllable onset and offset
In some diphthongs, there is no clear steady state in either the initial or final part of the syllable. The F2 point at the syllable initiation and at the syllable ending are treated as the onset and offset respectively of an F2 transition. The F2 transitions in /ou/ and /uo/, for example, were treated in this way in the previous chapters.

(b) Turning point
The point where the F2 trajectory changes direction in triphthongs has been regarded as one end of a transition. Very often, this point is a peak or valley of an F2 contour. This point was treated as the offset point of a preceding transition and coincident with the onset of a following transition. It should be emphasized that the F2 transition discussed here is the acoustically realized transition rather than the underlying F2 transition whose onset is always at the syllable initiation. The English words /wel/ and /dwel/, and the Chinese words /uei/ and /tuei/ are examples of this case.

(c) Intersection with a final steady state
The point where a transition intersects a final steady state such as that in the case of /ei/, /ai/ and /au/ was treated as the ending point of the transition.

(d) An initial steady state as a part of the transition
Since there are free variations between a straightforward a->i or a->u transition and a transition consisting of an /a/ steady state and a faster transition toward the /i/ target, the F2 transition rates in /ai/ tokens were measured as originating from the syllable initiation (i.e. from the onset of the /a/ steady state) to the /i/ target. The /au/ diphthong was similarly treated.

With regard to the mathematical representation of F2 transitions, a straight line approximation is usually used in the literature of diphthongs (Gay, 1968, Manrique, 1979, Dolan and Mimori, forthcoming). Representations using curves can also be found in the literature. An exponential function has been reported to fit F2 transitions well (Rabiner, 1968; Fujisaki, 1979). Other representations include polynomial representation (Yang and Cao, 1982) and cubic spline representations (Lindau, 1985).

In this chapter, I will measure F2 transition rates in the four Chinese diphthongs /ia, ua, ai au/ in a different way from that employed in Chapters 3-5. A constant difference in F2 between two adjacent points in LPC F2 tracks will be

used to distinguish an F2 transition from an F2 steady state. An exponential function with a time constant T, used by Rabiner and Fujisaki, will be used to represent the F2 transitions determined by this new criterion. The aim of this test is to check whether the basic properties of the F2 transition found earlier remain when the onset and offset of these F2 transitions are determined in this alternative way. I will also look in detail at the shape of an F2 transition and test the suitability of the exponential curve representation.

6.2. Procedures

The data used in this chapter include the diphthongs /ia, ua, ai, au/ (as four monosyllabic words 'duck, frog, sad, boil' respectively) read at three different tempos, as well as these same vowel sequences occuring across an intervening syllable boundary, i.e., /i#a, u#a, a#i, a#u/ (as in four disyllabic words /ni#a/'you guy', /k^{h}u#a/'rather bitter', /a#i/'aunty' and /ta#u/'big house', where only one of the two syllables in the words bears a high level tone). The same speakers were used. The starting point and the ending point of a transition in LPC F2 trackings were determined by establishing a criterion to distinguish the transition from a steady state. Subject to some exceptions described below, the first LPC point which differs from the following point by more than 10 Hz in the case of /ia, ai/ and by more than 5 Hz in the case of /ua, au/ is regarded as the starting point of the transition; the last point which differs from the preceding point by more than 10 Hz in the case of /ia, ai/ and by more than 5 Hz in the case of /ua, au/ is considered the ending point of the transition. The average range of overall F2 frequency changes (Δ F2) is 875 Hz for /ai/ and 869 Hz for /ia/. The F2 ranges are much smaller in /au/ (395 Hz) and /ua/ (370 Hz). Therefore, only the changes in F2 value between data points which are greater than approximately 1.5% of total F2 frequency change in these diphthongs are regarded as participating in the transition.

Exceptions were made when adjoining points showed that the difference was a local aberration. In some cases, although the difference between two adjacent points is greater than 10 Hz as in the case of /ia, ai/, we still did not consider the change to be part of the transition. For example, in the slow utterance of /ia/ of speaker B2, there is a sequence of points (in Hz): A=2149, B=2127, C=2062 ... L=1223, M=1207, N=1219, O=1220.... Although the difference between L and M is 16 Hz (1223 - 1207), we note that N and O are similar to L. It is likely, therefore, that L rather than M should be regarded as the ending point of the transition, because LMNO seems to be a portion of a steady state, with one deviant value at point M.

In some cases, the difference between two adjacent points is less than 10 Hz, but the two points were still considered part of the transition rather than part of the steady state; e.g., the slow speech /ai/ of speaker B2 shows a sequence: ...A=1328, B=1341...O=1873, P=1913, Q=1968, R=2017, S=2041, T=2053, U=2062, V=2084, W=2109, X=2132.... In this case, although STU shows slower frequency changes, we still consider it a part of the transition, because it is contained within a longer stretch which is changing at a quicker rate.

6.3. Exponential Function and Time Constant T

Figure 6.1 presents examples of the F2 transitions of one token each of /ia/, /ua/, /ai/ and /au/, read by speaker B1 at a moderate speech tempo. For each of

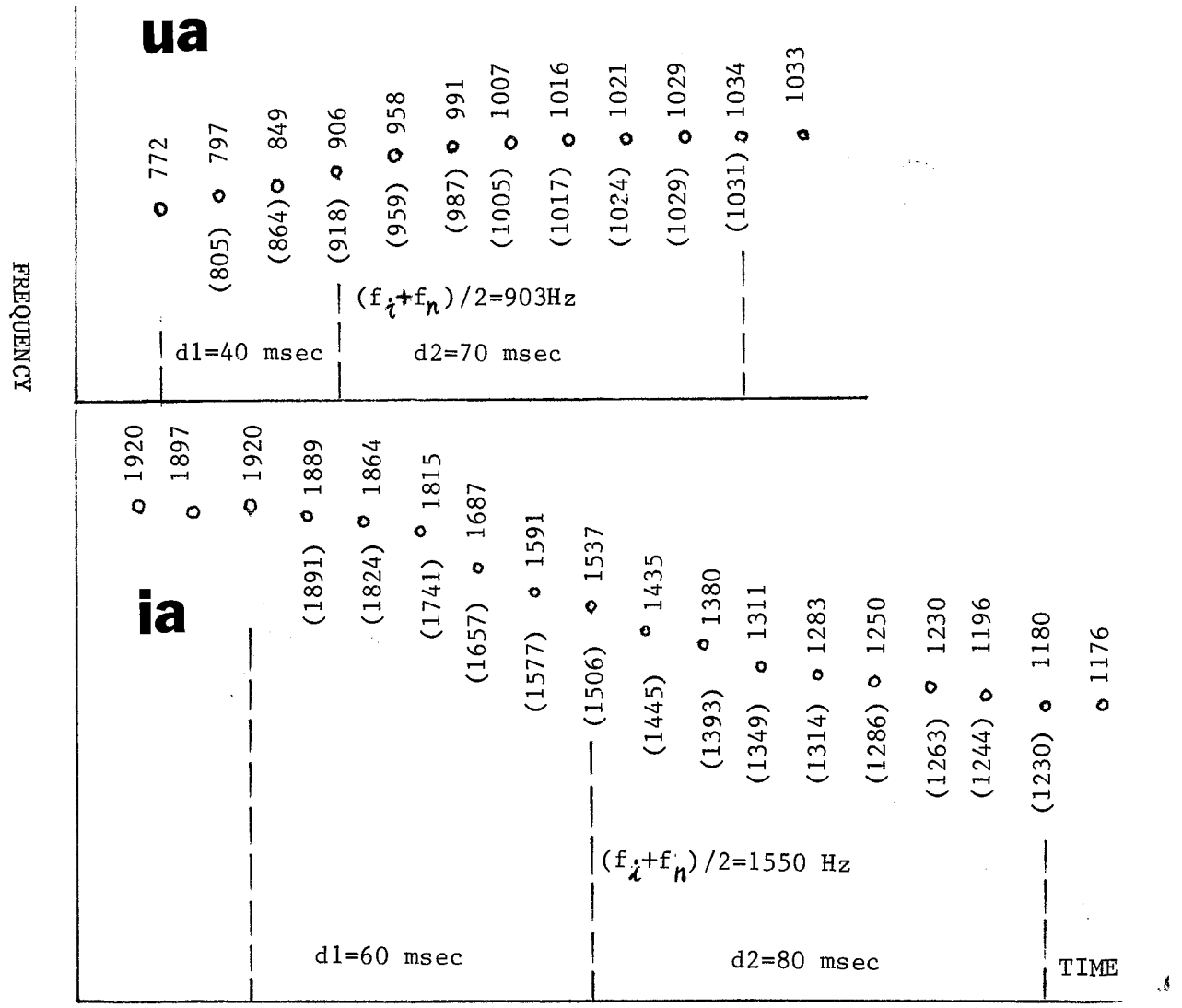


Figure 6.1. F2 transition samples of /ia, ua, ai, au/ from Speaker B1. The F2 frequency values (plotted with o) from the speech sample are given without parentheses; the F2 values calculated using Equation (6.1) are given in parentheses. The beginning, ending points and the mid-frequency points of the transitions are marked with a vertical dotted line. The mid-frequency values (mean of fi and fn), durations d1 and d2 are given as well.

ai
au
1271
(1287)
1287
(1329)
1347
(1384)
1382
(1445)
1414
(1508)
1446
(1569)
1431
(1626)
1646
(1678)
1682
(1725)
1876
(1766)
2005
1992
750
(809)
750
(819)
755
(830)
777
(843)
808
(857)
837
(873)
861
(890)
884
(909)
911
(928)
940
(948)
960
(966)
973
(983)
983
(995)
986
1002
d1=70 msec
d2=30 msec
$(f_i+f_n)/2$=1638 Hz
d1=75 msec
d2=55 msec
$(f_i+f_n)/2$=875 Hz
TIME
FREQUENCY

these tokens, the starting point and the ending point of the transition is marked by a vertical dashed line. In order to investigate the pattern of the distribution of F2 changes over time, the mid-point in frequency of the transition was calculated and the time point with the closest F2 frequency value was determined. The partial durations of the transition before and after this point were also calculated. These aspects of the F2 trajectory will be discussed in the next section.

The shape of the transitions was investigated by means of a time function with a time constant T. The frequency value at time t (taking the time of the starting point of the transition as 0) can be represented as:

$$(6.1)\quad f(t)=fi + (fn - fi)\left\{ 1 - \left(1 + \frac{t}{T}\right) \exp\left(- \frac{t}{T}\right)\right\} u(t)$$

where fi and fn are the initial and final freguency values respectively of a transition; u is a unit function, $u(t) = 0$ when $t < 0$, $u(t) = 1$ when $t > 0$.

Equation (6.1) represents the frequency value of a formant over time, namely f(t), when fi, fn and T are given for a particular diphthong or a two-vowel sequence. To use this equation to predict the formant trajectory over time, we have to determine T first based on the data in real speech. Our LPC data tell us quite directly fi and fn, the beginning and ending F2 frequencies of a formant transition of a given diphthong or hiatus. We have the frequency value at each instant in time as well. Now, we have to solve a new equation for the root T with known fi, fn, t and f(t). Here, fi and fn have unique values for a given diphthong sample; t and f(t) are a variable-function pair. Thus, any point specified with a time value and a frequency value can be used in the equation to provide a T value. For example, in /ua/, read by one speaker at a moderate speed as shown in Figure 1, fi and fn are about 770 Hz and 1035 Hz respectively for the F2 transition. Let us choose one point (t=20msec, f(t)=849 Hz) to provide a new equation with an unknown T:

$$(6.2)\quad 849=770+(1035-770)\{1-(1+20/T)\exp(-20/T)\}$$

To solve this equation, we have to use Newton-Raphson iteration (or Newton's method). It is desirable to calculate the root of this equation on a computer. The calculated T in (6.2) is 18.3 Hz.

The next point on F2 transition in /ua/ in Fig. 1 can be specified as t=30 msec and f(t)=906 Hz. Accordingly, a new equation can be formulated with the only unknown factor T:

$$(6.3)\quad 906=770+(1035-770)\{1-(1+30/T)\exp(-30/T)\}$$

The calculated T value is 17.4 msec. We can calculate T at any point specified by t and f(t) in the same way. The next two points provide T values of 16.1 and 15.4 respectively. In the ideal situation, namely, if the formant transition in real speech can be perfectly represented by equation (6.1), the T value calculated at any point along the formant transition should be the same. However, as we can see in the above example, the T values at different points are similar, not identical. We have to predict a formant transition by equation (6.1) with a T

value which is a mean of those calculated at some selected points in the real speech samples. For instance, the F2 transition in /ua/ can be represented by the following equation with the mean value T=16.8 msec (of the four T values 18.3, 17.4, 16.1, 15.3 msec, SD=1 msec):

$$(6.4)\quad f(t)=770+(1035-770)\{1-(1+t/16.8)\exp(-t/16.8)\}$$

Using equation (6.4), we can predict the F2 frequencies changing over time. The calculated f(t) values at every 10 msec are shown in Fig 6.1, too, within the parentheses.

I will choose 4-6 points in each token in a portion of the transition where frequency is changing most rapidly to provide a mean T, because we want to have our prediction of f(t) using equation (6.1) fit precisely that transition portion in which the frequency is changing more rapidly, since this portion is more sensitive to T variation than the slow frequency change portion in which frequency is changing more slowly.

Whether equation (6.1) fits the real speech samples can be seen by looking at the similarity among the T values calculated along each F2 trajectory in our diphthong samples. Among more than seventy tokens in our pilot study, the mean values of T calculated over a period of 30-40 msec usually have a standard deviation of 0-4 msec; the mean values of T calculated over a period of 50-80 msec usually have a standard deviation of 5-13 msec. A time function of F2 frequency values with a mean T value which deviates only slightly from those calculated with individual T values will also only deviate slightly from the real frequency data. Hence we can consider this time function to fit the real speech well.

We can also show the goodness of fit more directly. Table 6.1 shows the deviations of the calculated F2 values, namely, f(t), from those obtained by LPC F2 tracks for speaker B1.

The mean deviation in F2 values along the F2 trajectory ranges from 1.7% to 10.7 % of the Δ F2 of the syllable. In general, the exponential equation provides a reasonably good approximation. One factor that may decrease the accuracy of the predicted F2 frequency trace of a transition is the asymmetry of the F2 transition course distributed over time. This will be examined in the next section.

In addition to the fact that it fits LPC data from real speech well, the equation is also desirably simple since it can specify a transition in a diphthong or a vowel sequence using only one constant. T is in direct ratio with t in the sense that a formant with a larger T needs a proportionally longer time to reach a given frequency value (as implied in equation (6.1)). In a simple uniform motion, the velocity v is inversely proportional to time t in the sense that an object with a smaller v needs a proportionally longer time to reach a particular place. For the purpose of linguistic discussion, we may treat t in the above two relations as an intermediary relating T and the notion of velocity. A smaller T value means an overall faster movement (or overall fast formant transition rate), while a larger T value means an overall slower movement (or an overall slower formant transition rate).

Table 6.1. Mean deviations of the calculated F2 values from the LPC F2 values in /ia, ua, ai, au/ read by speaker B1 at a slow (S), moderate (M) and fast (F) speech tempo. The F2 values (f(T)) were calculated using Equation (6.1) with a T value in each token based on the average T of at least four LPC points. n is the number of LPC points in a diphthong.

	ia			ua			ai			au		
	S	M	F	S	M	F	S	M	F	S	M	F
Σ /d/ ----- n	44	31	32	36	5	16	103	64	69	19	11	22
% of Δ F2	4.2	4.2	4.5	10.7	1.7	6.5	8.6	7.6	9.4	7.4	2.4	8.7
n	16	14	11	20	10	9	27	15	10	34	34	13
SD	34	19	38	28	5	8	91	34	76	19	8	22

6.4. Distribution of the Frequency Change over Time.

It can be observed in the general pattern of LPC F2 tracks that diphthongs differ in the way that the frequency change is distributed over time. Some diphthongs concentrate the majority of the change toward the beginning of the diphthongs, while others concentrate it toward the end. The duration of the transition preceding the mid-point in frequency was compared with the duration following it. Let us call the first portion "d1" and the remainder "d2". The duration of these portions for each diphthong from each speaker at each speech tempo is given in Table 6.2, together with the difference d1-d2.

Table 6.2. Durations of d1 and d2 (demarcated by the mid-point in frequency) and the difference d1-d2.

	ia			ua			ai			au		
speaker/ /tempos	d1	d2	d1-d2	d1	d2	d1-d2	d1	d2	d1-d2	d1	d2	d1-d2
B1 slow	65	105	-40	70	110	-40	190	90	100	240	160	80
mod.	60	80	-20	30	70	-40	105	55	50	195	155	40
fast	50	70	-20	30	60	-30	70	30	40	85	55	30
B2 slow	50	60	-10	45	65	-20	115	125	-10	175	185	-10
mod.	50	50	0	30	50	-20	115	65	50	110	90	20
fast	45	85	-40	50	60	-10	85	75	10	95	55	40
B3 slow	75	95	-20	25	95	-70	145	115	30	140	140	0
mod.	55	85	-30	55	65	-10	120	110	10	120	110	10
fast	55	65	-10	30	50	-20	95	65	30	100	70	30
B4 slow	50	80	-30	70	100	-30	120	90	30	145	115	30
mod	60	60	0	40	60	-20	90	90	0	170	110	60
fast	55	95	-40	35	45	-10	70	70	0	110	90	20
B5 slow	45	65	-20	30	60	-30	150	40	90	190	80	110
mod.	50	60	-10	40	70	-30	80	70	10	120	80	40
fast	20	30	-10	20	80	-60	75	25	50	60	40	20
B6 slow	45	55	-10	25	95	-70	150	130	20	280	160	120
mod.	35	45	-10	20	50	-30	105	85	20	140	70	70
fast	40	50	-10	40	50	-10	100	70	30	80	60	20
Mean	50.3	68.6		38.1	68.6		110.0	77.8		141.9	101.4	
SD	11.9	19.8		15.1	19.5		32.4	30.1		57.6	43.0	

The four diphthongs can be divided into two groups, one having a transition of fast-then-slow type and the other having a transition of slow-then-fast type. We can easily see from Table 8.1 that, with a few exceptions, (d1-d2) has a negative value for /ia, ua/ and a positive value for /ai, au/. Paired T Tests show that d1 is significantly shorter than d2 for /ia/ ($T=6.007$, $p < 0.001$) and for /ua/ ($T=6.737$, $p < 0.001$). In contrast, d1 is significantly longer than d2 for /ai/ ($T=4.284$, $p < 0.01$) and for /au/ ($T=4.885$, $p < 0.05$). Here, more evidence can be seen that /ia, ua/ are not the same phonological elements as /ai, au/.

The mean of the (d1-d2) differences pooled across speech tempos for each diphthong, and for each of the two transition types established above at different speech tempos are listed in Table 6.3.

Table 6.3. Mean of d1-d2 pooled across (1) each diphthong and (2) each speech tempo for each of the two transition types (/ai, au/ and /ia, ua/).

Diphthong	(d1-d2)	SD	Speech rate		(d1-d2)	SD
/ia/	-18	13	/ia, ua/	Slow	-33	19
/ua/	-31	19		Mod.	-18	12
/ai/	34	27		Fast	-23	16
/au/	44	33				
			/ai, au/	Slow	61	41
				Mod.	32	22
				Fast	27	13

As shown in Table 6.3, /ia/ has the shortest d1-d2 difference and /au/ has the longest. We also note that speech tempo causes differences in d1-d2. The means are -33, -18 and -23 for slow, moderate and fast speech in /ia, ua/ , and 61, 32 and 27 for slow, moderate and fast speech in /ai, au/ (all these numbers are in msec). Slower speech tends to have longer d1-d2, and the mean for /ai, au/ is in fact almost double that at other speech tempos.

Taking the mid-point in frequency as reference, it is rather clear from the data that the F2 transition course is asymmetrical. However, this asymmetry is itself quite regular. Transitions are always slower at the portion close to the low vowel /a/. Note again that the vowel feature [+low] plays a role in the asymmetry of the F2 transition course. Since the low vowel /a/ may occur in either position of a diphthong, the asymmetry is bidirectional in time. Unlike the observed F2 transition pattern, Equation (6.1) demonstrates a time course with one starting point and no ending point. The time course will go closer and closer to an asymptote. In other words, the time course of Equation (6.1) is also asymmetrical in this regard. However, in the time axis, this exponential function is only asymmetrical in one direction. That is probably one of the reasons why the calculated F2 trajectory by Equation 6.1 fits the observed speech data only reasonably well. [1]

Recall that we have discussed the initial steady state of /a/ in the diphthongs /ai/ and /au/ at length in Chapter 4. We found free variations ranging from a pattern of a straight-forward a->i or a->u transition to a pattern of the long steady state /a/ followed by a faster a->i or a->u transition. In an approach with the straight line representation for the F2 transition, we treated the transition from the syllable initiation (namely, from the onset of the /a/ steady state) as an a->i or a->u transition. From the view point of a curve representation for F2 transition, the pattern of an initial /a/ steady state followed by a faster transition in /ai/ or /au/ might well be the extreme case of the asymmetry of the F2 transitions in these two diphthongs, with the very slow rate (sometimes zero rate) at the portion close to /a/.

6.5. T Values for Diphthongs at Different Speech Tempos

Since the mean T values were calculated by pooling across four to six T values at the fastest portion of the transition and since the fastest portion is the most important one in a transition pattern, we can assume that these mean T's are representative for the transitions. The predicted F2 trajectory based on this mean T value will be least affected by the asymmetrical nature of the transition. Table 6.4 shows the mean T values calculated using Equation 6.1.

Table 6.4. Mean T values in msec. pooled across 4-6 T values in the fastest portion of the F2 transitions in /ia, ua, ai, au/ read by 6 speakers at three speech tempos.

	ia			ua			ai			au		
	T	SD	syl.dur	T	SD	syl.dur	T	SD	syl.dur	T	SD	syl.dur
B1 slow	38	2	490	41	6	420	96	18	520	103	18	360
mod.	32	1	280	16	1	310	55	9	450	73	4	410
fast	27	2	190	17	3	190	43	5	160	47	5	190
B2 slow	29	7	420	26	2	400	62	11	480	80	12	540
mod	31	4	230	17	2	230	55	8	230	55	11	250
fast	26	2	230	30	2	270	46	10	230	48	7	250
B3 slow	46	8	440	20	2	350	88	9	450	81	10	280
mod	33	4	280	29	2	300	71	10	280	76	9	300
fast	34	3	170	16	1	180	55	8	180	55	5	190
B4 slow	28	0	360	38	3	270	63	13	350	109	16	290
mod	33	2	290	23	2	240	52	7	290	86	4	400
fast	31	3	230	21	2	190	33	3	230	62	8	250
B5 slow	31	3	320	17	1	290	94	6	350	96	13	430
mod	33	7	240	36	11	210	52	3	210	71	11	220
fast	11	2	170	22	3	170	30	8	170	27	6	190
B6 slow	25	1	440	18	3	470	87	9	460	151	23	470
mod	20	5	260	17	1	270	68	7	320	57	14	260
fast	27	4	190	21	3	200	59	3	240	55	7	270
Mean	29.7		290.6	23.6		275.6	61.6		312.2	74.0		308.3
SD	7.3		100.3	8.0		87.5	19.5		115.6	28.6		103.2

Based on the data in Table. 6.4, the correlation between the mean T value and the syllable duration was calculated for each diphthong. The correlation for /ai/ (r=0.7284) is much higher than that for /ia/ (r=0.4558). The correlation for /au/ (r=0.6849) is much higher than that for /ua/ (r=0.2132). This fact confirms the findings in the previous chapter that the transition rates are more tempo-sensitive in /ai/ and /au/ than in /ia/ and /ua/. The correlations between

T and syllable duration are similar to that between F2 transition rate and syllable duration in /ai/ and /au/. However, a considerable increase can be found in the correlations between T and syllable duration in /ia/ and /ua/, compared with that between F2 transition rate and syllable duration.

Arranging the data in another way, we have the mean values of T for each diphthong at the different speech rates as shown in Table 6.5.

Table 6.5. Means of T from each diphthong at different speech tempos.

Diphthong across speech rate			for each speech rate			
	T	SD			T	SD
/ia/	29.7	7	/ia/	Slow	32.8	8
				Mod.	30.3	5
				Fast	26.0	8
/ua/	23.6	8	/ua/	Slow	26.7	11
				Mod	23.0	8
				Fast	21.2	5
/ai/	61.6	19	/ai/	Slow	81.7	15
				Mod	58.8	8
				Fast	44.3	12
/au/	74.0	29	/au/	Slow	103.3	26
				Mod	69.7	12
				Fast	46.5	14

It can be seen in Table 6.5 that the mean T values in the three different tempos are similar (also with similar SD) in /ia/ and /ua/. There is no significant difference among different tempos. However, great difference can be found in mean T values among the three different tempos in /ai/ and /au/. T Tests show a highly significant pattern for /ai, au/, which can be represented schematically as:

(6.5) T(slow) > T(moderate) > T(fast).

The data here confirm the finding in Chapter 5 that, in diphthongs, the transition from a low vowel is tempo-sensitive, while that toward a low vowel is relatively tempo-insensitive.

6.6. T Values for Inter- and Intra-syllabic Transitions

In this section, a comparison will be made between the T values for syllable-internal transitions and those for intersyllabic (cross-syllable boundary) transitions in Chinese. If transitions are physiologically governed we would again expect no difference in T values to be found in these cases.

We measured and calculated d1, d2 and T in disyllabic hiatus /i#a, u#a, a#i, a#u/ as well as in their diphthongal counterparts. The results are given in Table 6.6.

Table 6.6. Means of d1, d2, d1-d2 and T in msec. in disyllabic hiatus at a moderate speech tempo, read by six speakers.

	i#a					u#a					a#i					a#u				
	d1	d2	d1-d2	T	SD	d1	d2	d1-d2	T	SD	d1	d2	d1-d2	T	SD	d1	d2	d1-d2	T	SD
B1	70	120	-50	41	6	40	40	0	25	1	75	65	10	45	8	60	70	-10	32	4
B2	45	65	-20	24	3	65	65	0	39	0	65	65	0	40	3	55	65	-10	34	2
B3	80	70	10	47	7	30	80	-50	18	1	45	85	-40	26	1	55	55	0	29	3
B4	80	100	-20	47	12	30	40	-10	20	3	100	50	50	46	15	75	45	30	43	3
B5	90	120	-30	62	14	40	60	-20	24	1	100	90	10	42	7	60	70	-10	33	3
B6	60	70	-10	40	7	65	75	-10	51	5	65	45	20	36	5	85	85	0	47	4
M	71	91		43.5		45	60		29.5		75	67		39.2		65	65		36.3	
SD	16	26		12		16	17		12		22	18		7		12	14		7	
(T in diph.)				29.7					23.6					61.6					74.0	
(SD)				7					8					19					29	

The patterns obtained from the individual speakers in these V#V sequences differ quite markedly. However, taking the means across speakers, one main observation emerges, namely, that the transition rate (reflected by the value T) is very similar for all four types of two-vowel sequences. Comparing the data in disyllabic hiatus with those in diphthongs, we can find some differences. First, there is no significant difference between d1 and d2 in any of the hiatus sequences. Second, T in the diphthongs /ai/ (58.8 msec, SD=8) and /au/ (69.7 msec, SD=12) are larger than T in hiatus sequences /a#i/ (39.2 msec, SD=7; $p < 0.005$) and /a#u/ (36.3 msec, SD=7; $p < 0.001$) respectively. However, T in the diphthongs /ia/ (30.3 msec, SD=5) and /ua/ (23.0 msec, SD=8) are not significantly different from T in hiatus sequences /i#a/ (43.5 msec, SD=12) and /u#a/ (29.5 msec, SD=12).

Again, we see a distinction between the two groups of diphthongs, concerning the difference in time constant T for the F2 transitions between the inter- and intra-syllabic position. The diphthongs /ia/ and /ua/ have a T which is not significantly different from that in /i#a/ and /u#a/ respectively while /ai/ and /au/ have a T which is significantly larger than that in /a#i/ and /a#u/ respectively. The data support the analysis of /ia, ua/ as being phonetically [ja, wa].

To conclude, the transition specifications are highly controlled linguistically. In addition to the cross-language difference in transition rates, a language may exhibit different specifications for different linguistic functions as regards syllable structure and boundary. Phonological ordering (associated with presence or absence of consonant elements [j, w]) and vowel features both play a role in the specification of transition rates. With respect to the low vowel, the transition from a low vowel is tempo-sensitive, has an initial slower rate portion (or steady state) followed by a faster rate portion, and differs in rate from the sequence across hiatus with the same components. In contrast, the transition toward a low vowel is relatively tempo-insensitive, has a faster portion followed by a slower rate portion, and does not differ in rate from the corresponding vowels in hiatus. Phonetically, the transition originates from a glide [j] or [w].

Footnote:

[1]. Equation (6.1) in the text (referred to as equation (1) in the footnote) represents a curve beginning with t=0 and f(t)=f1, but never reaches f2. This can be proved by the following procedure: assume that the curve have the value f2 at the time t2. Then, according to equation (1), f(t) could be equal to f2 only if

(2) $1-(1+t2/T)\exp(-t2/T)=1$

we get (3) from (2):

(3) $(1+t2/T)\exp(-t2/T)=0$

It should be either (4) or (5), derived from (3)

(4) $1+t2/T=0$

(5) $\exp(-t2/T)=0$

We may obtain a negative value t2=-T from (4), which would be ruled out by the unit function u(t) in equation (1). We may also obtain t2=-T(In 0). This is ruled out either by the wrong form (In 0) or by the unit function.

This fact indicates that t2 as the time where the curve attains f2 does not exist; we only have a curve getting closer and closer to f2. This property of equation (1) causes no problem for fitting the transition of fast-then-slow type as in /ia, ua/ ([ja, wa]). For slow-then-fast type, we need a factor put into equation (1). Taking /ai/ as an example, we assume that a new curve having the same variable t

(6) $at^2 + bt + c$

has to be superimposed onto (1) so that the curve for (1) can keep its original form at the first part of the transition and begins to raise its frequency value slowly after the mid point M (about 100 msec for /ai/) and increase by 10% more at the end of the transition. This can be represented by (7) and (8):

(7) $f(t) = at^2 + bt + c$

(8) f(0) = 1
    f(100) = 1.01
    f(200) = 1.1

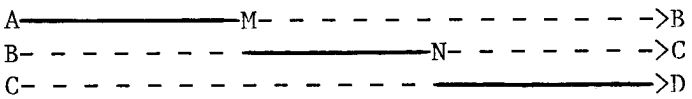
The solutions will be a=0.000004, b=-0.0003, c=1. Thus, we can modify equation (1) to (9) for better approximation in /ai/:

(9) $f(t)=f1+(f2-f1)\{1-(1+t/T)\exp(-t/T)\}(0.000004t^2-0.0003t+1)u(t)$

CHAPTER 7. CONCLUSION

Phonetic orthodoxy treats the acoustic realization of a syllable with complex vocalic elements as the concatenation of consecutive transitions from one element to the following one in the syllable. That is to say, for a syllable consisting of four elements A, B, C and D (n.b. the discussion does not include the syllables with post-nucleus cluster), the F2 trajectory as the most important acoustic feature, would simply be the concatenation of the A->B, B->C and C->D transitions, just as illustrated in Figure 7.1a. This study proposes an alternative view of the acoustic structure of this type of complex syllable based on LPC trackings of F2 in sets of selected Chinese and English syllables. The materials in this study suggest that the acoustic realization patterns can be accounted for by a new model, the truncation model, which may be schematically represented as in Figure 7.1b.

Here, the underlying A->B, B->C and C->D transitions all originate at the same temporal position -- the syllable initiation. If C is a nucleus and D is a post-nucleus element, the C->D transition may have an initial (quasi-)steady state. The first two transitions, A->B and B->C, truncate each other at point M, where they intersect. In other words, what is phonologically the first transition, A->B, connects what is phonologically the second transition, B->C, at the intersection point M, which is neither the offset target of the first transition nor the onset target of the second transition. Similarly, the second and third transitions, namely, the B->C and C->D, truncate each other at the intersection point N. The B->C transition has its realized portion between M and N, while the C->D transition only preserves the portion after N. Therefore, the F2 pattern of the syllable consisting of four phonological units, ABCD, would not be A->B->C->D but rather A-M-N->D, or more specifically as in the diagram below (n.b., only the horizontal dimension, the time, is relevant in this diagram):



In Figure 7.1b, the points M and N are actually the undershot B and C targets respectively. In a syllable having two adjacent rising (or falling) transitions the truncation process may result in an overshot target. The undershooting and overshooting of the target realizations can be conceptually explained and quantitatively predicted by the truncation model. Therefore, the F2 trajectory of the syllable can be determined in this approach.

In addition to the application of the truncation process, a good prediction depends upon two types of specification. The first is the proper specifications of underlying target values. If a language allows a syllable with the sequence ABCD (n.b. without final non-syllabic cluster), then, it also allows a BCD syllable and a CD syllable. Thus, the unrealized B and C targets in the syllable ABCD can be determined from the realized (and thus measurable) values in the simpler syllables BCD and CD respectively.

The second type of specification for a good prediction is the F2 transition rate for each pair of adjacent targets at different speech tempos.

Let us talk about the specifications of F2 transition rate first. With the target value known, the rates of the relevant F2 transitions play a central role in

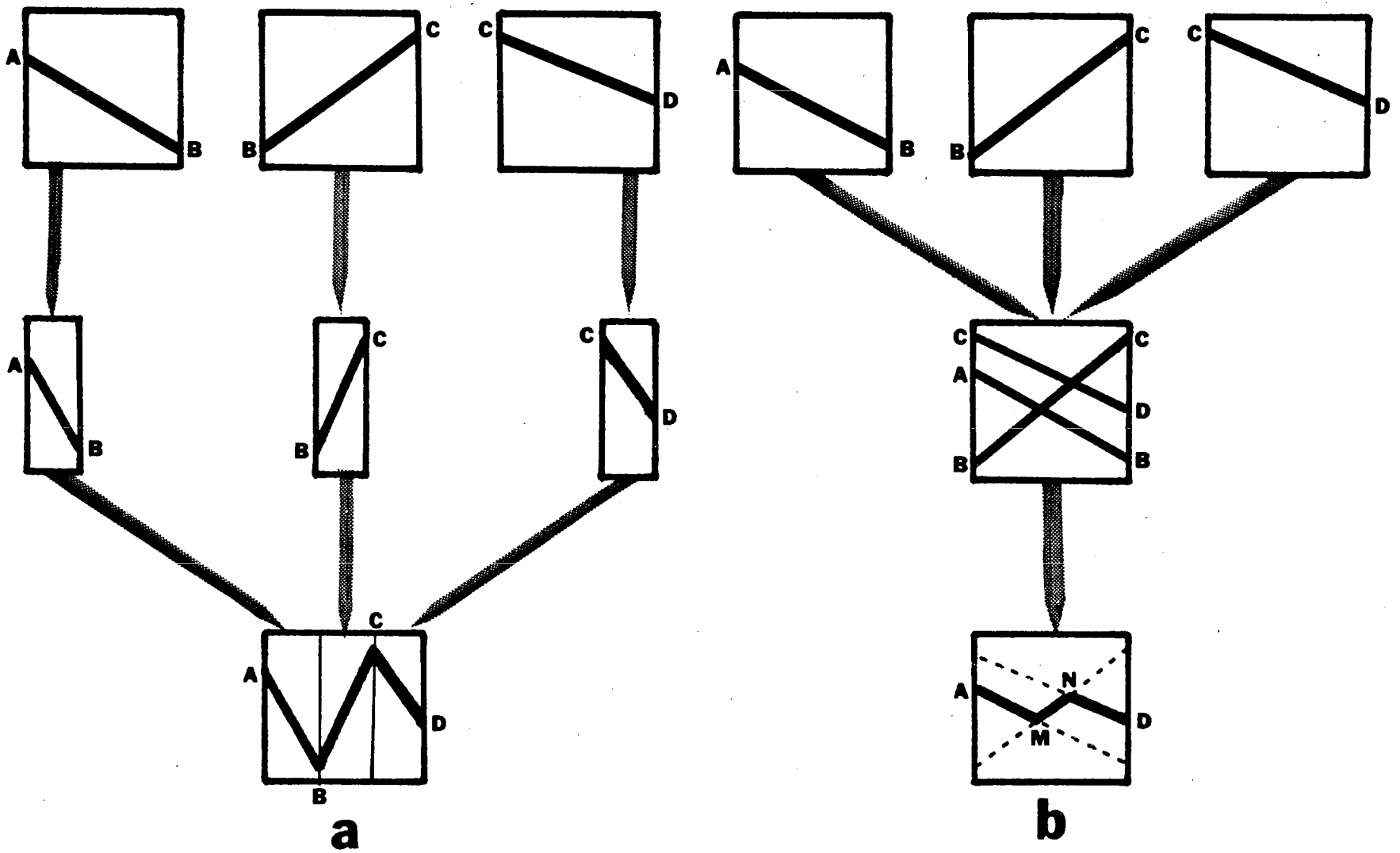


Figure 7.1. Two different views concerning the acoustic realization pattern of a syllable consisting of four components A, B, C and D. Suppose that there is no final non-syllabic cluster in this syllable.

determining the F2 trajectory in the syllable. Two issues concerning the specification of F2 transition rate have been investigated in this study. One is what determines the different rates for different diphthongs (or say, for different pairs of targets). In other words, we must know the cause of the diversity in transition rates among diphthongs. The other is what factors affect the F2 transition rate for a given diphthong. In other words, we try to find the cause for the variations in the individual diphthongs.

It has been found in the nine Chinese diphthongs that the transition rate is only moderately correlated with the difference between two target values (Δ F2). The Kent and Moll hypothesis of "the further the faster" is at work as a loose principle in the acoustic domain of the Chinese diphthongs. The correlation between the F2 transition rate and Δ F2 might be diminished by one strong factor -- the phonological ordering of the two targets. It appears that both phonological orders of a given pair of targets have deviated rate/Δ F2 ratios from the overall mean rate/Δ F2 ratio (0.0058) of the nine Chinese diphthongs.

Scrutinizing the rate/ Δ F2 ratios for different pairs of targets in opposite phonological orders, we find that some vowel features play a role with respect to patterning the rate/ Δ F2 ratios. When the vowel feature [low] is positively specified in the pair, the transition from a high vowel has much greater rate/Δ F2 ratio (about 0.008) than that of the transition toward a high vowel (about 0.004). The opposite result occurs in the diphthongs where there is no low vowel (except ou/uo pair). Taking the rate/ Δ F2 ratio 0.004 as the basic ratio, the F2 transition rates for /ai, au, ie, ye, ou, uo/ can be derived from the F2 range, and the those for /ia, ua, ei/ can also be derived by doubling the ratio. The fact that /ia, ua, ei/ have considerably faster transition rates than their counterparts with the reversed order suggests that they are phonetically [ja, wa, ei]. Leaving these three phonetical glide+vowel combinations out, the Kent and Moll hypothesis works quite well for the Chinese diphthongs.

It has been noticed that the specifications of F2 transition rates presented above for different diphthongs are language-specific. In Spanish, the phonological order of diphthong components does not have a significant influence on the F2 transition rate. The rates are similar or even sometimes identical for the two transitions with the same two targets in opposite orders. Furthermore, the difference in F2 transition rates is surprisingly great between some 'similar' diphthongs in Chinese and Spanish. For example, the rate of the a->u transition is four times as fast for Spanish (r=6.47 Hz/msec.) as for Chinese (r=1.62 Hz/msec.). It is clear that the specifications of F2 transition rates for different diphthongs are not automatic but linguistically controlled.

As regards the second issue concerning F2 transition rate, namely, the factors affecting the rate for a given F2 transition, we have focused on the possible variation of the F2 transition rates at different speech tempos. The data have shown that speech tempo does affect the F2 transition rates. There is in fact a range of F2 transitions, from the most tempo-sensitive to the relatively tempo-insensitive F2 transitions.

It is interesting that, in many diphthong pairs within which the components are in opposite orders, the F2 transition rate in one member is tempo-sensitive while that in the other is tempo-insensitive. Here again, the vowel feature [low] is relevant in patterning the variations at different speech tempos. When a low vowel occurs in a diphthong, the transition toward a high vowel tends to be tempo-sensitive while the transition from a high vowel tempo-insensitive. When

there is no low vowel in a diphthong, the transition from a high vowel (actually a glide) tends to be tempo-sensitive while the transition toward a high vowel tempo-insensitive. The fact that the same set of diphthongs, /ia, ua, ei/, stands out as the most tempo-insensitive supports the analysis that they are actually glide-vowel combinations.

In addition to the speech tempo, the complexity of the syllable may affect the F2 transition for a given pair of targets. Surprisingly, the F2 transition rate for a given pair of targets slows down when the number of the syllable components increases. For example, the e->l and w->e transitions in the English word /dwel/ are slower than the corresponding transitions in words /el/ and /wel/ respectively.

Other factors affecting the transition rate for a given pair of targets include the phonological positions of the two targets vis-a-vis a syllable boundary. In the case of diphthongs having a low vowel component, the rates for the F2 transitions from a high vowel (a glide) are quite similar in both syllable internal position and syllable boundary position. Also in this case the rates are tempo-insensitive and have a greater rate/Δ F2 ratio. However, the rates for the F2 transitions toward a high vowel are much slower in syllable internal position than in cross-syllable boundary position. In this case, the rates are tempo-sensitive and have a smaller rate/Δ F2 ratio.

So far we have summarized the specifications of mean F2 transition rates in diphthongs. For the acoustic realization of a syllable, the F2 transition shapes must also be specified. In the above discussion of rate specifications, we assume that the transition can be approximated as a straight line. However, some variations which make the transition differ considerably from the approximation of a straight line have also been a concern in this study.

In the diphthongs /ai/ and /au/, we found that the a->i and a->u transitions vary from a straight trajectory originating at the syllable initiation to a complex trajectory with an initial steady state /a/. There are intermediate states between these two variations. That is, the F2 trajectory can be a slower rising portion followed by a faster rising portion. In this study, the rate of the a->i and a->u transitions is defined as that for a transition from the syllable initiation. In a separate test in this study, the onset and offset of the F2 transitions in these diphthongs, as well as the corresponding diphthongs with opposite component order, /ia/ and /ua/, were carefully defined by a numerical criterion, namely, a constant difference between two LPC F2 values with a 10 msec interval. The results showed that the F2 transitions in these diphthongs were asymmetrical with regard to the mid-point in the F2 range. This asymmetry is itself rather regular. Transitions are always slower at the portion close to the low vowel /a/. Thus, a probable long steady state /a/ in the a->i or a->u transitions can be treated as the extreme case of this type of asymmetry. Note that there might be no asymmetry in the articulatory domain.

An attempt has been made in this study to use an exponential curve representation to match the F2 trajectory in the diphthongs /ia, ua, ai, au/. In general, the function can provide a reasonable approximation of the F2 trajectory. The time constant T was used as a measure of transition rate, and the complex patterns of the tempo effects on F2 transitions were confirmed in this approach. However, the exponential function curve has only its onset fixed, namely, the first target of the transition, while its offset goes infinitely close to an asymptote. The accuracy of the fit of the exponential curve is decreased by the asymmetry in the slow-plus-fast type of F2 transitions, e.g., the a->i and a->u transitions.

Let us return to the truncation model of the F2 structure in syllables. For a first approximation of F2 pattern prediction based on the phonological transcription of the syllable at a given tempo, we need a derivation which can be briefly stated as follows: The underlying target values can be specified by referring to the measured value in the simplest syllable. All the tauto-syllabic targets corresponding to pre-nucleus element(s) and nucleus element are located at the syllable initiation (usually the vowel onset). The transition rate between each pair of adjacent targets can be specified according to the rate/ Δ F2 ratio as well as the degree of sensitivity to the speech tempo. The tempo-insensitive transition rate can be specified merely by the rate/ Δ F2 ratio, while the most tempo-sensitive transition rates can be derived by r= Δ F2 / D_s (i.e., the F2 range divided by the syllable duration). A possible steady state may occur at the initial portion of a transition from a nucleus target to a post-nucleus target, such as in /ai/ and /au/. With all the transition rates and shapes specified, the truncation process takes place. The acoustically realized pattern will be similar to that in Figure 7.1b.

Let us go back to the issue of the phonetic characteristics of rising/falling diphthongs mentioned in Chapter 1. When we examine the dynamic property of acoustic patterns, the distinction between rising and falling diphthongs is remarkable in the case of /ia, ua/ versus /ai, au/. That is, the F2 transitions in these two rising diphthongs are relatively fast, tempo-insensitive and similar to the transition across syllable boundary; the F2 transitions in /ai, au/ are relatively slow, tempo-sensitive and considerably slower than those across syllable boundary. Since /ei/ is phonetically a rising diphthong [e̯i], it shares the same properties as /ia, ua/([ja, wa]).

This phonetic property of rising diphthongs is less obvious in /ou, ie, ye/ than /ia, ua, ei/. The diphthongs /ou, ie, ye/ are phonologically treated as combinations of a medial and a nucleus according to the assumption that the lowest (in terms of vowel height) vocalic component is the nucleus. Phonetically, they have a relatively slow and tempo-sensitive rate, and thus are grouped with the falling diphthongs /ai, au/. There must be some unknown factors involved in their phonetic realizations. One possible factor is that they all have a non-peripheral vowel (which is not /a/, /i/ or /u/) as their 'nucleus'. Therefore, their ending target is not the extreme value in the vowel space and the F2 range is relatively small. It is not very clear which element in these diphthongs is phonetically the syllabic or nucleus. Many tokens show a quasi-steady state of /i/ in /ie/. This may suggest that /ie/ is phonetically a falling diphthong.

With our phonetic data, we can summarize the dynamic characteristcs of phonologically defined rising and falling diphthongs as follows. The rising diphthongs with a peripheral vowel /a/, /i/ (or /u/?) as the nucleus or syllabic element have relatively fast and tempo-insensitive F2 transition rates. The rising diphthongs with a non-peripheral vowel, e.g., /o/ or /e/, and the falling diphthongs have a relatively slow, tempo-sensitive F2 transition rate.

The phonetic data gathered here show discrepancies between the phonological and phonetic forms, and thus call for some revision of the phonological analysis. The initial component at the onset of a fast and tempo-insensitive transition (in /ia/, /ua/ or /ei/) is actually a consonant ([j], [w] or [e̯]). The diphthong /ei/ has a dynamic characteristic which is markedly different from /ie, ou, uo/. In the phonology, /ei/ has been shown to be different from /ie, ou, uo/ in a very important phonological process in this language -- r-suffixation (Wang and He,

1985). In this process, /ie, ou, uo/ plus /r/ result in simple concatenation patterns /ier, our, uor/. However, /ei/ plus /r/ becomes /(e)r/. This again confirms that /ei/ is actually different in phonological and phonetic status from the other diphthongs which involve non-peripheral vowels. Based on observed phonetic facts of initial consonantal elements [w] and [e], we need a new analysis of the syllable structure in Chinese which allows two medial components in a syllable such that [w] and [e] in /uei/ ([wei]) can be accounted for. This is not a new proposal. It has been suggested in some phonological works on Chinese phonology (Hartman, 1944; Hockett, 1947; Pulleyblank, 1986) that a semi-vowel sequence /jw/ (or transcribed as /iu/) occurs before a nucleus. Finally, phonetic data challenge the phonological treatment of some non-peripheral vowels such as /e/ in /ei/, /uei/ and /ie/, /o/ in /iou/ as the nucleus of the syllable.

BIBLIOGRAPHY

Abercrombie, D. 1967. Elements of General Phonetics. Edinburgh University Press.

Alfonso, P. and Baer, T. 1982. Dynamics of Vowel Articulation. Language and Speech 25.2: 151-73.

Allen, J. 1983. Units in Speech Synthesis. Symposium 2. Proceedings of the Tenth International Congress of Phonetic Sciences. Dordrecht: Foris Publications: 151-156.

Anderson, H. 1972. Diphthongization. Language 48.1.

Beinum, F. J. K-V B. and M. H. Reitsma, 1976. The influence of the diminutive suffix on the preceding vowel. Proceedings of the Institute of Phonetic Sciences, University of Amsterdam, No. 4.

Bladon, R. A. W and A. Al-Bamerni, 1976. Coarticulation resistance in English /l/. Journal of Phonetics 4:137-50.

Bond, Z. S. 1978. The effects of varying glide durations on diphthong identification. Language and Speech 21.3.

Canh, N-P. 1974. A contribution to the phonological interpretation of the diphthongs in Modern Vietmanese. Acta Universitatis Cavolinae -- Philologica. Phonetica Pragensia IV. 133-142.

Cao, J. and S. Yang, 1984. An experimental study on diphthongs in Beijing Chinese (in Chinese). Zhongguo Yuwen.

Chan, M. K. M. 1986. On the status of "basic" tones. UCLA Working Papers in Phonetics 63:71-94.

Cheng, C. C. 1973. A Synchronic Phonology of Mandarin Chinese. The Hague.

Collier, R., Bell-Berti, F. and Raphael, L., 1982. Some acoustic and physiological observations on diphthongs. Language and Speech 25.4:305-23.

Delattre, P. 1965. Comparing the phonetic features of English, German, Spanish and French. Heidelberg: Julius groos Verlag.

Dolan, W. B. and Yoko Mimori, 1986. Rate-dependent variability in English and Japanese complex vowel F2 transitions. UCLA Working Papers in Phonetics 63:125-153.

Donegan, P. J. 1978. On the natural phonology of vowels. Working Papers in in Linguistics, Ohio State University. No. 23.

Dragonov, and Dragonova, 1955. Syllable structure of Standard Chinese (in Russian). The Chinese translation in Zhonguo Yuwen, 1958:11, 513-521.

Fowler, C. 1980. Coarticulation and theories of extrinsic timing control. Journal of Phonetics 80.8: 113-133.

Fromkin, V. 1985. (ed) Phonetic Linguistics, Essays in Honor of Peter Ladefoged. Orlando: Academic Press.

Fu, M. 1956. The phonemes in Beijing dialect and the Pinyin symbols (in Chinese).

Fujimura, O. 1981. Temporal organization of articulatory movements as multidimensional phrasal structure. Phonetica 38:66-83.

Fujimura, O. and Lovins, B., 1978. Syllables as concatenative phonetic units. syllables and segments. Bell et al, eds. North-Holland. 107-20.

Fujisaki, H., and Higuchi, N., 1979. Temporal organization of segmental features in Japanese disyllables. Proceedings of 9th international Congress of Phonetic Sciences, Copenhagen.

Garnes, S. 1976. Quantity in Icelandic: production and perception. Hamburger Phonetische Beitrage.

Gay, T. 1968. Effect of speaking rate on diphthong formant movement. JASA 44:1570-1573.

Gay, T. 1970. A perceptual study of American English diphthongs. Lang. Speech 13:65-88.

Gay, T. 1978. Articulatory units: segments or syllables? in Syllables and Segments ed. by A. Bell and J. B. Hooper. North-Holland Publishing Company.

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

Gerber, S. E. 1972. Perception of segmented diphthongs. Proceedings of the Seventh International Congress of Phonetic Science. Mouton, Le Hague, Paris.

Ginesy, M. and D. J. Hirst, 1975. Formant transitions and pitch-change in English diphthongs. Travaux de l'Institut de Phonetique d'Aix, Vol.2: 141-148.

Gleason, H. A. 1955. An Introduction to Descriptive Linguistics. New York: Henry Holt and Co.

Halle, M., 1964. On the bases of phonology. In The structure of Language. (Fodor and Katz eds). New Jersey: Prentice Hall. 604-612.

Han, M. S. 1968. Complex syllable nuclei in Vietnamese. Studies in the phonology of Asian languages, VI, University of Southern California.

Harris, K. S. 1984. Coarticulation as a component in articulatory description In Articulation Assessment and Treatment Issues. ed. by R. G. Daniloff. College-Hill Press.

Hartman, L. M. 1944. The segmental phonemes of the Peiping dialect. Language 20. I: 28-42.

Haudricourt, A. G. and J. M. C. Thomas. 1976. La notation des langues, phonetique et phonologie. Section of Vietnamien. Paris, p.125.

Hawkins, S. 1973. Temporal coordination of consonants in the speech of children: preliminary data. Journal of Phonetics 73.1: 181-217.

Heffner, R-M. S. 1949. General Phonetics. Madison.

Hibbitt, G. W. 1948. Diphthongs in American Speech. New York, Privately published.

Hockett, C. F. 1947. Peiping Phonology. Journal of American Oriental Society 67.4: 253-267.

Holbrook, A. and G. Fairbanks, 1962. Diphthong formants and their movements. Journal of Speech and Hearing Research 5.1.

Holmes, J. N. 1983. Speech Technology in the Next Decades. Proceedings of the Tenth International Congress of Phonetic Sciences, Dordrecht: Foris Publications, p. 125-40.

Hsueh, F. S. 1980. The phonemic structure of Pekinese finals and their r-suffixation. Bulletin of the Institute of History and Philology 80:9.

Huang, B and X. Liao, 1981. Modern Chinese (in Chinese).

Huang, Tai-Yi, Cai-Fei Wang and Yoh-Han Pao. 1982. A Chinese Text-to-speech synthesis system based on an initial-final model. IEEE 82: 1601-3.

Jones, D. 1922. Outline of English Phonetics. Second edition. New York.

Karlsson, F. 1970. Det Finska hogsprakets diftonger och vokalkombinationer. Publications of the Phonetics Department of the University of Turku. No. 9.

Keating, P. 1984. Phonetic and phonological representation of stop consonant voicing. Language 60.2.

Keating, P. 1986. CV phonology, experimental phonetics and coarticulation. UCLA Working Paper in Phonetics. No.62.

Kent, R. D. and K. L. Moll, 1972. Tongue body articulation during vowel and diphthong gesture. Folia Phoniat 24:278-300.

Kent, R. D. and F. D. Minifie, 1977. Coarticulation in recent speech production models. Journal of Phonetics 5:115-133.

Kozhevnikov, V.A. and Chistovich, L.A., 1965. Rech: Artikulyatsiya i Vospriyatiye (Moscow-Leningrad). Trans. Speech: Articulation and Perception. Washington, D.C.: Joint Publication Research Service, No.30, 543.

Ladefoged, P. 1975. A Course in Phonetics. New York: Harcourt Brace Jovanovich.

Ladefoged, P. 1977. The Abyss between Phonetics and Phonology. Chicago Linguistic Society papers.

Ladefoged, P. 1979. The phonetic specification of the languages of the world. Revue de Phonetique Appliquee 49-50:21-39.

Ladefoged, P. 1980. What are linguistic sound made of? Language 56.3: 485-502.

Ladefoged, P. 1984., Out of chaos comes order; physical, biological, and structural patterns in phonetics. Proceedings of the Xth International Congress of Phonetic Science. Foris Publications, 83-95.

Lehiste, I. 1967. Diphthongs versus vowel sequences in Estonian. Proceedings of the Sixth International Congress of Phonetic Sciences, Prague.

Lehiste, I. and G. E. Peterson, 1961. Transitions glides and diphthongs. JASA 33.3:268-277.

Li, Yenrui., 1984. A review of research on Putonghua phonemes. (in Chinese). Zhongguo Yuwen, 1984.4.

Li, Zhaotong and Xu, Siyi., 1981. Introduction of Linguistics. (in Chinese). China.

Liljencrants, J. 1968. The OVE III Speech Synthesizer. IEEE Transactions, Electroacoustics, AU-16:137-140.

Lindau, M, 1985. Hausa vowels and diphthongs. UCLA Working Papers in Phonetics 60:40-54.

Lindau, M and K. Norlin and J-O Svantesson, 1985. Cross-linguistic differences in diphthongs. UCLA Working Papers in Phonetics 61:40-44.

Lindblom, B. 1983. On the Origin and Purpose of Discreteness and Invariance in Sound Patterns. Draft.

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

MacNeilage, P., and P. Ladefoged, 1976. The production of speech and language. In Handbook of Perception, Vol. VII.

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

Malmberg, B. 1963. Structural Linguistics and Human Communication. Berlin: Springer-Verlag.

Malmberg, B. 1974. Manuel de Phonetique Generale. Paris: Edition Picard,

Manrique, A. M. B. de, 1976. Acoustic study of /i, u/ in the Spanish diphthongs. Language and Speech 19: 121-127.

Manrique, A.M.B de, 1979. Acoustic Analysis of the Spanish Diphthongs. Phonetica 36:194-206.

Martin, S. E. 1957. Problems of hierarchy and indeterminacy in Mandarin phonology. Bulletin of the Institute of History and Philology. Taipei, V.29.

Mattingly, I.G. 1974. Experimental methods for speech synthesis by rule. In T.A.Sebeok ed. Current trends in Linguistics. Vol 12. The Hague: Mouton.

Mattingly, I.G. 1981. Phonetic representation ans speech synthesis by rule. In The Cognitive Representation of Speech. ed.by Laver et al. North-Holland. 415-20.

Metreal, J-P. 1970. Description phonologique du dialecte de Gessenay (Saanen). Edition Herbert.

Moses, E. R. Jr. 1964. Phonetics (History and Interpretation). Englewood Cliffs, N. J., Prentice-Hall.

Ohman, S. E. G. 1966. Coaticulation in VCV utterances: spectrographic measurements. JASA 39:151-168.

Perkell, J. S. 1977. An overview of articulatory modeling and summary of the discussion. Proceedings of Articulatory Modeling Symposium. Grenoble.

Petursson, M. 1972. Peut-on interpreter les donnees de la radiocinematographie en fonction du tube acoustique a section uniforme? Travaux de l'Institut de Phonetique de Strasbourg. No. 4.

Piir, H. 1982. Recognition of Estonian Diphthongs. Academy of Sciences of the Estonian SSR, Division of Social Sciences. Tallinn.

Piir, H. 1983. On Estonian diphthong space. Abstracts of the Tenth International Congress of Phonetic Sciences. Dordrecht: Foris Publications.

Piir, H. 1983. Acoustics of the Estonian diphthongs. Estonian Papers in Phonetics, 82-83, Tallinn, 5-96.

Pike, K. L. 1947. On the phonemic status of English diphthongs. Language 23:151-159.

Pulleyblank, E. G. 1986. CV phonology and diachronic change as illustrated in the history of Chinese. Paper for the XIXth International Conference on Sino-Tibetan languages and Linguistics, Columbus, Ohio.

Rabiner, L., 1968. Speech Synthesis by Rule: An Acoustic Domain Approach. Bell System Tech. J. 47:17-37.

Ren, Hongmo, 1983. A Linguistic Model for Duration in Chinese. UCLA M. A. Thesis.

Ren, Hongmo, 1986. Linguistically conditioned duration rules in a timing model for Chinese. UCLA Working Papers in Phonetics 62:34-49.

Romeo, L. 1968. The Economy of Diphthongization in Early Romance. Janua Linguarum, Series Practica LV. The Hague-Paris.

Santerre, L and J. Millo, 1978. Diphthongization in Montreal French. in Linguistic Varation: Models and Methods, ed. by D. Sankoff, Academic Press.

Schwartz, R., Klovstad, J., Makhoul, J., Klatt, D. and Zue, V., 1979. Diphon synthesis for phonetic vocoding. IEEE 79:891-4.

Shuken, C. R. 1977. Syllable and diphthong identification in some Scottish Gaelic. Work in Progress. Department of Linguistics. Edinburgh University. No. 10.

Solomon, I and S. J. Sara, 1983. English diphthongs [ai, oi, ou]. Proceedings of the Tenth International Congress of Phonetic Sciences. Dordrecht: Folia Publication.

Suen, Ching Y., 1976. Computer Synthesis of Mandarin. IEEE 76:698-700.

Sweet, 1877. A handbook of Phonetics. Oxford: Clarendon Press.

Tatham, M. A. A., 1970. A speech production model for synthesis-by-rule. Working papers in Linguistics, No.6, Computer and information Sciences Research Center, Ohio State University.

Tatham, M. A. A.,1979. Some problems in phonetic theory. in Current issues in phonetic sciences. Harry and Hollien eds. Amsterdam-John Benjamins B.V.

Todo, A. 1958. The phonemes of Peiping Dialect. Project on Linguistic Analysis 4:1-18.

Trager, G. L. and H. L. Smith, 1951. An Outline of English Structure. Norman, Okla.

Tuller, B., and Kelso, S., 1984. The timing of articulatory gestures: Evidence for relational invariants. JASA 76.4:1030-6.

Wanner, D. 1975. A note on diphthongization. Studies in the Linguistic Sciences 5.2.

Warden, M. 1979. The phonetic realization of diphthongs in Toronto English. Toronto English: Studies in Phonetics. Studia Phonetica.

Warmuth, D., Mundie, J.and Vaughn, G., 1977., Speech synthesis by a programmable digital filter. IEEE 77:591-4.

Weeda, D. 1983. Perceptual and articulatory constraints on diphthongs in universal grammar. Taxas Linguistic Forum 22.

Wood, Sidney., 1982. The acoustical consequences of tongue, lip and larynx articulation in round palatal vowels. Working Papers, Lund Univ. 23:77-117.

Vaitkeviciute, V. 1983. On the duration of acute and circumflex diphthongs in Lithuanian. Abstracts of Tenth International Congress of Phonetic Sciences, Dordrecht: Foris Publications.

Xu, Shirong., 1980. Phonetics of Putonghua. (in Chinese). China.

Xu, M. 1963. On tonemes in the Beijing Phonology (in Chinese). Zhongguo Yuwen, 57.6.

Yang, S. and J. Cao, 1982. Dynamic properties of the diphthongs in Standard Chinese (in Chinese). Annual Reports on Experimental Phonetics, Institute of Linguistics, Chinese Academy of Social Sciences.

You, Rujie, Qian Nairong and Gao, Zhenxia. 1980. On the phonemic system of Putonghua. (in Chinese). Zhongguo Yuwen, 1980:5.

Zhou, Tongchun., 1982. The phonemic system of the Beijing dialect. (in Chinese). Beijing Normal University.

APPENDIX: CHINESE CHARACTERS OF THE TEST MATERALS

CHAPTER 3:	/ei/	/uei/	/tuei/	/phi#au/	
	欸	威	堆	皮襖	
CHAPTER 4:	/ai/	/uai/	/au/	/iau/	
	哀	歪	熬	腰	
	/ou/	/iou/			
	歐	优			
CHAPTER 5:	/ia/	/ua/	/uo/	/ie/	/ye/
	鴨	蛙	窩	耶	約
CHAPTER 6:	/ni#a/	/khu#a/	/a#i/	/ta#u/	
	你啊	苦啊	阿姨	大屋	