OSDL Ohio State
Dissertations in Linguistics

Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi

Shu-hui Peng

1996

For ordering information contact **OSDL** (Ohio State Dissertations in Linguistics):

OSDL
Department of Linguistics
222 Oxley Hall
1712 Neil Avenue
Ohio State University
Columbus, OH 43210-1298
U.S.A.

osdl@ling.ohio-state.edu

For dissertation abstracts and a listing of available titles visit our WWW server:

http://www.ling.ohio-state.edu/dissertations/

PHONETIC IMPLEMENTATION AND PERCEPTION OF PLACE COARTICULATION AND TONE SANDHI

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Graduate School of the Ohio State University

By

Shu-hui Peng, B.A., M.A.

* * * * *

The Ohio State University

1996

Dissertation Committee:

Mary Beckman

Keith Johnson

Robert Levine

Approved by

Mary Beckman

adviser
Department of Linguistics

To My Parents

ACKNOWLEDGMENTS

I express my sincere gratitude to my adviser Mary Beckman for introducing me to phonetics, for her guidance and encouragement throughout my entire graduate education, and for always being supportive and enthusiastic about my research. I have benefited very much from her teaching and her research. I am also grateful to the other members of my advisory committee, Keith Johnson and Robert Levine. Many thanks to Keith Johnson for generous technical support, valuable suggestions, and concern about my future career. I thank Robert Levine for his generous help and comments from a different perspective. In addition, I am grateful to Marjorie Chan for many interesting and inspiring discussions about Chinese languages. For comments and education, I would also like to thank Julie Boland, Robert Fox, Brian Joseph, Joyce McDonough, David Odden, John Ohala, and Mark Pitt. Gratitude is expressed to Sam Wang for help in recruiting subjects and in providing technical support for the perception experiments. The help of my friends who participated in the EPG experiment is also gratefully acknowledged.

I also owe thanks to my friends in the phonetics Laboratory at the Ohio State University: Gayle Ayers, Keven Cohen, Mariapaola D'imperio, Jamie Green, Stefanie Jannedy, Sun-Ah Jun, Sook-hyang Lee, Matt Makashay, Ho-Hsien Pan, Elizabeth Strand, Junko Tsumaki, Jennifer Venditti, and Kiyoko Yoneyama. I am especially grateful to Gayle Ayers and Jennifer Venditti for technical support. Thanks go to Mariapaola D'Imperio and Stefanie Jannedy for their suggestions about Italian and German.

I would like to thank my parents for their gift of education to me and for their support and encouragement all these years. To my husband, Chen-Jen, I offer sincere thanks for always being good company and for endless support and encouragement.

This research was supported by a Graduate Student Alumnae Research Award granted by the Graduate School of the Ohio State University, by an NIH grant #R29 DC01645-03 to Keith Johnson, and by the Department of Linguistics at the Ohio State University. I am very grateful to Department of Linguistics at the Ohio State University for providing me financial assistance throughout the past six years.

VITA

November 5, 1966	Born - Chia-Yi, Taiwan Republic of China
1989	B.A., Foreign Languages and Literature National Sun Yat-sen University, Kaohsiung, Taiwan
1990-1995	Graduate Assistant, Department of Linguistics, The Ohio State University, Columbus, Ohio
1991	M.A., Linguistics The Ohio State University, Columbus, Ohio
1994	Graduate Student Alumnae Research Award, The Ohio State University

PUBLICATIONS

Peng, Shu-hui (1995). Effects of prosodic position and tonal context on Taiwanese tones. *Proceedings of the Sixth North American Conference on Chinese Linguistics*.

Peng, Shu-hui (1995). Perceptual evidence of tonal coarticulation. *OSU Working Papers in Linguistics* 45, 160-165.

Peng, Shu-hui (1994). Effects of prosodic position and tonal context on Taiwanese tones. *OSU Working Papers in Linguistics* 44: 166-190.

Peng, Shu-hui (1993). Cross-language influence on the production of Mandarin /f/ and /x/ and Taiwanese /h/ by native speakers of Taiwanese Amoy. *Phonetica* 50: 245-260.

Peng, Shu-hui (under review). Production and perception of Taiwanese tones in different tonal and prosodic contexts. Submitted to *Journal of Phonetics*.

FIELDS OF STUDY

Major Field: Linguistics

Studies in Phonetics, Phonology, Phonological and Phonetic Interference, Psycholinguistics, and Sound Structure of Mandarin and Taiwanese

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

PHONETIC IMPLEMENTATION AND PERCEPTION OF PLACE COARTICULATION AND TONE SANDHI

BY

Shu-hui Peng, Ph.D.

The Ohio State University, 1996

Professor Mary Beckman, Adviser

This dissertation examines two cases of contextual variability for evidence concerning the nature of phonological and phonetic representations and the relationship between them. The first case is Mandarin tone sandhi whereby tone 3 becomes tone 2 when followed by another tone 3. The second case is place "assimilation" between Taiwanese coda and onset stops

In the tone sandhi study, the overall F0 of the derived tone 2 was slightly lower than that of the underlying tone 2. A forced-choice identification test between /tone 2 + tone 3/ and /tone 3 + tone 3/ showed that the derived tone 2 and the underlying tone 2 were perceptually indistinguishable to Mandarin speakers. On the other hand, concept formation test showed that most Mandarin speakers nonetheless categorized the derived tone 2 as tone 3, although they could be trained to categorize the derived tone 2 as tone 2 (with less consistent performance).

In the Taiwanese place "assimilation" study, EPG measurements showed a noncategorical gestural coarticulation. However, the latency of the second gesture with respect to the first decreased as speech rate increased, so that gestural "deletion" eventually occurred at faster speech rates. The dental gesture was "deleted" more frequently than the velar gesture. A concept formation test showed that coda /t/ and

coda /k/ had different patterns of categorization. As speech rate increased, the dental-velar sequence was categorized as /t/ less frequently, and as /k/ more frequently. By contrast, the velar-dental sequence was categorized as /k/ to about the same extent as for the velar-velar sequence. This difference in categorization may be attributed to different patterns of gestural coarticulation in the dental-velar sequence and the velar-dental sequence.

These results pose difficulties for all current models of phonological and phonetic representations. Models which make a hard and fast separation between (discrete categorical) phonological representation and (continuous) phonetic representation cannot explain the gradualness of the progression from "coarticulation" to "deletion". Models which collapse these two representations into one articulatory representation cannot explain the categorization of the sandhi tone.

CHAPTER I

INTRODUCTION

A complete description of the sound system of a language should include two parts, one phonological and one phonetic. The phonological part describes the units involved in the phonemic principles of contrast among words in the lexicon, and phonological and morphological alternations affecting the words when they connect together in the utterances of the language. The phonetic part shows the details of how the sound patterns are produced by the speakers and perceived by listeners. This dissertation examines two cases of contextual variability for the light that they can shed upon the nature of phonological representation and phonetic representation and upon the relationship between them. This is an extremely controversial issue in current phonetic theory, and there are many points of views. Some researchers believe that phonological representation and phonetic representation are distinct, whereas others believe they are not. Those who agree on the distinction between phonological and phonetic representations differ in their opinion about the relationship between the two types of representation. Some of them assume that the relationship between them is comparable to different derivational levels within generative phonology, whereas other argue that it is more like the relationship between syntax and semantics, and need not be derivational.

1.1 Models of the Phonetics-Phonology Relationship

To see how the points of view differ, consider how each would handle the influence of the following vowel on the place of articulation of velar stop consonants. The influence can take any of several forms, illustrated in Figure 1 (Öhman, 1966) and (1). In Swedish and English, [k] and [g] anticipate the tongue position of the following vowel, so the tongue makes a more forward constriction in the sequences [ki] and [gi] than in the sequences [ku] and [gu] (Ladefoged, 1982: 58). However, [k] or [g] in a similar phonetic environment shows even more extreme variation in some other cases. Velar softening, which is also called palatalization, refers to cases where velar stops [k] or [g] become palatal fricatives or affricates such as [ʃ, ʒ, t͡ʃ, d͡ʒ] before palatal vowels or glides such as [i, ɪ, e, y, j]. Examples are shown below in (1). Velar softening occurred as a the sound change in English and Mandarin. In Italian, the same sound change has left as a residue, morphophonological alternation in the singular vs. plural forms of some nouns. However, unlike Mandarin, the later introduction of [k] before [i] in English and Italian has made any "conditioned allophone" analysis impossible for these two languages. The velar stop phoneme in Acadian French shows variation all along the gradient involved in the sound change in Mandarin and English. When followed by front vowels, /k/ can be produced with the front constriction typical of English /k/, or with different degrees of more extreme palatalization including change to an alveo-palatal affricate (Hume, 1992).

(1) (a) Old English > Middle and Modern English

OE [kin] > ME [t͡ʃɪn] "chin"

(b) Shangtung Mandarin > Standard Mandarin

[feiki] > [feit͡ɕi] "airplane"

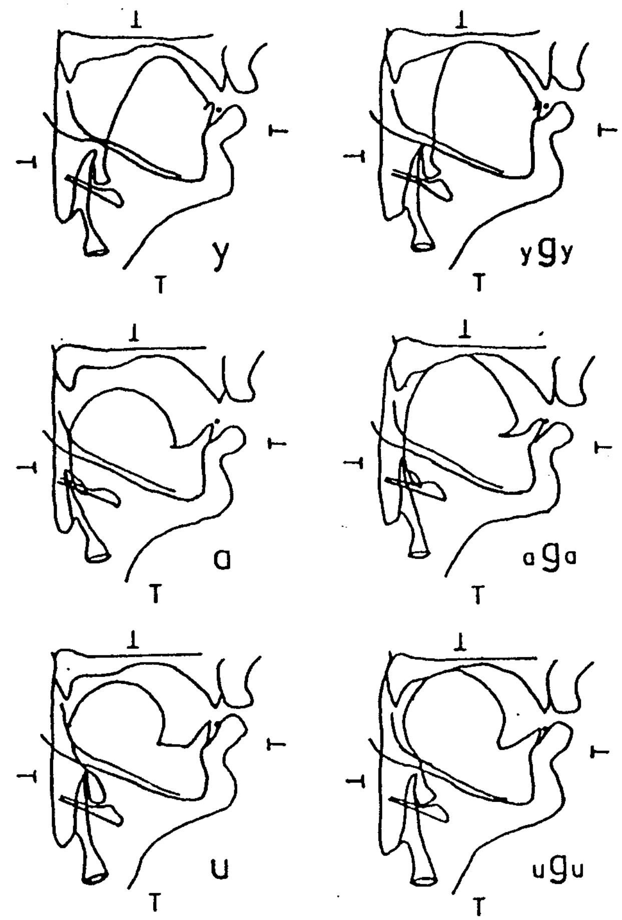


Figure 1. Contour tracings from x-ray motion pictures of [y, a, u] and [g] preceded and followed by these vowels (Öhman, 1966)

(c) Italian

[k]	amico	"friend"	[t͡ʃ]	amici	"friends"
	attico	"attic"		attici	"attics"
	chicchi	"crumbs for birds"			
	orecchio	"ear"			

(d) Acadian French

[ka]	'case'	[kɛ], [kʲɛ] or [t͡ʃɛ]	'quay'
[kut]	'cost'	[ki], [kʲi] or [t͡ʃi]	'who'
[cote]	'side'	[kø], [kʲø] or [t͡ʃø]	'tail'

1.1.1 Ladefoged

According to Ladefoged's (1980) view, phonological representation and phonetic representation are different in their functions and in the parameters used in their description. Phonetic representation employs articulatory parameters (which are related to motor control and articulator movements in production), and acoustic parameters (which are related to auditory effects in perception to describe phonetic details of speech sounds). On the other hand, phonological representation employs segments and categorical phonological features to describe sound patterns of language as a 'self-organizing social institution'. Each language has a small set of discrete phonemes which are characterized in terms of distinctive features which can be superficially related to either acoustic or articulatory features. However, Ladefoged (1980) maintained that the phonological features useful to phonological descriptions need not to correspond to the mental representations used by speakers and listeners while speaking and listening to a language. That is, phonology describes systems of contrasts, but not necessarily the behavior of any individual. Phonological rules are different from the rules acquired by speakers for speech production and speech perception. When speakers produce and perceive speech, they employ phonetic parameters, not phonological features.

In this and later work, such as Ladefoged & Wu (1984) and Ladefoged (1990), Ladefoged emphasizes that it is difficult to relate phonological descriptions to phonetic analysis. The difficulty is that phonology and phonetics have different aims. The aim of phonology is to describe the patterns of the contrasting units in a language. The aim of phonetics is to describe all sounds that occur in language and interpret them in terms of acoustics or physiology. The relation between phonological description and phonetic data is even more difficult to establish due to the fact that the average number of phonemes is much smaller than the number of auditorily distinct sounds that the human vocal apparatus can generate. Thus, phonological features must be distinct from phonetic parameters. They can not be in a one-to-one relationship. One or two phonological features can be interpreted in terms of a single articulatory parameter, a small number of others can be interpreted in terms of a single primary acoustic parameter. However, most phonological features correspond to articulatory parameters and acoustic parameters in a many-to-many fashion. For instance, the feature [coronal] has to be defined by at least two articulatory parameters: tongue tip raising and tongue tip fronting. The feature [strident] may correspond to the frequency and amplitude distribution of the turbulence noise that arises when air passes through a narrow oral channel and strikes the incisors.

According to Ladefoged's (1980) model, then, the patterns in Figure 1 and Table 1 will not have a uniform representation. The variation in the place of articulation for the velar stops in different vowel contexts that occurs in Swedish and English requires a phonetic representation to include such fine subphonemic articulatory detail. Because the place feature of velar stops do not change into another distinctive feature, the variation cannot be described by a phonological rule. A better way to represent this variation is using Öhman's (1966) articulatory speech synthesis model. Velar softening, by contrast, can be described by a phonological rule indicating that the place and manner features of the consonant have changed from a velar stop to a palatal affricate. In English and Italian, /k/ and /t͡ʃ/ are distinctive phonemes, because [k] also appears before [i]. Phonetic representation has nothing to say about the alternation, except that the stop will follow coarticulation rules for velar stops and there will be a different phonetic representation for the affricate. This means that cases such as Acadian French alternations will be handled only with extra stipulation since they are not typical categorical changes. Because one of the functions of phonetic representation is to describe the utterances of individuals, it can interpret the free variation in Acadian French velar stops as different degrees of coarticulation used by speakers.

1.1.2 Pierrehumbert (1990)

Pierrehumbert's (1990) shares Ladefoged's (1980) view that phonological and phonetic representations are distinct and have different functions. Phonological representation describes categorical contrasts in language in terms of discrete features and syntactic rules. Phonetic representation describes physical events of speech, that is, the quantitative properties of articulation, acoustics, and audition. In general, there are two main domains of phonetic representations: acoustic and articulatory. In each of these domains, there are some subdomains. For instance, in the acoustic signal, a phonetic representation can be completed in the time-domain of the wave form (amplitude over time) or in the spectral domain (amplitude over frequency).

In contrast with Ladefoged's (1984) view that phonological representation is not likely to be a mental representation, Pierrehumbert (1990) argues that phonological representation is concerned with the implicit knowledge stored in the mind of speakers and listeners. For example, for language to be a tool of communication, there must be a

cognitive status in the association between the phonological forms of words and their meanings.

Pierrehumbert (1990) also emphasized that phonetic interpretation of phonological categories is context-dependent, and the principles that explain the relationship between phonological representation and phonetic representation are semantic rather than syntactic. That is, the symbolic representation employed in phonology is interpreted by the phonetic events and phenomena in the real world, namely articulation and acoustics. The relation between phonological representation and phonetic representation is not established by rewrite rules in derivational chains.

The strict separation between phonetic and phonological representations makes it difficult for Pierrehumbert's (1990) model to deal with cases such as Acadian French velar softening. The alternations of the velar stop cannot be described only by a discrete feature change in a syntactic rule. They cannot be described as just a coarticulation process between the velar stop and the following vowel, either.

1.1.3 Pierrehumbert (1994)

However, in her recent work, Pierrehumbert (1994) argues that some phonological alternations have quantitative variation both in occurrence and in phonetic implementation, so that they cannot be fully described by discrete syntactic rules feeding into gradient phonetic rules. For instance, in English, syllable-final voiceless stop consonants may be glottalized, or even be reduced to a glottal stop before certain consonants, e.g. /t/ in "artwork". However, the glottalization only occurs some of the time. When it does occur, the probability of glottalization varies depending on segmental context. This irregularity makes it impossible to fully describe glottalization in English using only one phonological rule, since a single rule must either omit some cases or over-generalize outputs. On the other hand, if separate rules are to be used for each segmental context to include all possible glottalization environments, each rule will also need to be specified with its own probability distribution and hence must miss any generalizations about analogous prosodic context and rate effects.

Pierrehumbert (1994) also argued that some phonological alternations which are considered to be obligatory categorical changes by traditional phonological theories can be explained better by a gradient model which allows gradient application of the alternations. For example, in many languages, phonotactics disfavor the concurrence of homorganic consonants in the same morpheme. This phenomenon is described as the Obligatory Contour Principle for place of articulation (OCP-Place) and is treated as categorical change by most phonological analyses (e.g. McCarthy, 1986). However, Pierrehumbert (1994) argued that the application of the OCP-Place varies according to the distance and similarity between the consonants, so a gradient "similarity" model is more appropriate than a categorical presentation of absolute adjacency to describe the OCP-Place effect.

1.1.4 Keating

Keating's (1990) model of phonological and phonetic representations described the structure of language using a generative grammar. Unlike most of the models which have two types of representations: one phonological and one phonetic, Keating's model includes one type of phonological representation and two types of phonetic representations: one categorical and one quantitative. Phonological representation describes distinctive categories which are defined in a relative manner at the level of overall contrasts within a language. Categorical phonetic representation describes

'natural phonetic categories' in any specific context. Quantitative phonetic representation describes parametric details of utterances in languages.

In Keating's (1984) model, natural classes used in phonological rules are organized in terms of binary phonological features. These binary phonological features are implemented as phonetic categories selected from a universally specified set. For example, for voicing there are three major phonetic categories in the fixed category set: (1) voiced, (2) voiceless unaspirated, (3) voiceless aspirated. Languages choose different categories from this set. Physical details of phonetic categories implemented in specific contexts in a language are described using continuous articulatory and acoustic parameters.

Phonological representation and the two types of phonetic representation are derivationally related in a model of speech production and perception in Keating's (1984; 1990) model. Phonological representation generates categorical representation. Categorical phonetic representation then serves as the input to articulatory planning of speakers and the output of speech processing of listeners. A speaker will use a categorical phonetic representation which is derived from a mental representation—phonology—to arrive at an articulatory phonetic representation. A listener will use an acoustic phonetic representation to arrive at a categorical representation which then will be used to derive a phonological representation.

Parametric phonetic representations are derived from the categorical phonetic representation using a model—" the targets and interpolation model". Phonological features used in categorical representation are mapped to physical parameters in a many-to-many manner. That is, feature values are mapped to targets (articulatory or acoustic). The mapping is specified in terms of universal phonetic categories. However, the phonetic details are also language-specific to some extent. A target does not necessarily refer to an absolute quantitative value, but rather a range of quantitative values. They are connected to each other by interpolation functions which are usually language-specific and/or parameter-specific. A single feature may be related to more than one parameter. For instance, speakers can implement the feature Voice by manipulating articulatory parameters such as expansion of oral cavity, lowering of larynx, relaxing of the vocal tract walls, respiratory effort and lowering of the velum. Listeners, on the other hand, can derive a feature value such as [+ spread glottis] from perceiving values of acoustic parameters, e.g. closure duration, voice onset time, or aspiration amplitude. Different languages may have different realizations of the same feature. For example, Disner's (1978) study suggests to Keating that a high front unrounded vowel is 'interpreted' differently (with different target values for formant frequencies) in different languages.

The derivational relationship between the different types of representations in Keating's model (1990) makes it difficult for the model to interpret the alternations in Acadian French. If a phonological feature will generate some phonetic category selected from a universal set, then there should be a phonetic category chosen for the feature [velar]. However, there seem to be more than one category that can be selected by Acadian French speakers for the feature [velar]. Even if we assume that there is only one phonetic category which has been selected, there is more than one qualitatively distinct set of parametric phonetic outputs which can be derived from the categorical representation.

1.1.5 Browman and Goldstein

Articulatory representation is phonological representation as well as phonetic representation according to the theory proposed by Browman and Goldstein (1989;

1990a; 1990b; 1992). Their model is designed to handle sequence effects such as assimilation based on a computer simulation of speech production. Abstract control parameters which define 'gestures' are interpreted by a system of dynamic equations into observable articulator movements which then generate the acoustic output of the computational system. Articulatory gestures are defined in terms of speech tasks which include the formation and release of articulatory constrictions and the underlying dynamics that characterize the motions of articulators. In this view, then, articulatory gestures are the basic atoms of phonological structure. There are gestures for tongue tip, tongue body, lips and so on, and each gesture is defined by dynamic parameters referring to three or four dimensions: constriction degree, constriction location, stiffness and constriction shape (which is only used for oral gestures). Gestures are organized into coordinate structures in speech utterances.

Inter-articulator coordination (for instance, lower lip and jaw coordination in producing bilabial consonants) is defined in the gestural specifications and in general definitions of 'tract variables' like lip aperture. Inter-gestural coordination (for instance, lip constriction and velar constriction in producing /u/) is defined in gestural scores which include two dimensions: articulatory tiers and time. The timing of gestures plays an important role in the gestural score. Different patterns of overlap for successive articulatory gestures can produce different types of phonetic and phonological variation. One gesture might be completely overlapped by neighboring gestures and therefore hidden acoustically. Discrete activation intervals of gestures are marked to show the durations of the gestures in the gestural score. There is a gestural descriptor which is the pointer to the articulator tier and to the dynamic parameters within the gesture's activation interval. Gestural descriptors can also serve as distinctive features employed in phonology, such as [closed labial], [narrow pharyngeal]. Gestures using distinct tract variables such as lip protrusion and lip aperture are categorically distinct within a sound system inherently. Gestures using the same tract variable can also be categorically distinct by virtue of the different values of the dynamic parameters such as different degrees of constrictions. There are some stable ranges with the parameters. These ranges probably can be explained in terms of the non-linear relationship between the parameter values and acoustic properties proposed in Stevens's quantal theory (1972; 1989). Browman and Goldstein (1989) and Goldstein (1989) indicated that the quantal articulatory-auditory relationship can be used to partition continua of gestural organization, but not to constrain talking-precision in every act of speech production.

Browman and Goldstein's model (1989; 1990) can describe the anticipatory coarticulation between velar stops and the following vowels as gestural coarticulation. According to their model, neighboring segments may preserve their own places of articulation, while simultaneously showing overlapping of articulatory gestures. Gestural hiding occurs when one gesture is hidden behind another gesture. Gestural overlap may also lead to blending of two neighboring gestures which may cause undershooting articulatory targets.

Previous research indicated that segmental coarticulation is a gradual process of gestural overlapping. The extent of gestural overlapping is affected by speech rate. As speech rate increased, more gestural overlapping was found (Munhall & Löfqvist, 1992; Nolan, 1992; Byrd, 1994). Munhall and Löfqvist (1992) found different patterns of laryngeal gestural overlapping between an alveolar fricative and an alveolar stop. Two separate glottal openings were found at slow speech rates. The underlying glottal opening gestures for the fricative and the stop were blended into one single movement at fast rates. At intermediate rates, the gestures for the fricative and the stop were partially overlapped.

However, Browman and Goldstein's model cannot explain velar softening very well. The change from dorsal to coronal gesture cannot be easily described as the results of gestural overlapping or undershooting of articulatory targets. The merging of phonological representation and phonetic representation into one articulatory representation makes it more difficult for this model to describe the phonetic variations of the velar stop occurring in Acadian French. Multiple gesture representations will be needed to describe all the variations. Since articulatory representation is also phonological representation, there will also be multiple phonological representations for the velar stop. This will predict multiple distinctive categories while only one distinctive category is needed for the phonemic system of the language.

In summary, neither the two models which propose separate representations for phonological alternations and phonetic variations (Ladefoged, 1980; Keating, 1990; Pierrehumbert, 1990) nor the model which collapses these representations into one articulatory representation (Browman & Goldstein, 1989; 1990a; 1990b; 1992) can fully describe the various kinds of vowel effects on the place of articulation of velar stop consonants. Models which distinguish phonological representation from phonetic representation can handle cases such as anticipatory coarticulation and simple velar softening, but it has difficulty in dealing with mixed cases such as the free variation of the velar stop in Acadian French. Articulatory phonology can provide detailed articulatory descriptions for anticipatory coarticulation using gestural overlapping and blending, but it has problems to describe velar softening and more complicate cases such as Acadian French stop variations.

1.2 Two Case Studies: Tone Sandhi and Place Assimilation

This thesis explores these questions about phonological representation and phonetic representation in two case studies. The first case study explored a Mandarin tone sandhi rule in which the low-dipping tone (i.e. "tone 3" or [214]) becomes a high-rising tone (i.e. "tone 2" or [35]) when followed by another low-dipping tone in a two-syllable word or phrase. For example, /y^{214} san^{214}/ —> [y^{35} san^{214}] "umbrella". (Numerical values represent the pitch height on a five-point scale created by Chao, 1948, 1968.) The second case study explored Taiwanese stop place "assimilation". In Taiwanese, the first of two adjacent stops seems to assimilate to the second stop in place of articulation, for example /set$^{\underline{21}}$ ke^{21}/ —> [sek$^{\underline{53}}$ ke^{21}] "to design". The tone sandhi has been described as a discrete phonological alternation in which one tone simply changes to another tone in a specific environment. By contrast, the place assimilation may be a gradual process in which articulatory gestures for the two stops overlap to different extents, creating an illusion of categorical change in cases of extreme overlap.

The patterns of place assimilation are examined using Electropalatography (EPG). The changed tone in tone sandhi is compared to the underlying tone 2 using measurements of fundamental frequency. The perception of segmental coarticulation in place assimilation and the changed tone in tone sandhi are examined using two different methods: the traditional forced-choice identification and the concept formation method. Concept formation is an experimental paradigm which can reveal perceptual categorization which results from high-level speech processing. The comparison of these two perceptual testing methods may show different patterns of categorization and is more informative than either method alone.

Previous work on production of stop consonant clusters across morpheme boundaries supports a gradient degeneration of the place cues of the first consonants. Nolan (1992) showed that place assimilation in English alveolar-velar stop sequences separated by word boundaries as in "late calls" forms a continuum of gestural overlap between the two consonants. The articulation of the alveolar stop can be roughly classified into four patterns for four different extents of gestural overlap: full alveolar, residual alveolar, zero alveolar, and nonalveolar. Full alveolar refers to tokens with a complete closure at the alveolar ridge. Residual alveolar includes tokens with contact along the sides of the palate which is further forward than the nonalveolar, but without a complete closure at the alveolar ridge. Zero alveolar tokens show contact which is not further forward than the nonalveolar token.

If place assimilation is gradual as shown in previous studies, how is the articulatory continuum perceived by listeners? Nolan's (1992) study of English stop consonants revealed that the identification of a word-final stop consonant by listeners varied depending on the extent of the gestural overlapping between the consonant and the following initial stop consonant. The identification accuracy ranged from highly reliable when the gestural cue of the stop was shown clearly, to completely non-identifiable when the gestural cue of the stop consonant completely disappeared due to gestural coarticulation or assimilation.

Tone sandhi is similar to alternation of articulatory gestures in that tonal features are changed by surrounding tonal environments. However, Mandarin third tone sandhi has been described only as a discrete change. The third tone changes to tone 2 when followed by another tone (Chao, 1948). Tone sandhi and gestural coarticulation are also different in the possible cause for the alternation. Most segmental interactions can be attributed to assimilations stemming from coarticulation during speech production, whereas tonal alternations are more difficult to account for in this way and sometimes even appear to be dissimilatory in nature. It is not very clear how articulatory mechanisms such as cricothyroid contraction for vocal fold elongation are manipulated to produce the coarticulation-like assimilation of tones. Given these differences, the perceptual categorization of the sandhi tone may be different from that of gestural coarticulation.

The most well-known Mandarin tone sandhi rule is: tone 3 becomes tone 2 when followed by another tone 3. The four phonologically distinctive tones in Mandarin can be represented using two tone levels: tone 1 /H/, tone 2 /LH/, tone 3 /L/, and tone 4 /HL/. The tone sandhi can be described with the following rule. L —> LH/__ L. Zee (1980) found that tone 2 derived from tone 3 through the tone sandhi rule on average has a lower F0 offset than an underlying tone 2. It was also found that a phrase with tone 2 followed by tone 3 is perceptually identical with a phrase with underlying tone 3 followed by tone 3 (Wang & Li, 1967). For example, /fən^{214} tʂ^{h}aŋ^{214}/ "flour factory" was not distinguished from /fən^{35} tʂ^{h}aŋ^{214}/ "grave yard" in Wang and Li's (1967) identification test. This finding indicates that these two combinations of tones are phonetically very similar, but it does not show how the sandhi tone is represented phonologically or how Mandarin speakers categorize it. Based on these studies, it is still not clear whether the tone sandhi rule creates a complete neutralization between the underlying tone 2 and the derived tone 2.

Chapter II investigates Mandarin third tone tone sandhi in a production experiment and two perception experiments. Chapter III investigates place assimilation in Taiwanese stop consonants with an EPG experiment and a perception experiment. Chapter IV provides some general discussion and conclusion about how well the models describe the two case studies.

CHAPTER II

CASE STUDY ONE: TONE SANDHI

This chapter investigates the Mandarin third tone tone sandhi using a production experiment and two perception experiments. The production study compared the underlying tone and the sandhi tone in acoustic measurements. An identification test was used to show whether the underlying tone 2 and the tone 2 derived by tone sandhi are perceptually neutralized. Native Mandarin listeners were tested to see whether they can distinguish 'tone 2 + tone 3' sequences from 'tone 3 + tone 3' sequences. Another perception test using the concept formation method (which can access the phonological knowledge of listeners) was designed to show how the derived tone 2 is categorized by native Mandarin speakers—that is, which phonological category is assigned to the derived tone 2?

2.1 Experiment One

This experiment compared F0 contours of two tone sequences: /35 + 214/ and /214 + 214/ produced in different speech rates. The comparison shows the similarity and difference between the first syllables in terms of fundamental frequency (F0) and duration.

2.1.1 Methods

2.1.1.1 Experimental Materials

Twelve pairs of Mandarin two-syllable phrases were used. The two phrases in each pair had identical segments, but one of the phrases has tone 2 followed by tone 3, whereas the other had tone 3 followed by tone 3. After the application of the tone sandhi rule, the two phrases in each pair were homophones.

Twenty-four phrases with high familiarity were used. Subjective familiarity of ratings for 88 pairs of Mandarin phrases with these two kinds of tone combinations were elicited from a survey of 93 Mandarin speakers who were college students in Taiwan. They rated a written list of the 176 phrases in random order on a seven-point scale ranging from 1 for phrases which were unknown to 7 for phrases which were most familiar (i.e. they knew the meaning very well). Twelve pairs of phrases with the highest and approximately equal familiarity rating were chosen.

2.1.1.2 Subjects

Ten native Taiwanese Mandarin speakers, five male and five female, served as speakers in this experiment. They were all born and grew up in Taiwan, but none of them spoke Taiwanese. They spoke English as a second language. None of them had hearing problems by self-report.

2.1.1.3 Recording and Measurements

The recordings were made in the sound booth at the Linguistics Department of the Ohio State University. The 24 phrases were read three times in three blocks. In each block, speakers were asked to read each phrase starting with a slow speech rate and then increase the speech rate gradually to their fastest possible rate. The recordings were

digitized (16 kHz sampling, 16 bits/sample) and F0 was extracted at 10 ms intervals using the dynamic programming auto-correlation-based pitch tracking algorithm of the Entropic *Waves+* speech analysis package. Nine of each speaker's repetitions of each of the 24 phrase were selected to be measured for duration and F0: three tokens from each of the slowest, mid, and fastest speech rates. Duration and time-normalized F0 trajectories of each target syllable (the first syllable of each phrase) were extracted from the raw f0 trajectories. In these time-normalized trajectories, F0 values of each target syllable were measured by linear interpolation at ten equally spaced points of the raw F0 contour. When the F0 contour of the token did not show a clear separation between the first syllable and the second syllable due to the continuation of voicing through the initial segment of the second syllable, the token was also displayed on a wide-band spectrogram to determine the offset of the first syllable.

2.1.2 Results

Figure 2 and Figure 3 show mean F0 values of tone 2 and the sandhi tone at 10 measurement points averaged over speech rates for each speaker. A three-way repeated-measures ANOVA (tone x speech rate x measurement point) showed that tone 2 and the sandhi tone were not very different from each other. Figure 4 shows mean F0 values of tone 2 and the sandhi tone at 10 measurement points averaged over speakers and speech rates. The effect of tone type was marginally significant, [$F(1,9)=6.01$, $p < 0.05$]. The overall F0 value of the sandhi tone (178.33 Hz) was slightly lower than that of a underlying tone 2 (180.66 Hz). The effect of the measurement point was significant, [$F(9, 81)=28.25$, $p < 0.001$]. The general tone shape of both tones--the rising contour was preserved. The interaction of tone and measurement point was not significant,

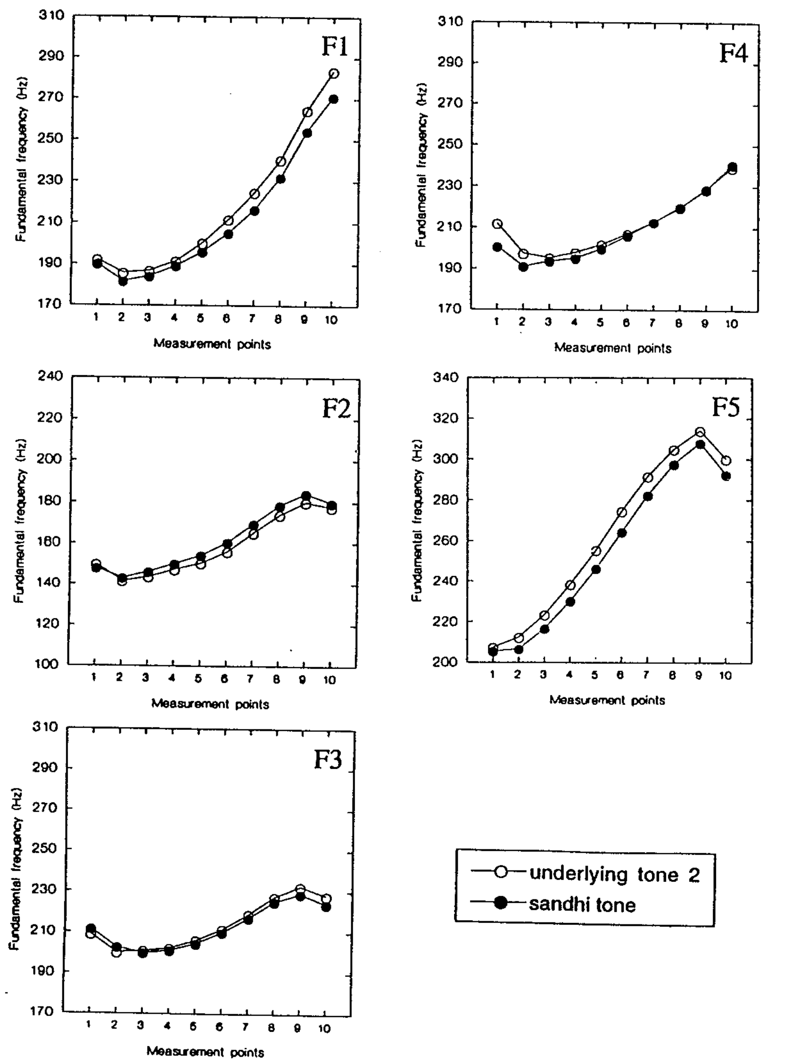


Figure 2. Mean F0 values of tone 2 and the sandhi tone at 10 measurement points for each female speaker

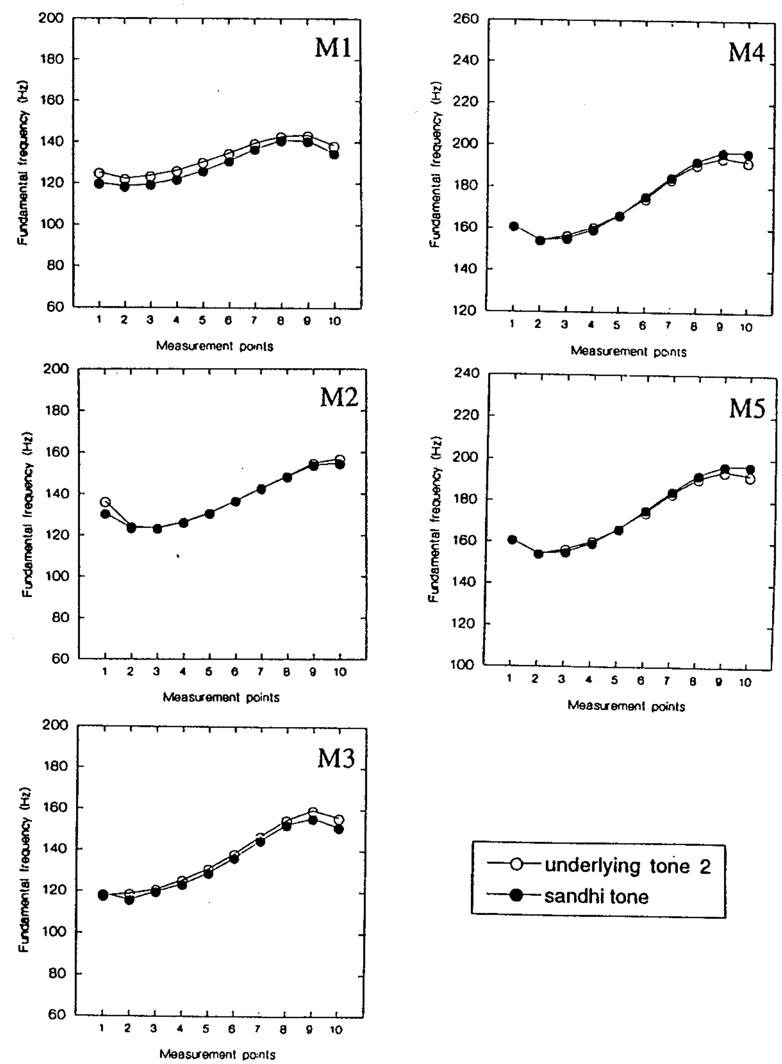


Figure 3. Mean F0 values of tone 2 and the sandhi tone at 10 measurement points for each male speaker

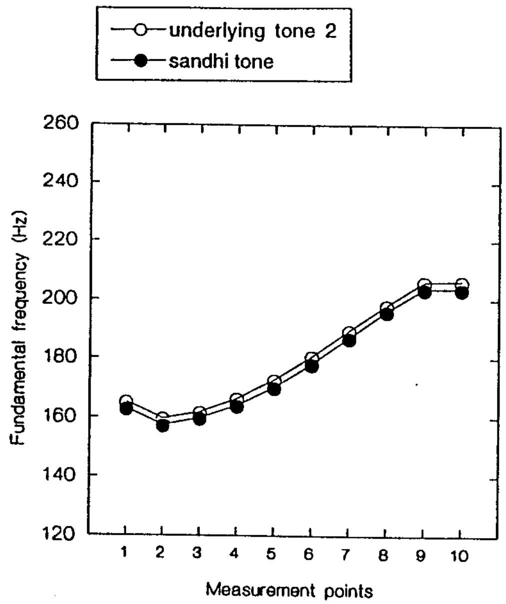


Figure 4. Mean F0 values of tone 2 and the sandhi tone at 10 measurement points averaged over speakers and speech rates

[$F(9, 81)=0.07$, $p = 0.25$]. The tone shape of the sandhi tone was the same as that of the underlying tone 2.

The pitch range of both tones varied according to speech rate (Figures 5 & 6). The interaction of speech rate and measurement point was significant, [$F(18, 162)=17.11$, $p < 0.001$]. The extent of the pitch rise from the onset to the offset of the tone decreased as speech rate increased. The pitch contour of tokens spoken in faster speech rate ended at a lower F0 value than those of tokens spoken in slower speech rate. The initial part of the pitch contour in fast speech rate also had a higher F0 than those in normal and slow speech rates.

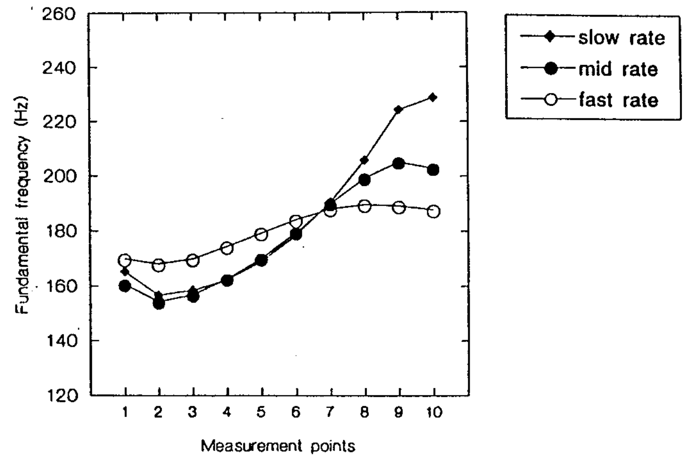


Figure 5. Mean F0 values of tone 2 at 10 measurement points averaged over speakers

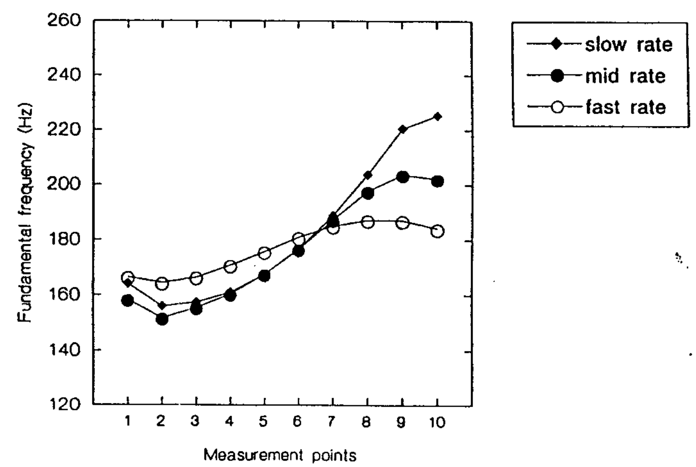


Figure 6. Mean F0 values of the sandhi tone at 10 measurement points averaged over speakers

2.1.3 Discussion

The sandhi tone is not completely neutralized with tone 2 acoustically. The pitch height of the sandhi tone was not only lower than tone 2 in the same tonal environment at the offset of the pitch contour (as found in Zee's 1980 study), but also lower in overall pitch height. This difference in pitch height may be because the low onset of the sandhi tone (original 214) was maintained in the surface sandhi form. The sandhi form began and ended at lower pitch values while the rising slope stayed the same.

Tonal reduction occurred in phrases read at the fast speech rate. The pitch range of the pitch contour of both the sandhi tone and tone 2 varied as the function of speech rate. The rising slope of both the sandhi tone and tone 2 became shallower as speech rate increased. The short duration of the syllable seems to be incompatible with the realization of a large pitch rise. Therefore, the tone targets were undershot in tones produced in normal and fast speech rate.

2.2 Experiment Two

An identification test was designed to see whether underlying tone 2 is perceptually distinguishable from tone 2 derived by tone sandhi. This experiment tested whether the small acoustic difference between the underlying tone 2 and the sandhi tone found in Experiment 1 can be detected by native Mandarin speakers.

The identification test was also designed to show whether speech rate affected the differentiation between tone 2 and the sandhi tone. The increasing of speech rate makes the duration difference between the two tones smaller, so they may become neutralized in fast speech. Since tone 3 in citation form or phrase-final position is longer than tone 2 in duration (Ho, 1976), we might expect that tokens spoken in slow speech rate may be biased toward identification as tone 3 (regardless of whether they are underlyingly

tone 2 or derived from tone 3 by tone sandhi). Tokens spoken in faster speech rate, on the other hand, may be biased toward identification as tone 2. A perception study of Mandarin tone 2 and tone 3 showed that variation of syllable length affected the identification of F0 contours intermediate between tone 2 and tone 3 (Blicher, et al., 1990). The identification boundary was shifted toward tone 3 by longer syllable duration. Syllable-lengthening enhanced the F0 cues of tone 3.

A similar result was found in a German identification study (Jannedy, 1994) distinguishing the German minimal pair 'braten' "to fry" and 'beraten' "to advise". In slower speech rates, 'braten' was sometimes mis-identified as 'beraten' by German speakers—that is, slower rates introduced the precept of a reduced vowel breaking up the initial consonant cluster. In faster speech rates, 'beraten' (with the first vowel much reduced) was sometimes identified as 'braten'.

2.2.1 Methods

2.2.1.1 Experimental Materials

One male speaker (speaker M1) from Experiment One whose productions were most similar to the average production of the 10 speakers was chosen to be used in the perception test. The stimuli were 216 tokens (24 disyllabic phrases x 3 rates x 3 repetitions) which were divided into three blocks according to speech rate: slow, normal, and fast. Therefore, there were 72 tokens in each block. Tokens in each block were placed in random order and dubbed to an audio tape.

2.2.1.2 Subjects

Fifteen native Mandarin speakers (12 female, 3 male) participated in the identification test. Some of them also speak Taiwanese and/or English poorly. The listeners were all born and grew up in Taiwan and none had hearing problems by self report.

2.2.1.3 Procedure

Listeners were run individually in a quiet room. The stimuli were played to the listeners through headphones. They identified the phrases by circling answers on answer sheets in a forced-choice identification between /tone 2 + tone 3/ and /tone 3 + tone 3/. The two possible answers were disyllabic phrases written in Chinese characters. The subjects heard the block with tokens read in slow speech rate first, then tokens read in normal speech rate, and finally tokens read in fast speech rate.

2.2.2 Results

The production of speaker M1 was similar to the average production of the 10 speakers. A three-way ANOVA (tone x speech rate x measurement point) showed that the mean F0 value of the underlying tone 2 (132.60 Hz) was slightly higher than that of the sandhi tone (128.95 Hz) in the production of speaker M1, [$F(1, 2100) = 35.84$, $p < 0.01$]. The effect of measurement point was significant, [$F(9, 2100)=77.97$, $p < 0.01$). The general tone shape of both tones--the rising contour was maintained. In addition, the effect of speech rate was also significant, [$F(2, 2100) = 33.13$, $P < 0.01$]. The F0 values for the mid (129.17 Hz) and fast speech rates (128.86 Hz) were slightly lower than the F0 value for the slow speech rate (134.27 Hz). The pitch ranges of both tones varied according to speech rate. The interaction of speech rate and measurement point was significant, [$F(18, 2100)=10.04$, $p < 0.01$].

The measure A' which is a non-parametric measurement of perceptual sensitivity (Grier, 1971) was calculated for each listener's responses in different speech rates. A' (a non-parametric analog of d') is calculated by the formula shown in (2)

(2) $A'=0.5+(y-x)(1+y-x)/4y(1-x)$

where x is the ratio of false alarms (i.e. identification of tone 3 for underlying tone 2, and of tone 2 for underlying tone 3) and y is the ratio of correct responses (i.e. underlying tone 3 identified correctly as tone 3, and tone 2 as tone 2).

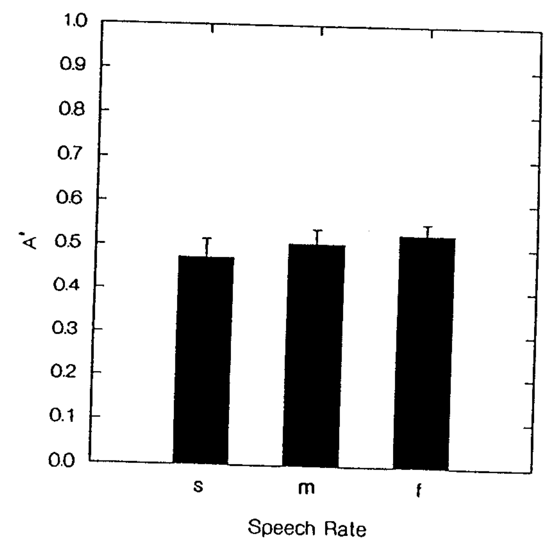


Figure 7. Mean A's with standard error for the identification of the sandhi and non-sandhi phrases for each speech rate block. s = slow speech rate, m = mid speech rate, f = fast speech rate

The values of A' range from zero to one. A value of one refers to perfect performance. For example, if a listener identifies the phrases with complete accuracy without any false alarm responses, then the A' value for that listener is one. An A' value of 0.50 occurs if responses were random.

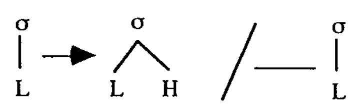
The identification of the sandhi phrases and the non-sandhi phrases was at about chance. The mean A' of the identification test was 0.50 with a standard deviation 0.13. Figure 7 shows the mean A' of identification for each speech rate block. A one-way ANOVA indicated that the effect of speech rate was not significant, [$F(2, 42)=0.71$, $p = 0.495$].

2.2.3 Discussion

Mandarin speakers cannot distinguish /tone 2 + tone 3/ phrases from /tone 3 + tone 3/ phrases regardless of speech rate. This finding replicated the findings of the earlier perception study of Mandarin tone sandhi by Wang and Li (1967). Mandarin speakers could not perceive the small pitch difference between the sandhi tone and the underlying tone 2 found in Experiment 1. The sandhi tone is perceptually indistinguishable from tone 2.

For a phonological representation following Keating's model, the tone sandhi can be described using a rewrite rule such as in (3).

(3)



The output derived from this phonological rule is the phonetic category LH (i.e. tone 2). This categorical phonetic representation will then generate a parametric phonetic representation. Since the sandhi tone and the underlying tone 2 are the same at the level of categorical phonetic representation, the parametric phonetic representation for both tones should be the same. As predicted by Keating's model, the sandhi tone and the underlying tone 2 were almost the same acoustically and therefore were perceptually indistinguishable. If the relationships between the two types of phonetic representations and the phonological representation are derivational as suggested by Keating's model, then for listeners, the same parametric phonetic representation will generate the same phonetic category (tone 2) for the sandhi tone and the underlying tone 2. The same phonetic category will then generate the same phonological representation. That is, Keating's model will predict that the sandhi tone and the underlying tone will be categorized as the same category by listeners.

2.3 Experiment Three

Experiment 2 showed that native speakers cannot distinguish tone 2 + tone 3 from tone 3 + tone 3 in an identification task. However, the question still remains whether the sandhi tone is categorized as tone 3 or tone 2. Experiment 3, using concept formation as an experimental paradigm, was designed to show how the sandhi tone is categorized by native speakers.

Concept formation is an experimental method that has been used by some researchers such as Jaeger (1986) for categorization. Jaeger (1986) used concept formation to investigate listeners' categorization of the English stop consonant [k] in the syllable-initial consonant cluster [sk] in words such as 'school'. The concept formation paradigm is composed of two listening sessions: a learning session and a test session.

In the learning session, listeners are trained to form or use a linguistic category. which includes stimuli in a target category chosen by the experimenter and stimuli in other categories. Those tokens in the target category are positive tokens, and the others are negative tokens. After each stimulus, feedback indicating whether or not the stimulus is a member of the target category is given to listeners. It is expected that the listener will form or make use of an existing category during the learning session. The learning session is followed immediately by a test session which includes positive tokens, negative tokens and test tokens which have ambiguous (to the experimenter) category identity. This paradigm provides an unbiased indication of the category membership of the ambiguous items, as well as a way to test the degree to which listeners can include the ambiguous stimuli in one category or another.

In addition to categorization of the sandhi tone, this experiment was also constructed to compare the difficulty in training Mandarin speakers to categorize the sandhi tone as tone 2 to as tone 3. Therefore, listeners participating in the perception test were divided into two groups in this experiment. One group was trained to learn tone 2 as the target category and the other group tone 3. The stimuli were two-syllable phrases. Phrases beginning with unambiguous tone 2 or tone 3 were positive tokens for the two groups respectively. Words beginning with any of the other three Mandarin tones were negative tokens. /tone 3 + tone 3/ phrases were test tokens. Unlike the more typical concept formation test such as Jaeger's (1986), in which feedback is given to subjects only in the learning session, feedback was also given to the listeners in the test session to train them to categorize the sandhi tone as tone 2 in one group and as tone 3 in the other group.

There is one more issue that needs to be considered in this experiment than in Jaeger's (1986) experiment for /k/ categorization. In Jaeger's (1986) experiment, the

test tokens—words beginning with the 'sk' sequence— only provide in a neutralized environment. The tested [k] never appears in anything but the [sk] environment. In these words, there is no evidence of phonological alternation. By contrast, the sandhi tone in surface form (tone 2) may appear in other non-sandhi environments. There is a clear alternation from one tone (tone 3) to another tone (tone 2). This may make it easier for listeners to detect the change of tone in the tone sandhi environment. However, the fact that many syllables with the sandhi tone have homophones makes the categorization of the sandhi tone more complicated than the /k/ categorization in Jaeger's (1986) experiment.

It is often assumed that the lexicon of Mandarin is represented in terms of single-syllable morphemes. To test whether syllables with the sandhi tone are represented as monosyllabic morphemes or as parts of the /tone 2 + tone 3/ phrases in the lexicon of Mandarin, two types of test tokens were used. For half of the test tokens, the target syllables in surface form had homophones whereas the other half did not have. For example, the first syllable of [yn^{35} ɕy^{214}] (<— /yn^{214} ɕy^{214}/) 'permission' was homophonous with /yn^{35}/ 'cloud'. By contrast, the first syllable of [tau^{35} yεn^{214}] (<— /tau^{214} yεn^{214}/) 'movie director' has no tone 2 homophone. If syllables with the sandhi tone are represented as monosyllabic morphemes, there will be two candidate morphemes (/yn^{214}/ or /yn^{35}/) available in the homophone condition and only one candidate morpheme (/tau^{214}/) available in the non-morpheme condition for listeners to choose. Therefore, listeners may make more mistakes on test tokens in the homophone condition than on those in the non-homophone condition by choosing either one of the candidate morphemes.

2.3.1 Methods

2.3.1.1 Experimental Materials

The ten native Mandarin speakers who participated in Experiment 1 also read the phrases used in this perception test. Each of the speakers read 1/10 of the 156 phrases. The phrases were read in the same session as Experiment 1. The three types of tokens used in this perception test are shown in Table 1. There were 15 positive tokens and 15 negative tokens in the learning session. The positive tokens for the tone 2 group were disyllabic phrases beginning with tone 2 followed by any of the four tones: tone 1, tone 2, tone 3, and tone 4. The negative tokens for the tone 2 group were disyllabic phrases beginning with tone 1, tone 3, or tone 4 followed by any of the four tones except the case where tone 3 is followed tone 3. The positive tokens for the tone 3 group were disyllabic phrases beginning with tone 3 followed by tone 1, tone 2, or tone 4. Phrases beginning with tone 1, tone 2, or tone 4 followed by any of the four tones were negative tokens for the tone 3 group.

Table 1. Three types of tokens used in Experiment Three Case Study One. T stands for tone.

positive tokens	negative tokens			test tokens
T2 + T1	T1 + T1	T3 + T1	T4 + T1	
T2 + T2	T1 + T2	T3 + T2	T4 + T2	
T2 + T3	T1 + T3		T4 + T3	T3 + T3
T2 + T4	T1 + T4	T3 + T4	T4 + T4	

The test session included 30 positive tokens, 50 negative tokens, and 20 test tokens. The test tokens were disyllabic phrases beginning with tone 3 followed by tone 3. The first syllables of half of the test tokens in surface forms had homophones. The other half did not have. However, unlike in Experiments 1 and 2, none of the /tone 2 + tone 3/ or /tone 3 + tone 3/ disyllabic phrases had homophones. All the 156 phrases are frequently-used phrases.

Each trial started with a stimulus followed by a 3-second silence, then auditory feedback indicating membership in the target category. Trials were separated by a 3-second pause and blocks of 10 trials were separated by a 6 second pause.

2.3.1.2 Subjects

Thirty native Mandarin speakers (20 female, 10 male) participated in the perception tests. They were randomly divided into two groups. Fifteen listeners were trained to learn tone 2 as the target category. The other 15 subjects were trained to learn tone 3 as the target category.

2.3.1.3 Procedure

Listeners were instructed to concentrate on the first syllables of the phrases and find the features of sound which all the "yes" tokens share. They were asked to circle the answers ("yes" or "no") on the answer sheets after they heard the stimulus and before the feedback. Feedback was given to listeners in both the learning session and the test session. After the learning session, their answers were checked by the experimenter to see whether they had learned the category. If the last ten trails were correctly answered, then the listener was judged to have learned the category. Listeners who did not meet this criterion were asked to redo the learning session until they passed the criterion. (Five subjects in the tone 2 group and nine subjects in the tone 3 group did not reach the criterion on first session.) The test session started after they learned tone 2 or tone 3 as the target category.

2.3.2 Results

Listeners of the tone 2 group and the tone 3 group showed different patterns of categorization of the sandhi tone. Figure 8 shows each group's mean ratios of correct responses for all test tokens according to the trial order of the test tokens. The tone 2 group showed a gradually increasing correct response pattern through the test. The most rapid increase in correct response rate was found in the first five tokens. It dropped to below 0.8 in the next three tokens and raised again to the high response range. The biggest drop of correct response ratio occurred in the first half between trial 25 and trial 37 as well as in the second half between trial 70 and trial 73. The correct response ratio dropped 0.13. By contrast, the tone 3 group had high correct response ratio starting from the beginning of the test. The biggest drop in the correct response ratio (0.13) occurred in the first half session between trail 21 and trial 27. After this drop, the correct response ratio was maintained well above 0.9, even during the long interval between trial 80 and trial 93.

The results for each tone session was divided into first half and second half responses (the first 50 trials, compared to the last 50 trials) to show whether the categorization by listeners improved during the test as a function of feedback. A three-way repeated-measures ANOVA (tone x half x morpheme type) showed significant effects of tone [$F(1, 592)=28.75$, $p < 0.01$] and session half [$F(1, 592)=20.59$, $p < 0.01$]. The mean correct response ratio of tone 3 was higher than that of tone 2. The performance of listeners was better in the second half of the test session than in the first half. The interaction of training tone and session half was also significant, [$F(1,$

$596)=11.17$, $p < 0.001$]. A Tukey post-hoc test showed the first half was significantly different from the second half in the responses of the tone 2 group. However, the tone 3 group showed no change in performance during the test session. The categorization in the two tone sessions differed in the first half session, but not in the second half session. The effect of morpheme type (homophone condition vs. non-homophone condition) was not significant. The mean ratio for the homophone condition and the non-homophone condition were 0.90 and 0.91. The categorization of the test tokens was not affected by whether the first syllables of the test tokens had homophones or not. The mean correct response to the first test token (i.e. prior to any feedback or training about the experimenter-imposed categorization of the sandhi tone) was 0.33 for tone 2 and 0.93 for tone 3.

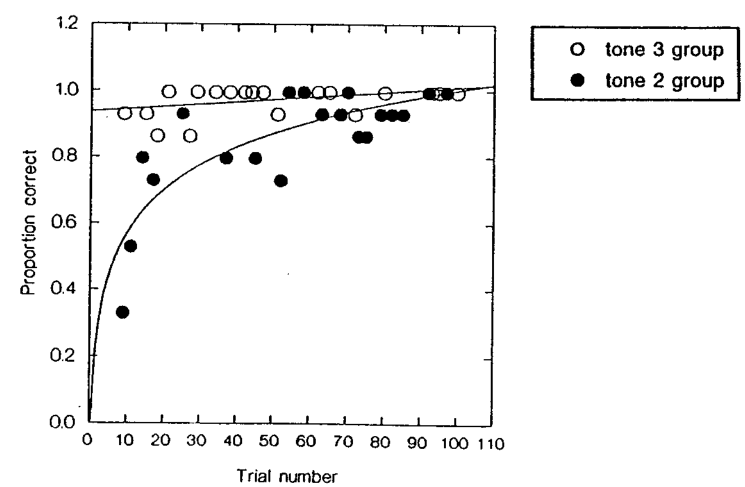


Figure 8. Mean ratios of correct responses for the categorization of the sandhi tone by the tone 2 group and the tone 3 group

2.3.3 Discussion

The sandhi tone was categorized as tone 3 by most Mandarin listeners despite the fact that it is perceptually indistinguishable from tone 2. Listeners categorized the sandhi tone as a low tone (tone 3) even though they were presented with a rising pitch contour which they could not differentiate from an underlying tone 2 in the identification test. The sandhi tone carries the same perceptual information as tone 2 at the phonetic level, but belongs to the tone 3 category at the phonological level. This pattern of categorization suggests that the categorization of the sandhi tone does not result from the derivational mapping from phonetic categories to phonological representation as argued in Keating's model (1990). The surface tone 2 derived from the tone sandhi was not categorized as tone 2 by most listeners.

Morphological information of the syllables does not influence the categorization of the sandhi tone. Listeners who were trained to categorize the sandhi tone as tone 3 consistently categorized the sandhi tone as tone 3 in both the homophone condition and the non-homophone condition. Listeners who were trained to categorize the sandhi tone as tone 2 did not make fewer mistakes in the homophone condition than in the non-homophone condition. These results shows that the categorization of the sandhi tone does not simply arise directly from accessing the monosyllabic morpheme in the lexicon and looking for the underlying tone for the surface tone 2. The fact that most listeners did not assign the same category for the underlying tone 2 to the sandhi tone suggests that the two-syllable phrases as wholes have provided some lexical information for listeners to distinguish the sandhi tone from the underlying tone 2. That is, the two-syllable phrases may be stored as wholes in the lexicon and therefore are activated as wholes. This finding also casts doubts on the assumption that Mandarin lexicon is represented in terms of single-syllable morphemes.

The sandhi tone is acoustically so similar to tone 2 that listeners can also be trained to categorize the sandhi tone as tone 2. However, in order to categorize the sandhi tone as tone 2, listeners apparently needed to repress their intuitive phonological knowledge and at the same time learned to attend to the phonetic information of the sandhi tone, since their performance was not as consistent as the performance of listeners who were trained to categorize tone 3 as the target category. The pattern of responses over the test session also suggests that the formation of this temporary category stored in working memory was affected by the interval duration between test stimuli which contained the sandhi tone. In the first half of the test session, the biggest drop in correct response ratio occurred between the two test tokens separated by the largest number of foil tokens.

Literacy may have some effects on the categorization of the sandhi tone. The listeners who participated in this categorization experiment all had formal school-education in Taiwan. In the elementary schools of Taiwan, an onset rhyme-based phonographic system which is usually called Chu-yin-fu-hao System is taught to the first grade students, before the formal orthographic system is taught. The Chu-yin-fu-hao System includes a tonal marking system (tone 1 unmarked, tone 2 /, tone 3 √, tone 4 \ and the light tone ˙). The tone 3 tone sandhi rule is sometimes also taught in the teaching of the spoken form. However, the sandhi tone is still marked as tone 3 using the Chu-yin-fu-hao System. This may make it easier for the listeners to categorize the sandhi tone as tone 3. On the other hand, the learning of the tone sandhi rule may explain why some listeners categorized the sandhi tone as tone 2 before any reinforcing training and why listeners can be trained to categorize the sandhi tone as tone 2 while most of them initially categorize it as tone 3.

CHAPTER III

CASE STUDY TWO: PLACE COARTICULATION

This chapter investigated Taiwanese place assimilation in an EPG experiment and a perception test using the concept formation method. The production study used different EPG measurements to examine different patterns of gestural coarticulation between velar and dental stops in phrases spoken in various speech rates. The perception test was to show whether and how different patterns of gestural coarticulation affect the categorization of the stop consonants. Regression analyses were performed to show how well the EPG measurements predict the categorization patterns of the stop consonants.

3.1 Experiment One

In the first experiment, Electropalatography (EPG) was used to investigate gestural coarticulation of Taiwanese stop consonants. Taiwanese voiceless stops contrast in four different places of articulation in syllable-final position: /p/, /t/, /k/ and /ʔ/. The stop closure is never audibly released in producing these stop consonants. In syllable-initial position, stop consonants contrast in place of articulation, voicing, and aspiration (voice onset time): /p t k/, /p^h t^h k^h/, /b l (d) g/.

3.1.1 Methods

3.1.1.1 Experimental Materials

Eight two-syllable phrases were used. They are listed in Table 2. The first syllable ends with a voiceless alveolar stop /t/ or a voiceless velar stop /k/. The second syllable starts with a voiceless aspirated alveolar /t^h/ or velar /k^h/ stop or a voiceless unaspirated alveolar /t/ or velar /k/ stop. These 8 consonant sequences were classified into 4 groups according to place of articulation: **tk** (/tk tk^h/), **kt** (/kt kt^h/), **tt** (/tt tt^h/) and **kk** (/kk kk^h/). Consonant sequences with the same place of articulation **tt** and **kk** are used as control tokens.

Table 2. Taiwanese phrases used in Case Study Two Experiment One.

/pat$^{\underline{53}}$ taŋ^{55}/ "another year"	/pat$^{\underline{53}}$ t^{h}aʔ$^{\underline{21}}$/ "another pile"
/pat$^{\underline{53}}$ kaŋ^{55}/ "another day"	/pat$^{\underline{53}}$ k^{h}aŋ^{55}/"another hole"
/lak$^{\underline{53}}$ taŋ^{55}/ "ten years"	/lak$^{\underline{53}}$ t^{h}aʔ$^{\underline{21}}$/ "ten piles"
/lak$^{\underline{53}}$ kaŋ^{55}/ "ten days"	/lak$^{\underline{53}}$ k^{h}aŋ^{55}/"ten holes"

3.1.1.2 Subjects

Six native speakers of Taiwanese: 3 male and 3 female participated in the experiment. All the speakers also spoke Mandarin and English as their second and third languages. None of them had speech problems (by self-report).

3.1.1.3 Recording & Measurements

Articulatory data were recorded using the Kay Elemetrics EPG system. In this system, the productions of speakers are filtered with custom pseudo-palates (Figure

10a) in which 96 electrodes are systematically arranged over the surface of the alveolar ridge hard palate. Each of the type was spoken at several different speaking rates. The speakers started by repeating the phrase five times at slow rate and then increased the speaking rate gradually until reaching his/her fastest rate within the interval of 10 repetitions.

Speech waveforms were directly digitized (10 kHz sampling rate, 16 bits/sample). The duration of the vowel in the first syllable was measured as the reference for speech rate. Duration was measured on a wide-band spectrogram using Kay CSL from the onset of the vowel to the beginning of the stop closure as shown by the cursors in Figure 9.

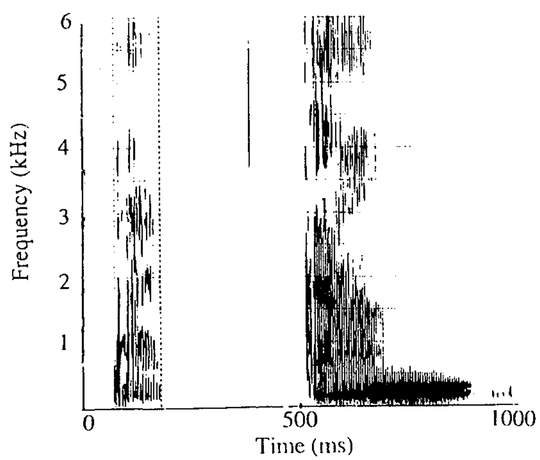


Figure 9. A sample spectrogram showing the measurement of the duration of the first vowel

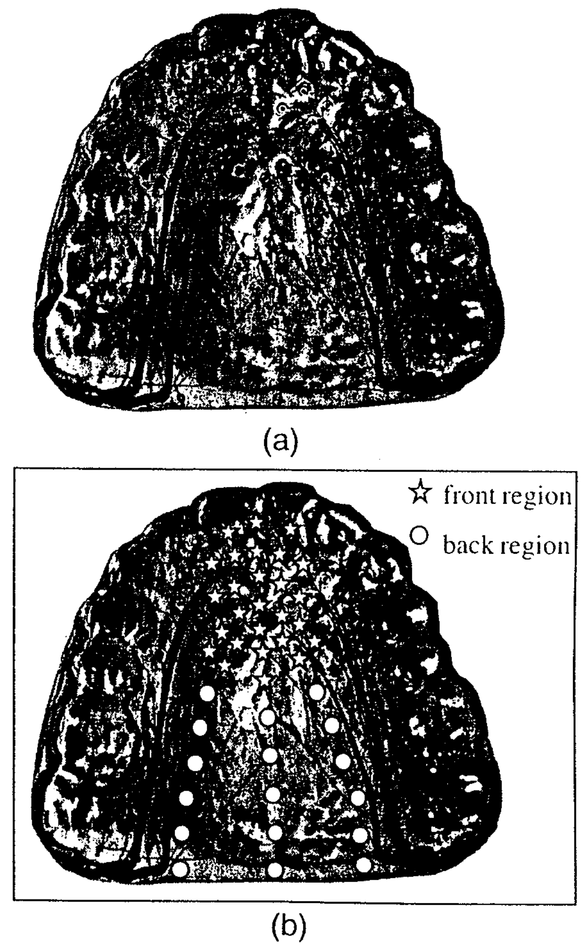


Figure 10. (a) A typical artificial acrylic palate used in Experiment One Case Study Two (b) An artificial palate with defined contact regions

Each speaker's palate was divided into two regions, front and back, as shown in Figure 10b. The boundary between the two regions was set at the beginning of the hard palate, just behind whethere the slope from the alveolar ridge reaches the highest point of the hard palate. The contact patterns at the front region and the back region were displayed separately in terms of percentage of electrodes contacted. The percentage was calculated at every 10 millisecond time frame. To focus on contact patterns associated with complete consonant closures, only electrodes along the midline of the vocal tract were included in the calculation as shown in Figure 10b. For example, when making a front contact in producing /t/, contact along the sides of the palate which does not contribute to distinguishing places of articulation was not included. So, for each speaker, there were 27 electrodes in the front region and 17 electrodes in the back region in total.

EPG measurements included duration of contact and percent of electrodes contacted in each region. The percentage of electrodes contacted was measured at the peak contact of each region--the highest percentage of the electrodes contacted. For each token, the following temporal points were measured: beginning of contact (C1 begin & C2 begin), end of contact (C1 end & C2 end), and end of peak contact (C1 maxend & C2 maxend). Based on these measurements, numbers were calculated as shown in (4) and (5). These points are illustrated in schematized contours as shown in Figure 11. The closing gesture was from the beginning of the contact to the end of the maximum contact where the proportion of electrodes contacted started to decrease.

(4) Values calculated for each individual gesture:

1. C1 closing gesture duration (C1 closing) = C1 maxend - C1 begin
2. C2 closing gesture duration (C2 closing) = C2 maxend - C2 begin

(5) Values calculated to show gestural coarticulation:

1. C2 onset latency relative to C1 onset (Δ onset) = C2 begin - C1 begin
2. C2 Peak contact latency (Δ peak) relative to C1 peak contact = C2 maxend - C1 maxend
3. C2 offset latency relative to C1 offset (Δ offset) = C2 end - C1 end
4. Duration of simultaneous contact at the front region and the back region during the closing gestures (Δ overlap)= C1 maxend - C2 begin.
5. Proportion of the first closing gesture overlapped by the second closing gesture (% overlap) = Δ overlap / C1 closing
6. Gestural phasing between C1 and C2 (phase) = Δ onset / C1 closing

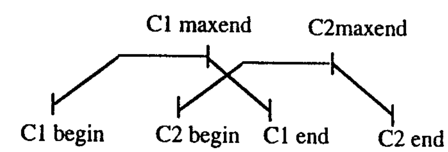


Figure 11. Schematic contact contours and measurement points

3.1.2 Results

As expected, some speakers' productions of the coarticulated /t/ showed no contact in the front region and some of the productions of coarticulated /k/ showed no contact in the back region. The percentages of /t/ or /k/ at the onset or coda positions with no contact in the front region or the back region were listed in Table 3. No-contact tokens occurred mostly in the fastest speech rates, but it also occurred in two speakers' productions in the "normal" speech rate at the middle of the rate continuum (F1 & M2).

/tk^h/ was produced by one speaker (M2) with no contact in the front region in all 15 utterances. As shown in Table 3, more consonants with complete absence of contact were found at the coda position than at the onset position. In fact, for /k/ at the onset position, only 2 speakers (F1 & M1) had any tokens with no contact in the back region. There were more tokens with no contact in the front region for /t/ than there were tokens with no contact in the back region for /k/. The largest percentage of tokens with no contact were those with no contact in the front region for /t/ at the coda position. Tokens with no contact in the front region or in the back region were not included in the statistical analyses of other measurements of the contact pattern in that specific region. However, they are shown in the figures.

Table 3. Percent of tokens with no front or back contact.

	front region		back region	
subject	coda	onset	coda	onset
F1	53%	0	0	3%
F2	10%	0	23%	0
F3	7%	0	3%	0
M1	10%	0	3%	3%
M2	83%	0	47%	0
M3	33%	0	0	0
pooled	33%	0	13%	1%

The percentage of electrodes contacted at maximum contact and the duration of the closing gestures were regressed against consonant group (a nominal variable) and

speech rate (a continuous variable measured in terms of the duration of the vowel in the first syllable of the phrase). Two dummy variables were created for the three consonant groups, with **tt** or **kk** was used as the "reference" group.

Figure 12 shows percent of maximum contact in the front region for each consonant group according to speech rate for each speaker. The results of regression are listed in Table 4. The regression accounted for a significant percentage of the variance in the production of each speaker except for speaker M2 who showed a high percentage of tokens with no contact for **tk** tokens (Note the very large line of tokens at zero contact ranging from rate 200 to 50 ms.). In all five other speakers, the percent of maximum contact varied as a function of both speech rate and consonant group. Percent of maximum contact decreased as speech rate increased. The percent of front contact for **kt** and **tk** was about the same as or slightly less than **tt** in the production of most speakers. Nevertheless, **tk** had a higher percentage than **tt** in the productions of two speakers.

Table 4. Adjusted R^2 and partial coefficients for a linear fit for percent of maximum front contact regressed against speech rate and consonant groups. **=$p<0.01$, *=$p<0.05$.

subject	R^2	rate	V1 (**kt**)	V2 (**tk**)	constant (**tt**)
F1	.293**	.054**	1.764	9.88**	40.397**
F2	.158**	.089**	-9.321**	-12.285**	56.084**
F3	.55**	.249**	-2.043	15.962**	17.852**
M1	.471**	.239**	-8.822*	-20.389**	62.775**
M2	.009	-.034	4.594	-13.345	57.634**
M3	.385**	.242**	2.535	2.814	26.084**

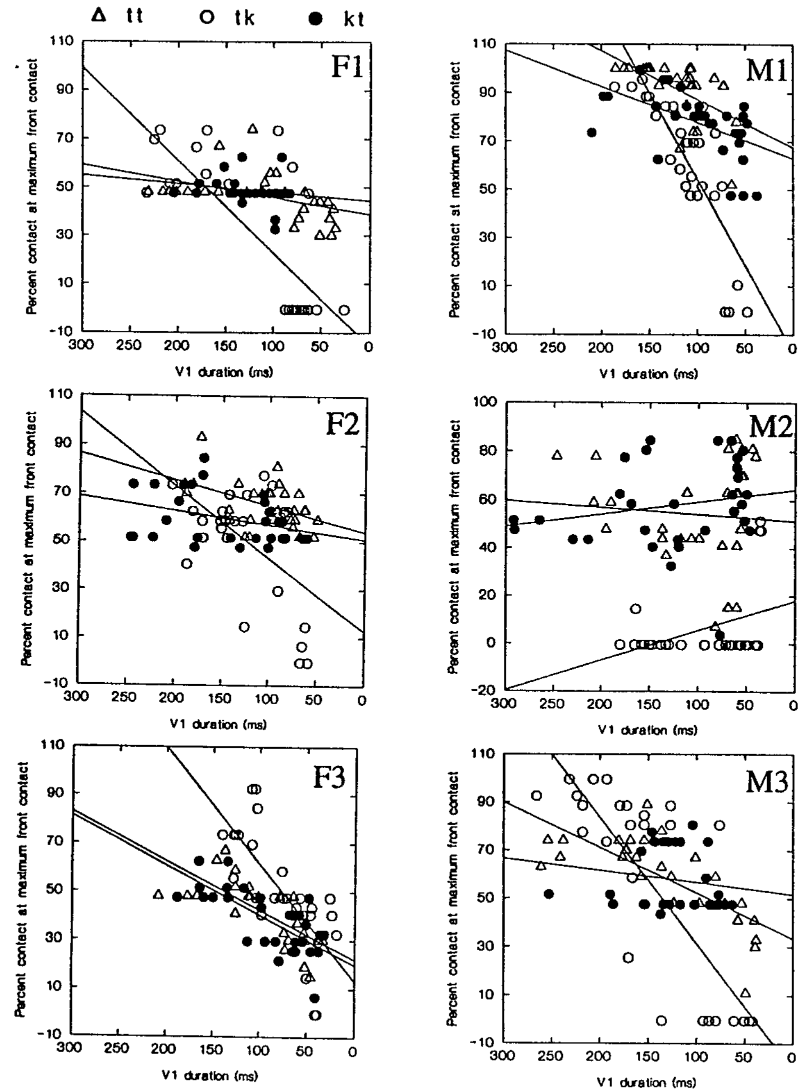


Figure 12. Percent of electrodes contacted at the maximum front contact of **kt**, **tk** and **tt** for each speaker shown from slow to fast speech rate

Figure 13 shows the maximum contact in the back region according to speech rate for each speaker. The results of regression of percent of maximum contact against speech rate and consonant group are shown in Table 5. The regression analysis was significant for all speakers again except speaker M2. The contact in the back region varied according to speech rate in the production of most speakers, and especially for the **tk** sequence. The percent of maximum contact decreased as speech rate increased. There was more contact for **tk** than **kk** in the production of all speakers except speaker M2. Three speakers also had more contact for **kt** than for **kk**.

Table 5. Adjusted R^2 and partial coefficients for a linear fit for percent of maximum back contact regressed against speech rate and consonant groups. **=$p<0.01$, *=$p<0.05$.

subject	R^2	rate	V1 (kt)	V2 (tk)	constant (kk)
F1	.354**	.083**	5.016	21.145**	3.179
F2	.577**	.02**	18.752**	40.612**	15.442**
F3	.901**	.096**	9.907**	58.662**	16.322**
M1	.596**	.19**	29.72**	35.931**	6.891
M2	.026	-.082*	-7.594	-2.472	53.126**
M3	.663**	.223**	5.346	29.496**	7.067

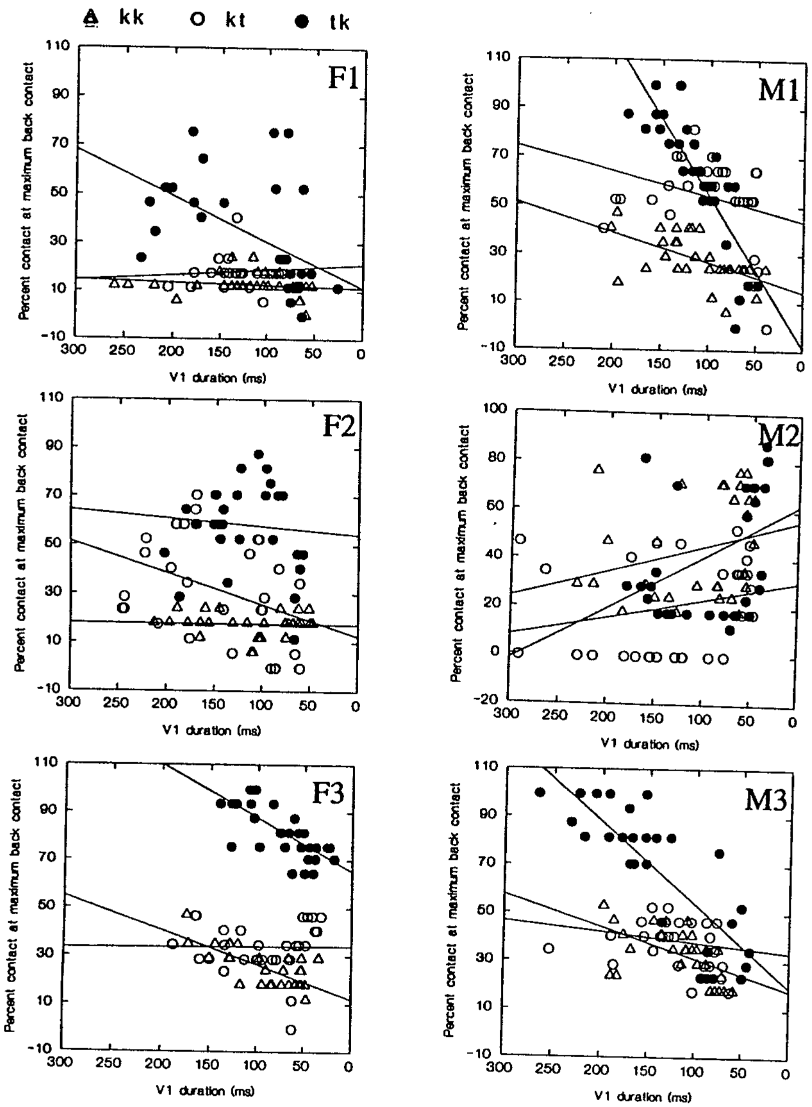


Figure 13. Percent of electrodes contacted at the maximum back contact of **kt**, **tk** and **kk** for each speaker shown from slow to fast speech rate

Duration of the dental closing gesture in all consonant groups for each speaker is shown from slow to fast speech rate in Figure 14. Table 6 shows the results of regression against consonant group and speech rate. The regression analysis was significant for all speakers except speaker M2. The closing gesture duration decreased as speech rate increased. The duration of **tk** and **kt** was shorter than **tt**'s.

Table 6. Adjusted R^2 and partial coefficients for a linear fit for duration of dental closing gesture regressed against speech rate and consonant groups. $**=p<0.01$, $*=p<0.05$.

subject	R^2	rate	V1 (**kt**)	V2 (**tk**)	constant (**tt**)
F1	.607**	1.526**	-110.677**	-194.216**	63.607**
F2	.581**	1.181**	-96.414**	-27.889*	63.942**
F3	.556**	1.584**	-67.908**	-41.941*	17.665
M1	.611**	2.083**	-95.866**	-93.884**	2.851
M2	.01	-.327	-5.363	-54.984	175.147**
M3	.317**	1.901	12.236	-123.063**	30.709

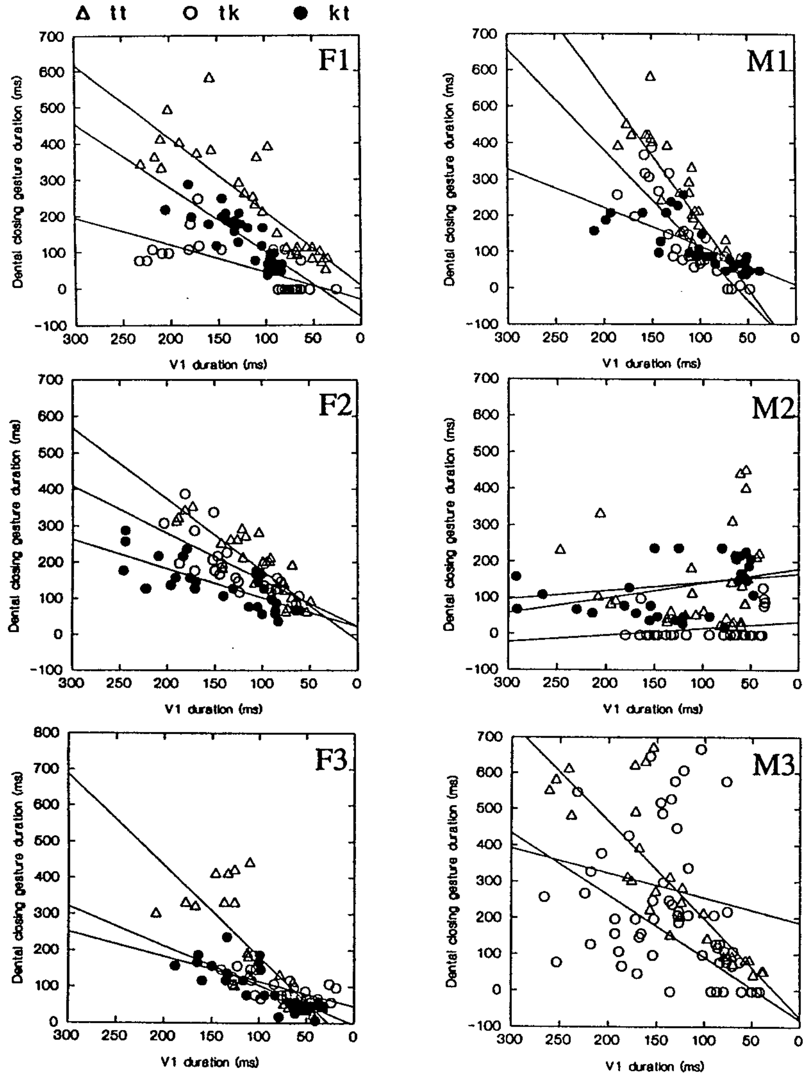


Figure 14. Duration of dental closing gesture of **kt**, **tk** and **tt** for each speaker shown from slow to fast speech rate

As with the dental closing gesture, duration of velar closing gesture was affected by speech rate and consonant group (Figure 15). The regression was significant for each speaker (Table 7). Velar closing gesture duration decreased as speech rate increased except for speaker M2, who had very many no contact tokens in which no contact duration could be measured. However, the duration of the velar closing gesture was less affected by consonant group than that for the dental closing gesture in Figure 14. The duration of **tk** was about the same as **kk** except for the case of speaker F1 where the duration was shorter than **kk**. The duration of **kt** was also shorter than **kk** in the production of three speakers, but the difference was smaller than that found in dental closing gesture.

Table 7. Adjusted R^2 and partial coefficients for a linear fit for duration of velar closing gesture regressed against speech rate and consonant groups. $^{**}=p<0.01$, $^{*}=p<0.05$.

subject	R^2	rate	V1 (kt)	V2 (tk)	constant (kk)
F1	.283**	.764**	-39.068*	-60.239**	84.151
F2	.349**	1.129**	-38.991	-15.966	43.298
F3	.55**	1.307**	-45.187**	-11.423	59.061
M1	.698**	2.002**	-2.939	-22.141	-28.194
M2	.04	-.31*	-19.388	-14.038	184.688
M3	.461**	1.857**	-126.401**	-43.591	49.883

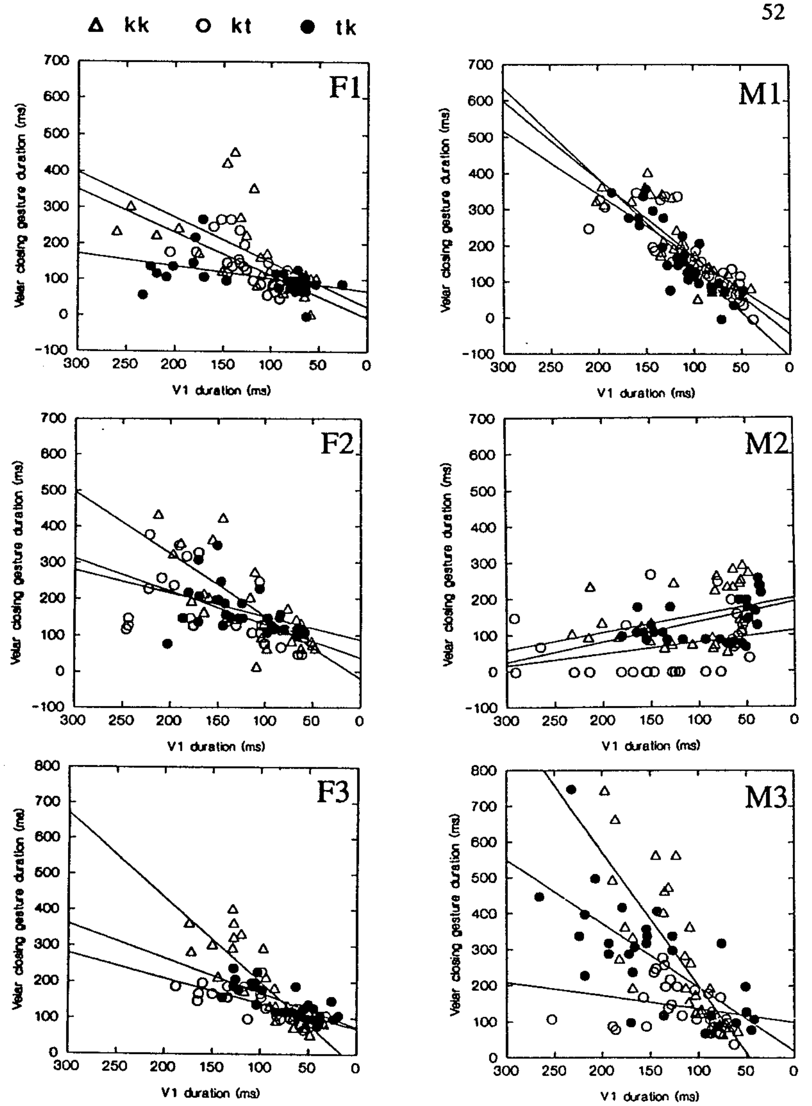


Figure 15. Duration of velar closing gesture of **kt**, **tk** and **kk** for each speaker shown from slow to fast speech rate

Figures 16-18 show the onset, peak, and offset latency of the gesture of the second consonant relative to the gesture of the first consonant as a function of speech rate for each speaker. These measurements of inter-gestural timing were regressed against speech rate and consonant group. One dummy variable was created for the two consonant groups **tk** and **kt**.

The regressions (Tables 8-10) for the latency measurements were significant in the productions of most of the speakers indicating that the relative timing of the two gestures varied as a function of both speech rate and consonant group. The latency decreased as speech rate increased, especially in **kt**. The onset latency of **kt** was longer than that of **tk** indicating that the initiation of the velar gesture relative to the preceding dental gesture in **tk** was relatively earlier than the initiation of a dental gesture relative to a preceding velar gesture in **kt**. In some productions of some speakers, the onset latency was even negative for **tk**. That is, the velar gesture started earlier than the dental gesture in fast speech. This phenomenon where the second gesture started earlier than the first gesture only occurred in very few cases of the **kt** sequence. However, the offset latency for **kt was** shorter than for **tk**. The completion of the velar gesture was set later than that of the dental gesture relative to the completion of their preceding gestures. The peak contact latency in the **kt** sequence was not longer than in the **tk** sequence except for the productions of two speakers.

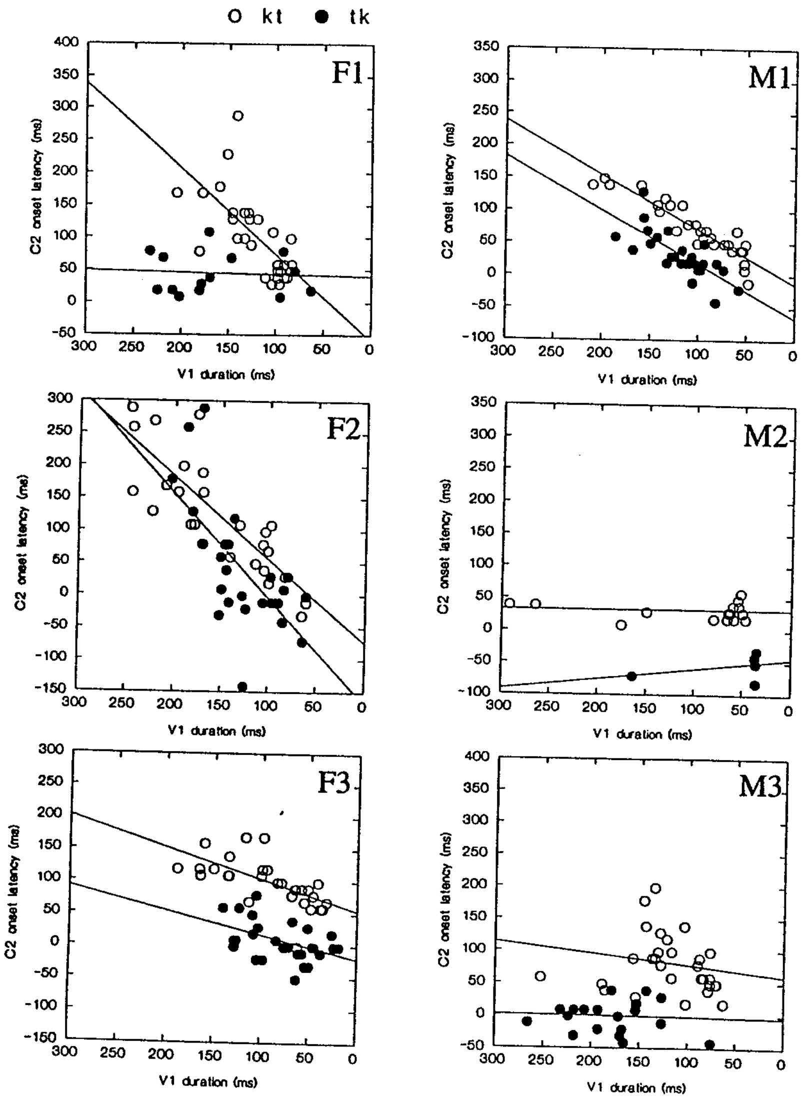


Figure 16. C2 onset latency of **kt** and **tk** for each speaker shown from slow to fast speech rate

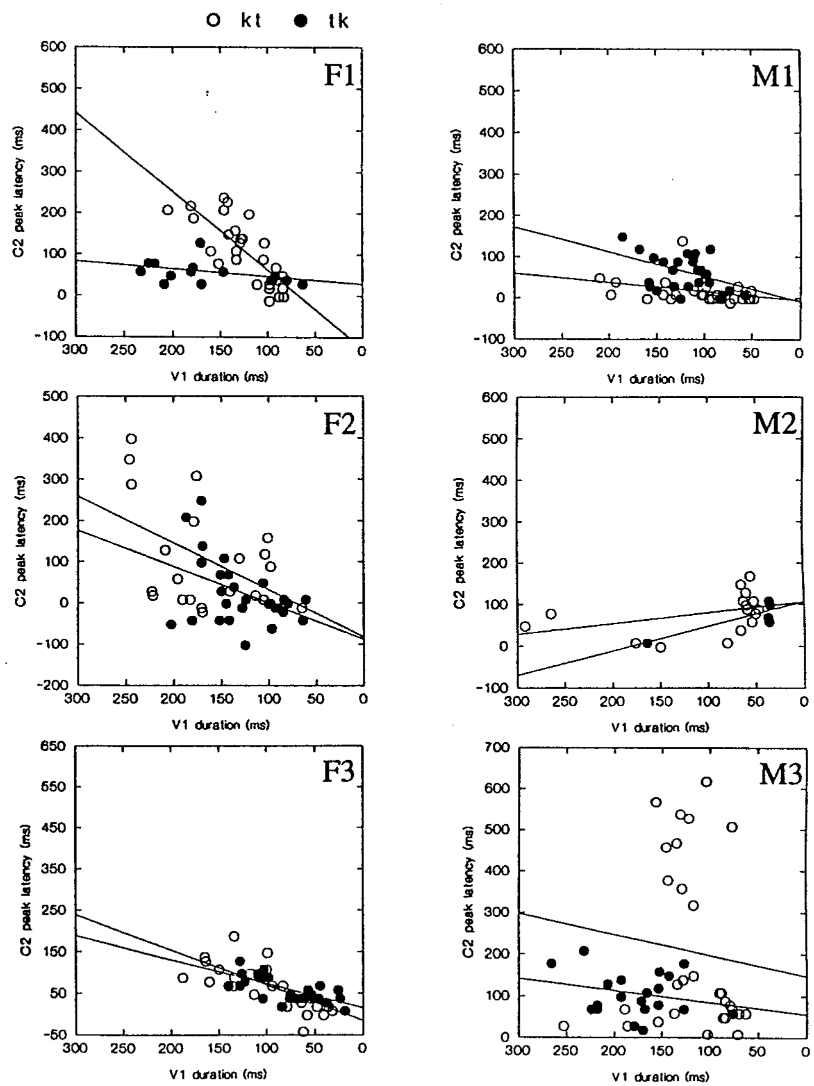


Figure 17. C2 peak latency of **kt** and **tk** for each speaker shown from slow to fast speech rate

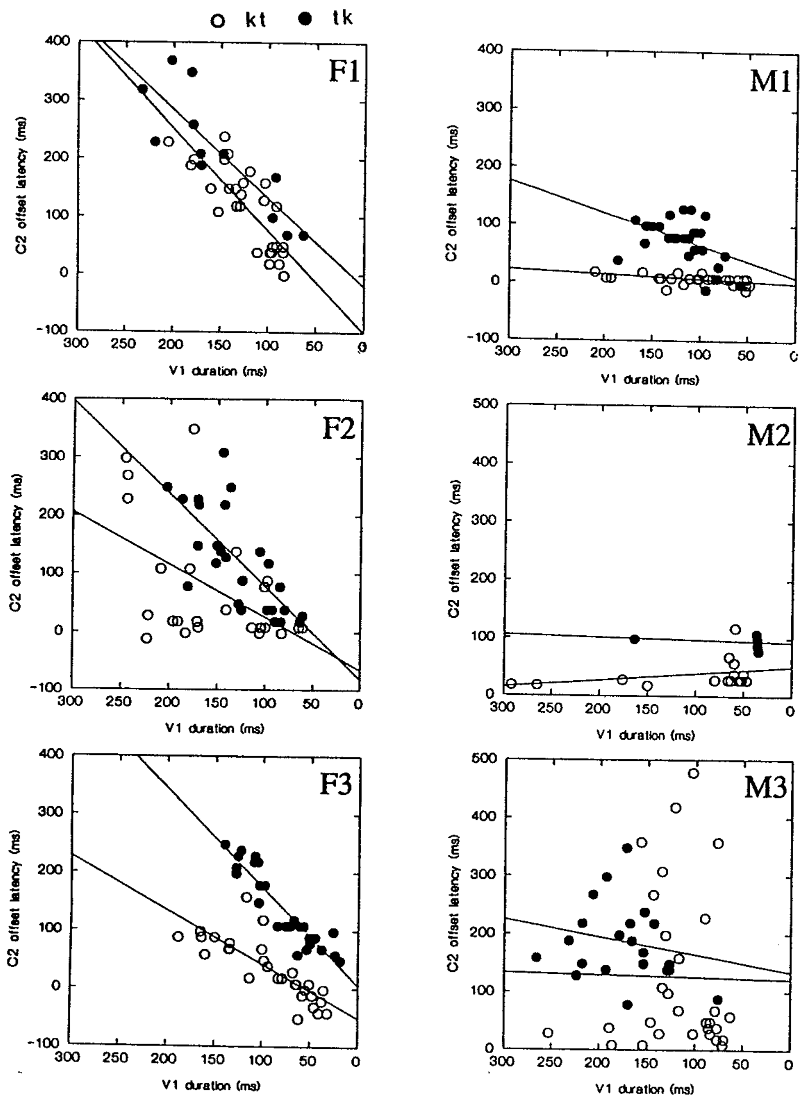


Figure 18. C2 offset latency of **kt** and **tk** for each speaker shown from slow to fast speech rate

Table 8. Adjusted R^2 and partial coefficients for a linear fit for onset latency regressed against speech rate and consonant groups. **=p<0.01, *=p<0.05.

subject	R^2	rate	V1 (**kt**)	constant (**tk**)
F1	.313**	.556**	80.847**	-45
F2	.624**	1.389**	47.597**	-137.467**
F3	.757**	.446**	84.608**	-26.157**
M1	.788**	.835**	53.47**	-66.776**
M2	.858**	-.02	87.256**	-52.775**
M3	.533**	.114	88.585**	-20.653

Table 9. Adjusted R^2 and partial coefficients for a linear fit for peak latency regressed against speech rate and consonant groups. **=p<0.01, *=p<0.05.

subject	R^2	rate	V1 (**kt**)	constant (**tk**)
F1	.353**	.883**	82.154**	-84.999*
F2	.299**	1.053**	41.221	-110.727**
F3	.505**	.749**	-8.106	1.728
M1	.376**	.325**	-40.922**	23.681
M2	.143	-.3*	21.416	88.493**
M3	.06	.413	124.493*	32.893

Table 10. Adjusted R^2 and partial coefficients for a linear fit for offset latency regressed against speech rate and consonant groups. $**=p<0.01$, $*=p<0.05$.

subject	R^2	rate	V1 (kt)	constant (tk)
F1	.782**	1.79**	-56.873**	-46.817
F2	.364**	1.127**	-78.333**	-21.08
F3	.83**	1.209**	-120.206**	43.465**
M1	.65**	.207*	-64.503**	50.988**
M2	.547**	-.1	-52.81**	102.181**
M3	.03	.144	-52.802	162.536*

The phasing of the gestures is shown in Figure 19 according to speech rate. The "phasing" value referred to the time of onset of the second gesture relative to the duration of the first gesture. C2 onset latency was expressed as a ratio of C1 duration. If the second gesture was initiated after the closing gesture of the first consonant, then this phasing value is larger than one. For phase values less than one, the smaller the value, the earlier the onset of the second gesture relative to the gesture of the first consonant was. The productions of all speakers were affected by speech rate and consonant group. The regression was significant for each speaker (Table 11). However, only two speaker's (M1 & F2) productions were significantly affected by speech rate. For these speakers, as speech rate increased, phase decreased: the second gesture occurred earlier in the first gesture. **kt** had a larger phase than **tk**. The phasing of the velar gesture in **tk** was earlier than the dental gesture in **kt**.

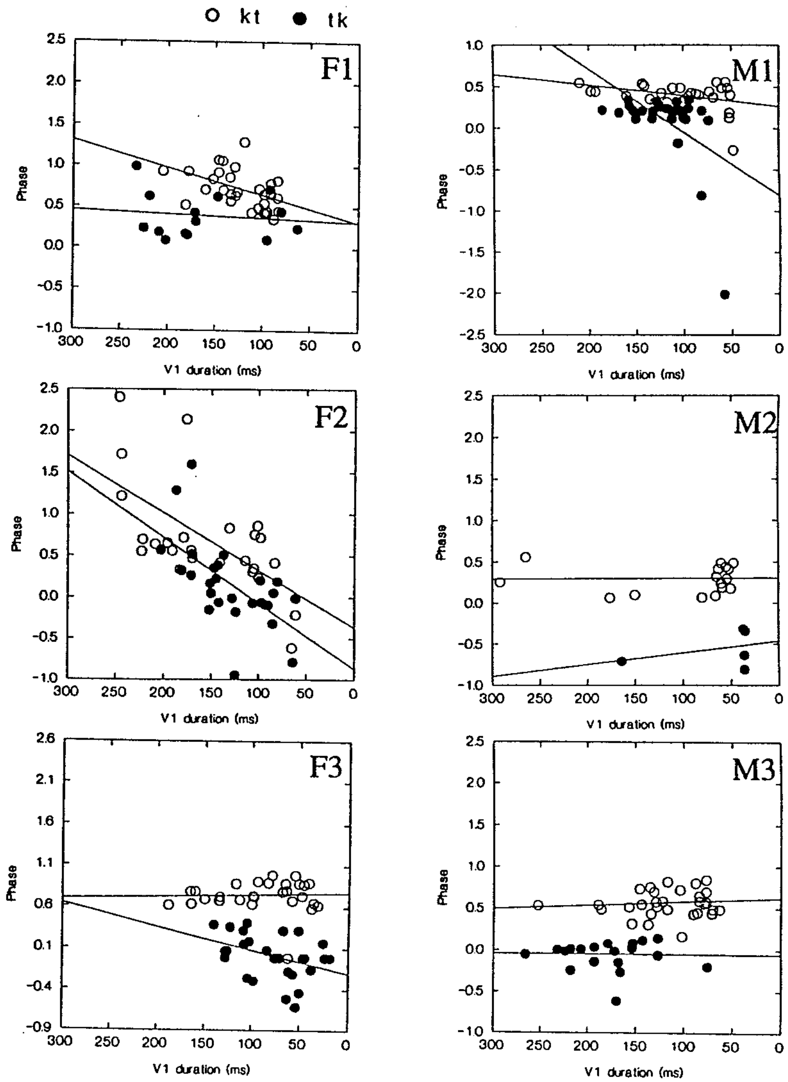


Figure 19. The phasing of the gestures of **kt** and **tk** for each speaker shown from slow to fast speech rate

Table 11. Adjusted R^2 and partial coefficients for a linear fit for phase regressed against speech rate and consonant groups. **=p<0.01, *=p<0.05.

subject	R^2	rate	V1 (**kt**)	constant (**tk**)
F1	.314**	.002	.39**	.128
F2	.48**	.007**	.36**	-.779**
F3	.723**	.001	.725**	-.052
M1	.239**	.003**	.355**	-.27
M2	.802**	0	.861**	-.541**
M3	.773**	0	.604**	-.003

Figures 20 and 21 show the duration of the first gesture and the proportion of the gesture of the first consonant overlapped by the gesture of the second consonant. The duration and proportion of overlapping were significantly affected by speech rate in the productions of three speakers. As speech rate increased, raw duration of the interval of overlap decreased, but the proportion of the first gesture overlapped by the second gesture increased. This overlap proportion in the **kt** sequence was significantly smaller than in the **tk** sequence (Table 12). The duration of overlap in the **kt** sequence was also significantly smaller than in the **tk** sequence except for two speakers (Table 13). A larger proportion of the dental closing gesture was overlapped by the following gesture than the velar gesture. That is, there was a smaller proportion of the dental closing gesture which had only front contact than the velar closing gesture which had only back contact.

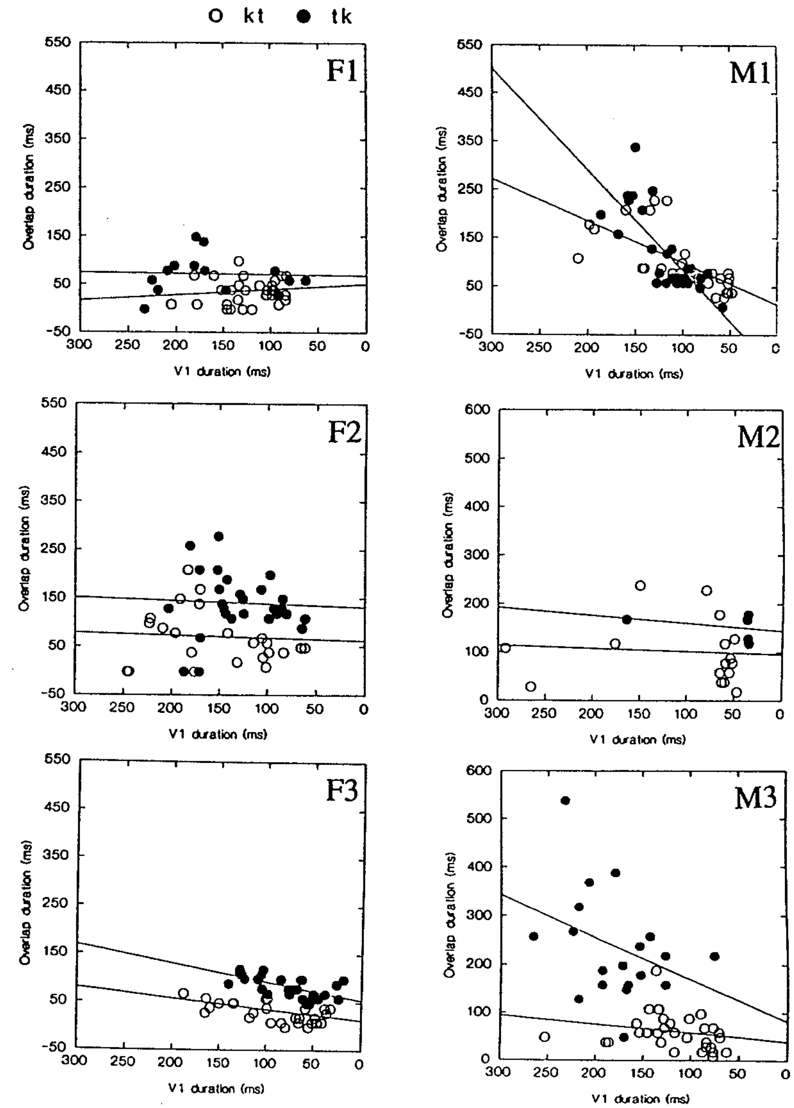


Figure 20. The duration of overlap of **kt** and **tk** for each speaker shown from slow to fast speech rate

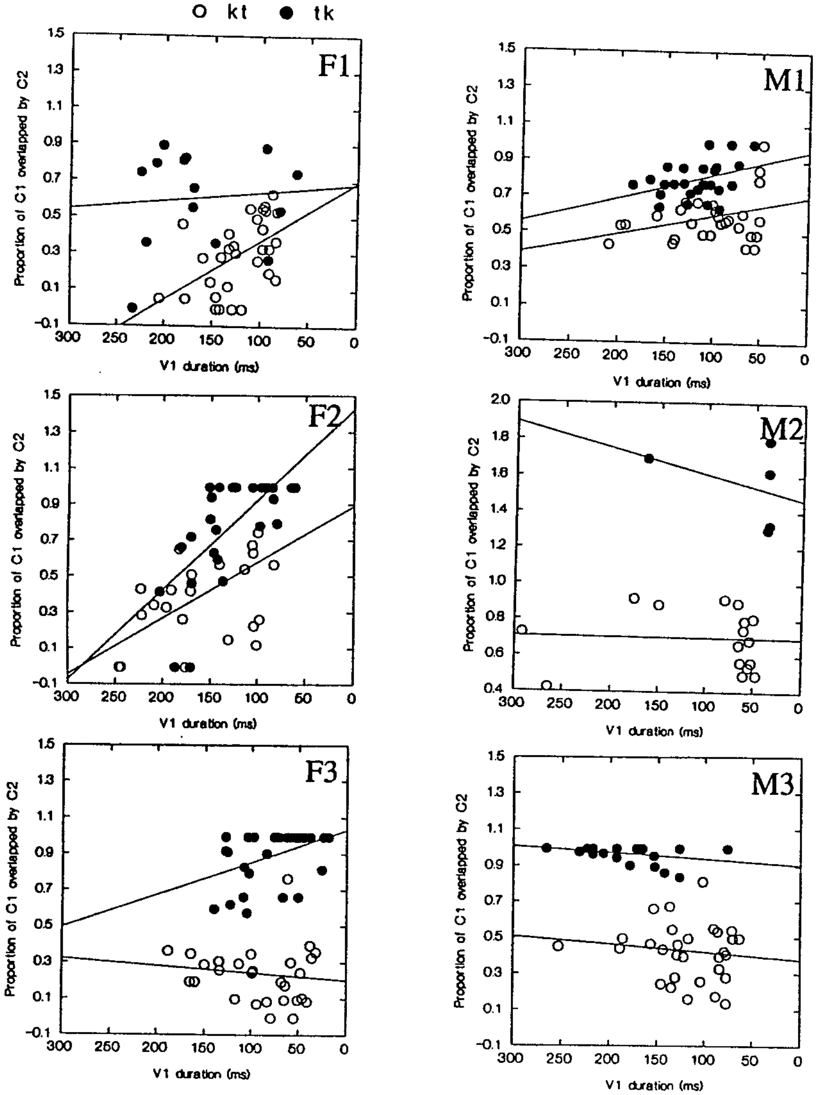


Figure 21. The proportion of overlap of **kt** and **tk** for each speaker shown from slow to fast speech rate

Table 12. Adjusted R^2 and partial coefficients for a linear fit for proportion of overlap regressed against speech rate and consonant groups. $**=p<0.01$, $*=p<0.05$.

subject	R^2	rate	V1 (kt)	constant (tk)
F1	.334**	-.002	-.374**	.867**
F2	.563**	-.004**	-.275**	1.262**
F3	.822**	0	-.647**	.925**
M1	.493**	-.001**	-.221**	.931**
M2	.802**	0	-.861**	1.541**
M3	.817**	0	-.515**	.895**

Table 13. Adjusted R^2 and partial coefficients for a linear fit for duration of overlap regressed against speech rate and consonant groups. $**=p<0.01$, $*=p<0.05$.

subject	R^2	rate	V1 (kt)	constant (tk)
F1	.185**	-.036	-36.173**	77.242**
F2	.219**	.059	-70.694**	132.423**
F3	.665**	.28**	-54.2**	62.131**
M1	.444**	1.208**	-.217	-22.083
M2	.039	.072	-54.873	149.548**
M3	.599**	.463	-147.363**	153.505**

3.1.3 Discussion

Some consonants spoken in faster speech rates showed no sign of contact in the current study. This phenomenon may be interpreted in two different ways. It may be that the gesture is deleted due to place assimilation. The underlying specification of place of articulation for the first consonant is changed to that of the second consonant, so the underlying gesture for that consonant is not even intended in the production. For example, the dental stop in **tk** assimilates to the velar stop, so the dental gesture is deleted. However, Browman & Goldstein's articulatory model (1989; 1990) and Lindblom's (1963) study of vowel reduction have suggested that articulatory reduction may sometimes lead to undershoot of articulatory targets or formant targets. For example, dental gesture may be intended and the tongue may even move toward the dental target, but without any actual contact at the dental area achieved. Like gestural deletion, tokens like this also will not show any sign of contact shown in the EPG measurements. There is no way to differentiate these two interpretations given only the EPG data. Therefore, in the text following, the term "gestural deletion" will be used to refer to tokens with no observed contact, but no interpretation should be inferred. That is, the term does not imply a "real" deletion of the gesture as opposed to undershoot of the gesture.

Gestures were "deleted" more frequently at the coda position than the onset position. Farnetani and Busà (1994) found this same asymmetry for "gesture deletion" in the Italian consonant clusters /nk/ and /kn/, where the gesture for the nasal consonant was frequently "deleted" in /nk/ but not in /kn/. One possible account for such an asymmetry of place assimilation is based on acoustic-auditory factors rather than articulatory coarticulation. In Taiwanese, stop consonants at the coda position (which are sometimes described as "unreleased" stops) do not show clear acoustic cues of the

consonant at the release of the closure, hence they are acoustically "weaker" than stop consonants at the onset position. This may cause the gestures of syllable-final stops to be reinterpreted as being not necessary in the intent, so that the gesture might actually be deleted. In fact, this is Ohala's (1990; 1993) account and he attributes this asymmetry to the difference between CV and VC in the saliency of place cues. The place cues in the transition between the consonant and the vowel in CV are stronger than in VC. Therefore, in VCCV, heterorganic consonant clusters may be heard as homorganic clusters of the second consonant by listeners. The place cues of CV dominate the place cues of VC. Listener-turned speakers then produce the heterorganic clusters as homorganic clusters or singletons.

However, this explanation cannot explain why most speakers only showed complete place assimilation in some of their productions in the current study. If intent changes, it is not changed categorically. The same speaker can produce truly assimilated coda stops, or can produce only gradiently "weakened" stops.

The dental gesture was "deleted" more frequently than the velar gesture. This difference in "gestural deletion" can be explained by the general articulatory difference between /t/ and /k/. The movement toward an apical closure is shorter and the stop can be released rapidly (Stevens, Keyser, & Kawasaki, 1986). The release of /t/ can be accomplished rapidly by the forward and downward movement of only the front part of the tongue blade. For /k/, the constriction is made by the tongue dorsum, so the movement toward the closure takes longer and is released less rapidly than /t/. Thus, the duration of the dental closing gesture was shorter than the duration of the velar closing gesture. As speech rate increased, the duration of gestures became shorter. Because the dental closing gesture had shorter duration in general, it may be reduced to be no contact at all in fast speech rate.

Gestural reduction occurred both in time and in space in this study. Contact area became smaller and contact duration became shorter as speech rate increased. Coarticulation , on the other hand, caused the gestures to be reduced in duration, but not in contact area. The dental gesture or the velar gesture had a shorter closing gesture duration when followed by a stop with the other place of articulation as compared to the geminate consonants in which the dental gesture or the velar gesture extended from the coda to the onset. Nevertheless, the percent of electrodes contacted at maximum contact of the dental gesture or the velar gesture in **tk** and **kt** was not smaller than in the geminates in the productions of most speakers. More speakers' productions showed increase in back contact than in front contact in **tk** and **kt** (comparing to the geminates). The contact area of the velar gesture and the dental gesture was increased by the coarticulation with the other gesture.

The increase in the front contact and the back contact for **tk** and **kt** may be attributed to articulatory constraints of the articulator. The dental gesture and the velar gesture were produced by two parts of the same articulator--the tongue. When the tongue tip and tongue dorsum are raised to make the dental gesture and the velar gesture almost simultaneously, the dental contact and the velar contact were extended backward and forward respectively by the raising of the other part of the tongue.

The temporal relationship between the dental gesture and the velar gesture was assessed by measuring the latency of the second gesture, and the "phasing"--the time of the onset of the second gesture relative to the duration of the first gesture and percentage of overlapping between the two gestures. The latency of the second gesture relative to the first gesture decreased as speech rate increased. The second gesture was initiated earlier in faster speech. The relative phasing and the proportion of overlap between the two gestures were less affected by speech rate. Only the productions of two speakers

had the second gesture occurring earlier in the first gesture and the percent of overlap increasing as speech rate increased. The duration of the first gesture and percent of its duration overlapped by the second gesture were also less affected by speech rate than was the latency. In some speakers' productions, while the initiation of the two gestures became closer spaced in time as speech rate increased, the phasing and the percent of overlapping between the two gestures were mostly maintained.

Latency, phasing and overlapping also differed between **tk** and **kt**. The onset latency of the velar gesture relative to the dental gesture was shorter than that of the dental gesture relative to the velar gesture. The velar gesture started earlier than the dental gesture relative to their preceding gesture. This may be due to the articulatory difference between the velar gesture and the dental gesture. Tongue dorsum is larger and heavier than the tongue tip, so it takes more time for the tongue dorsum to make a velar closure than the tongue tip to make a dental closure. Therefore, the velar gesture in **tk** needed to start earlier than the dental gesture in **kt**. The velar gesture may even need to start before its preceding gesture when speaking faster.

The velar gesture is in general longer than the dental gesture, so the peak latency and offset latency of the velar gesture relative to the preceding gesture was longer than the dental gesture. The phasing of the velar gesture occurred earlier in the preceding gesture than that of the dental gesture, so the dental gesture was overlapped more by the following velar gesture than the velar gesture.

3.2 Experiment Two

This experiment was designed to investigate how different patterns of gestural assimilation and reduction are interpreted by native speakers using the same concept formation paradigm as was used in Experiment 3 of Case Study 1. The task tested the categorization of final consonants in the syllables /pat/ and /lak/ which showed gestural coarticulation with the following initial consonants in Experiment 1. It was designed to answer the question: Do different degrees of gestural overlapping or deletion affect the categorization of the stop consonants?

3.2.1 Methods

3.2.1.1 Experimental Materials

Ten single syllable words which end with one of the stop consonants /p, t, k, ʔ/ were used as positive tokens and negative tokens. They are listed in Table 14. These words were read in isolation in slow, normal, and fast speech rates by a female subject (F1) who also participated in the production study (Experiment 1). The words were read in the same session as the words used in Experiment 1. Two repetitions of each word in slow and fast speech rates were chosen to be used in the perception test.

Table 14. Monosyllabic Taiwanese words used in Case Study Two Experiment Two.

/pat53/	'another'	/pak21/	'north'
/tsat53/	'section'	/tsak21/	'disturb'
/lat53/	'power'	/lak53/	'six'
/lap53/	'pay'	/tsap53/	'ten'
/paʔ21/	'hundred'	/tsaʔ21/	'bring'

For the test in which /t/ was the target category, the three words ending with a /t/ were positive tokens. Words ending with any other stop consonant were negative tokens. When /k/ was the target category, the three words ending with /k/ were positive

tokens and words ending with any other stop consonant were negative tokens. Test tokens in the test session were the first syllables excised from the disyllabic words produced by subject F1 in Experiment 1. Each excised token consisted of the entire first syllable and all of the following onset consonant up to the release of the stop closure.

In the learning session of each test, there were 12 positive tokens and 14 negative tokens. There were 6 positive tokens and 28 negative tokens in the test session. Eleven repetitions (token 5-token 15) of each of the 6 consonant sequences /tt, tt^h (kk, kk^h), tk, tk^h, kt, kt^h/ from slow to fast speech rate were test tokens. The test session included 7 repetitions of all the 100 tokens (6 + 28 + 66 = 100). Tokens were randomized in each repetition. Each trial started with the stimulus followed by a 4 second pause before the feedback which was only given to listeners in the training session. There was also a 4 second pause between trials. Every ten tokens formed a block. There was a 10 second pause between blocks.

3.2.1.2 Subjects

Thirty native Taiwanese speakers (11 male and 19 female) who were born and grew up in Taiwan participated in the perception test. (They all also speak Mandarin as a second language.) The listeners were randomly divided into two groups. One group was trained to respond 'yes' to syllables ending with /t/, and the other group was trained to respond 'yes' to syllables ending with /k/.

3.2.1.3 Procedure

The same procedure as in Experiment 3 of Case Study 1 was used. Listeners were run individually in a quiet room wearing headphones. Each trial was answered by circling 'yes' or 'no' on answer sheets. Listeners were trained until they learn the category /t/ or /k/--as measured by correctly answering ten consecutive trials. In the test session, after listening to each stimulus, subjects are asked to answer "yes" or "no" without feedback to indicate whether the stimulus contained the target category.

3.2.2 Results

Unlike the categorization of the Mandarin sandhi tone, the categorization of Taiwanese coda /t/ and /k/ never reach one hundred percent correct. Listeners had some difficulty in recognizing each coda stop consonant as a distinct category. The categorization of the coda /t/ in **tk** and **kt** was the most difficult. The mean proportions of /t/ categorization for **tk** and **kt** were near chance (0.526 and 0.414).

Figure 22 shows the proportion of /t/ categorizations for each consonant group with stimuli arranged according to the speaker's intended rate from first (slowest) production to last (fastest) production. The categorization of /t/ and /k/ varied depending on consonant group and stimulus rate from the slowest production to the fastest production. A two-way ANOVA (consonant group x stimulus rate) showed that the consonant group had an effect on the proportion of /t/ categorizations, [$F(2, 957) = 97.649$, $p < 0.01$]. A Tukey post-hoc test showed that the proportion of /t/ categorizations for **kt** was lower than that for **tk** which in turn was lower than for **tt**. The categorization of the coda /t/ in **tk** was affected by the following onset **k**. The effect of stimulus rate was also significant, [$F(10, 957) = 2.386$, $p < 0.01$]. Patterns of categorization for different stimulus rates were different. The interaction of consonant group and stimulus rate was significant, [$F(20, 957) = 1.746$, $p = .022$]. The number of /t/ categorizations for **tt** and **tk** decreased gradually along the stimulus rate, while the number for **kt** showed an increasing trend.

Figure 23 shows the proportion of /k/ categorizations for each consonant group and stimulus rate. The effect of consonant group was significant, $F(2, 957) = 16.892$, $p = 0.000$]. A Tukey post-hoc test indicated that **tk** was categorized as /k/ less often than either **kk** or **kt**, but the number of /k/ categorizations was not significantly different for these two. The effect of stimulus rate was also significant, [$F(10, 957) = 4.79$, $p < 0.01$]. There was no interaction between consonant group and stimulus rate.

The proportion of /t/ or /k/ categorizations was regressed against the speech rate measurement (the duration of the first vowel) and the articulatory measurements taken in Experiment 1. Table 15 shows the results of the regression for correlations of /t/ and /k/ categorizations with speech rate measurement for each consonant group. As speech rate increased (shorter duration of the first vowel), the number of /t/ categorizations for **tt** decreased . The number of /t/ categorizations for **kt** increased as speech rate increased. On the other hand, the number of /k/ categorizations for both **kk** and **tk** increased as speech rate increased. However, the effect was stronger for **tk**. percentage of /t/ categorizations for **tk** and /k/ categorizations for **kt** did not significantly correlated with the speech rate measurement.

The results of regression analyses of /t/ and /k/ categorization rates against the EPG measurements showed that the EPG measurements did not correlate with the /t/ and /k/ categorization rates very well. Only the /t/ categorization rate for **tt** and the /k/ categorization rate for **tk** could be predicted from a few EPG measurements. For **tt**, the /t/ categorization rate was significantly correlated with the duration of dental closing gesture, [adjusted $R^2 = 0.149$, standard error = 0.121, standardized regression coefficient = 0.436 ($T = 2.165$, $p < 0.05$)], and percent of electrodes contacted at the maximum contact, [adjusted $R^2 = 0.383$, standard error = 0.103, standardized regression coefficient = 0.643 ($T = 3.75$, $p < 0.01$)]. The longer the duration of the dental closing gesture and the more electrodes contacted at the maximum contact, the more tokens of /t/ were categorized for **tt**. The low correlation between the /t/ categorization rate for **tk** and the EPG measurements might be due to the large number of tokens with no contact for the dental gesture. Tokens with no contact were excluded in the regression analyses.

The /k/ categorization rate for **tk** was significantly correlated with duration of the velar closing gesture, [adjusted $R^2 = 0.203$, standard error = 0.089, standardized regression coefficient = -0.491 ($T = -2.522$, $p < 0.05$)], peak latency, [adjusted $R^2 = 0.66$, standard error = 0.082, standardized regression coefficient = -0.853 ($T = -3.269$, $p < 0.05$)], and offset latency, [adjusted $R^2 = 0.843$, standard error = 0.055, standardized regression coefficient = -0.935 ($T = -5.274$, $p < 0.01$)]. The /k/ categorization rate for **tk** was higher for tokens which were spoken in faster speech rate and so had shorter duration of the velar closing gesture. When the peak latency and offset latency of the velar gesture relative to the dental gesture in **tk** was shorter, the /k/ categorization rate was higher.

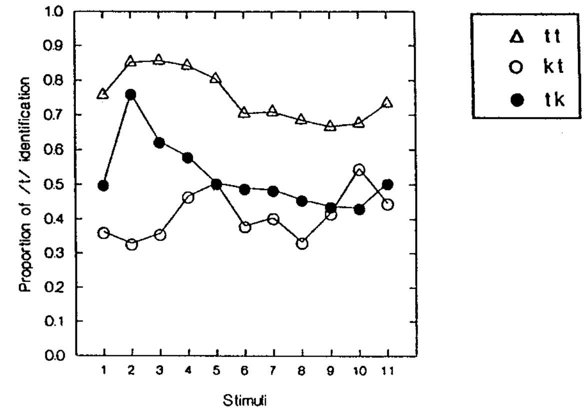


Figure 22. The proportion of /t/ categorization for **kt**, **tk** and **tt** with stimuli arranged according to the speaker's intended rate from first (slowest) production to last (fastest) production

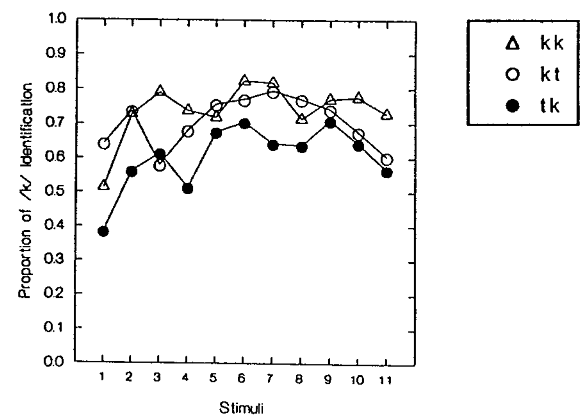


Figure 23. The proportion of /k/ categorization for **kt**, **tk** and **tt** with stimuli arranged according to the speaker's intended rate from first (slowest) production to last (fastest) production

Table 15. Adjusted R^2 with standard error of estimate and standardized coefficients for /t/ and /k/ categorization rates of each consonant group regressed against speech rate.

category	consonant group	adjusted R^2 (standard error)	standardized coefficient
/t/	tt	0.188 (0.118)	0.476 (T = 2.42, p < 0.05)
	kt	0.192 (0.081)	-0.48 (T = -2.449, p < 0.05)
	tk	0.000 (0.118)	0.038 (T = 0.168, p = 0.868)
/k/	kk	0.221 (0.099)	-0.508 (T = -2.64, p < 0.05)
	kt	0.077 (0.081)	-0.351 (T = -1.636, p = 0.118)
	tk	0.394 (0.078)	-0.65 (T = -3.829, p < 0.01)

3.2.3 Discussion

Different patterns of gestural coarticulation in **tk** and **kt** contributed to different patterns of categorization of the syllable-final /t/ and /k/. Because the dental gesture was overlapped more by the following velar gesture, the mean /t/ categorization rate for **tk** was lower than that for **tt**. **tk** was also categorized as /k/ in more than fifty percent of the responses on average. As speech rate increased, the **tk** sequence was categorized as /t/ less frequently, and as /k/ more frequently. By contrast, the /k/ categorization rate for **kt** was about the same as for **kk**. **kt** was categorized as /t/ in less than fifty percent of the responses. The difference in the categorization between **tk** and **kt** may be attributed to the difference between them in timing and overlapping of the gestures. The onset latency of the velar gesture relative to the preceding dental gesture was earlier than that of the dental gesture relative to the preceding velar gesture. The dental gesture was also more overlapped by the following gesture than the velar. In fact , the dental gesture was "deleted" frequently in the coda position. Therefore, the **kt** sequence was categorized as

/k/ most of the time, while the **tk** sequence was categorized as /t/ or /k/ depending on degree of gestural overlap.

The concept of coda /t/ and /k/ categories seems to be difficult for Taiwanese speakers. Listeners' performance in the categorization of these coda stop consonants was not very good comparing to Jaeger's (1986) previous study on the categorization of English /k/ using concept formation. One possible reason for the difficulty in forming these Taiwanese stop categories was that there are no clear acoustic cues for the release of stop closure in Taiwanese coda stop consonants. Linguistically naive listeners can hardly separate the stop consonant from the preceding vowel in perception. They are very likely to treat the coda stop consonant as a part of the preceding vowel. Another possible reason is that the Taiwanese language does not have a native writing system. Jaeger (1986) in a study of the /k/ categorization of the English [sk] sequence suggested that English listeners' categorization for the [sk] sequence might be a phonological categorization based on the English orthographic system. Therefore, they categorized the [k] in the word 'school' as /k/. The fact that Taiwanese speakers do not know Taiwanese in terms of the writing system makes it difficult for Taiwanese speakers to recognize the 'unreleased' coda stop consonants.

CHAPTER IV

GENERAL DISCUSSION AND CONCLUSION

At first glance, the tone sandhi study seems clearly to support the idea shared by some researchers (Keating, 1984; 1990; Ladefoged, 1980; 1984; Pierrehumbert, 1990) that phonological representation and phonetic representation are distinct. If phonological representation describes distinctive categories at the level of overall contrasts and categorical alternations within a language, whereas phonetic representation describes details of speech utterances. The tone sandhi rule fits two characteristics of a typical phonological rule. Case study 1 showed that the tone sandhi rule applies wherever the environmental condition described in the rule is met and is not affected by speech rate or speaker variation. If phonological rules are "syntactic", as suggested by Pierrehumbert (1990). a "syntactic" rule in which discrete elements are used can be constructed to describe this Mandarin tone sandhi , i.e. tone 3 —> tone 2 / __ tone 3. However, the rule does not describe the consistent acoustic difference between the sandhi tone and tone 2: the sandhi tone had a slightly lower F0 value than tone 2. These details of utterances require parametric phonetic representations which can specify F0 target levels and F0 transitions.

The categorization of the sandhi tone also indicated that phonological representation is a cognitive representation stored in the mind of speakers and listeners. The sandhi

tone was categorized as its underlying form—tone 3 in the tone sandhi environment—, although it was changed into a surface form which is perceptually indistinguishable from tone 2. This contradicts what is assumed in Ladefoged's model (1980; 1984): phonological representation is not likely to be a mental representation.

The derivational relationship between categorical representation and parametric representation argued for in Keating's model (1990) cannot explain the acoustic difference between the sandhi tone and tone 2. According to her model, the phonological representation (tone 3 in a following tone 3 context) will generate a categorical representation (tone 2) which will generate a parametric phonetic representation (a rising F0 contour appropriate for tone 2). Because the categorical representation of the sandhi tone is the same as that of the underlying tone 2, the sandhi tone should have the same parametric phonetic representation as the underlying tone 2. However, the sandhi tone is actually slightly different from the underlying tone 2 in its measured acoustic parametric representation.

The categorical phonetic representation in Keating's (1990) model is difficult to fit to the representation of tone sandhi. According to this model, phonological tonal features will be implemented as phonetic categories chosen from a universally-specified set of categories. That is, the phonetic categories are separated by absolute boundaries. However, distinctive tones are necessarily relational categories. [+High] tone is higher in pitch than [-High] tone other things being equal, but other things rarely are equal. Each language chooses its own way to divide the pitch range into contrastive tones. Moreover, even after accounting for "speaker normalization", tonal context and more global pitch range variation affects the tone target height, in language specific ways. For example, [L] of tone 3 (L) is much lower than [L] of tone 2 (LH). Each tone is defined relative to other tones in each language.

If phonological representation as well as phonetic representation are articulatory representations as suggested by Browman and Goldstein (1989), then one needs to know how to represent tones in gestural scores describing glottal activities. If tone can be represented using articulatory gestures, then what are tone gestures? The location, degree, stiffness and shape of the constriction at the vocal fold will need to be specified in the gestures. For tone sandhi, the underlying tone 3 and the sandhi tone may be represented using different gestural scores. However, the concept formation study suggested that at a more abstract level, all tone 3's including those in the sandhi and non-sandhi environments were the same. This seems to be inconsistent with gestural phonology in which phonetic representation and phonological representation are not distinct.

The other case study poses its own different problems. A discrete alternation account which might explain Mandarin tone sandhi does not necessarily explain place assimilation in Taiwanese. Place assimilation between Taiwanese stop consonants cannot be described by traditional phonological rules for the following reasons. First, we interpret the "no-contact" cases in terms of categorical gesture deletion and hence assimilation, place assimilation did not always occur. It was conditioned by speech rate and speaker variation. Place assimilation occurred often in normal or fast speech, but rarely in slow and careful speech except for one speaker. Secondly, the probability for place assimilation to occur was different for consonants with different places of articulation and independently varied for different syllable positions, so it cannot be stated by only one general 'variable rule'. The dental gesture was deleted more frequently than the velar gesture. Gestures were deleted more frequently at the coda position than at the syllable-initial position. Thirdly, if place assimilation did occur, it was not always a categorical or discrete change. There was quantitative variation within

the phonetic implementation of the place assimilation. Categorical change, such as a dental stop becomes a velar stop, only occurred in some cases produced by speakers in normal or fast speech rate, rarely in slow speech rate. In many cases, the target of the contact made by the tongue for a stop consonant was shifted by the abutting stop consonant with a different place of articulation, but the shift of position was not big enough to change the place of articulation classification of the consonants. As speech rate increases, the dental contact moved further back and the velar contact was fronted. Therefore, if traditional phonologically approach using distinctive features would be used to describe the contextual variation of Taiwanese stop consonants, multiple phonological rules specified with appropriate probability distributions for different consonants and different prosodic positions will need to be stated. However, this approach not only cannot include cases with a non-categorical change, but also misses the generalization about rate influences and so on. It fails as badly as when it applies to Glottalization in English (Pierrehumbert, 1994).

The stop consonant place coarticulation found in the current study can be well described by Browman and Goldstein's (1989; 1990a; 1990b; 1992) model of phonological and phonetic representation using articulatory gestures. In their model, both phonological representation and phonetic representation are rooted in the coordinative structure of articulatory gestures. Gestural scores which describe the motion of articulators for each gesture as well as the timing of gestures can represent different degrees of place assimilation. The non-categorical change of place features can be represented by different patterns of gestural overlapping. The more one gesture was overlapped by another gesture, the less reliable the identification of the gesture was. As more of the dental gesture of /t/ was overlapped by the following velar gesture of /k/, the less frequent it was categorized as /t/. One gesture may be overlapped by the following gesture to the extent that it was hidden acoustically. Extreme cases where no contact was detected for a gesture might be attributed to be the endpoint of a continuum of gestural reduction instead of deletion of a gesture (Browman & Goldstein, 1992). However, this articulation-based model does not predict the differences between two gesture types found in the current coarticulation study. The dental gesture was more subject to deletion than the velar gesture. The latency of the velar gesture relative to the preceding gesture was shorter than that of the dental gesture.

In conclusion, as suggested by Pierrehumbert (1994), the description of the sound system of language is a complicated issue. It cannot be fully described either by models which make a hard and fast separation between (discrete categorical) phonological representation and (continuous) phonetic representation, or by models which collapse these two representations into one articulatory representation. More sophisticated studies of speech categorization need to be done before a better model can be constructed. Issues concerning the psychological reality of phonological categories and how categorization relates to the physical and psychophysical parameters of speech production and perception especially need to be explored more in depth to account for language sound system as a tool of communication.

REFERENCES

Blicher, D., Diehl, R. & Cohen, L. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics* 18, 37-49.

Browman, C. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201-251.

Browman, C. & Goldstein, L. (1990a). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18, 299-320.

Browman, C. & Goldstein, L. (1990b). Tiers in articulatory phonology, with some implications for casual speech. In K. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I: Between Grammar and Physics of Speech* (pp. 341-376). Cambridge, U.K.: Cambridge University Press.

Browman, C. & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica* 49, 155-180.

Byrd, D. (1994). *Articulatory Timing in English Consonant Sequences.* Ph.D. Dissertation, UCLA.

Chao, Y.R. (1948). *Mandarin Primer.* Cambridge, MA: Harvard University Press.

Chao, Y.R. (1968). *A Grammar of Spoken Chinese.* Berkeley, CA: University of California Press.

Disner, S. (1978). Vowels in Germanic Languages. *UCLA Working Papers in Phonetics* 40.

Farnetani, E. & Busa, M. G. (1994). Consonant-to-consonant interactions in clusters: categorical or continuous processes? *Consiglio Nazionale delle Ricerche Quandemi del centro di Studio per le Ricerche di Fonetica* Volume XIII, 220-245.

Grier, J. (1971). Nonparametric indexes for sensitivity and bias: computing formulas. *Psychological Bulletin* 75(6): 424-429.

Goldstein, L (1989). On the domain of the quantal theory. *Journal of Phonetics* 17, 91-97.

Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica* 33, 353-367.

Hume, E. (1992). *Front Vowels, Coronal Consonants and Their Interaction in Nonlinear Phonology.* Ph.D. Dissertation, Cornell University.

Jaeger, J. (1986). Concept formation as a tool for linguistic research. In J. Ohala & J. Jaeger (eds.), *Experimental Phonology* (pp. 211-237). Orlando: Academic Press.

Jannedy, S. (1994). Rate effects on German unstressed Syllables. *OSU Working Papers in Linguistics: Papers from the Linguistics Laboratory* 44, 105-124.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60(2), 286-319.

Keating, P. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics* 18 321-334.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language* 56 (3), 485-502.

Ladefoged, P. (1982). *A Course in Phonetics.* New York: Harcourt Brace Jovanovich.

Ladefoged, P. (1984). Out of chaos comes order: Physical, biological, and structural patterns in phonetics. In M. P. R. Van den Broecke & A. Cohen (eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences* (pp. 83-95). Dordrecht: Foris Publications.

Ladefoged, P. (1990). On dividing phonetics and phonology: comments on the papers by Clements and by Browman and Goldstein. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, (pp. 398-405). Cambridge, U.K.: Cambridge University Press.

Ladefoged, P. & Wu, Z. (1984). Places of articulation: an investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12, 267-278.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of Acoustical Society of America* 35, 1773-1781.

McCarthy, J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry* 20, 71-99.

Munhall, K. & Löfqvist, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics*, 20, 93-110.

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In G. J. Docherty & D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 261-289). Cambridge, U.K.: Cambridge University Press.

Ohala, J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 258-275). Cambridge, U.K.: Cambridge University Press.

Ohala, J. (1993). Coarticulation and phonology. *Language and Speech* 36(2), 155-170.

Öhman, S. E. (1966). Coarticulation in VCV utterances: spectrographic Measurements. *Journal of Acoustical Society of America* 39, 151-168.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics* 18, 375-394.

Pierrehumbert, J. (1994). Knowledge of variation. *CLS* 30, 1-25.

Stevens, K. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (eds), *Human Communication: A Unified View* (pp. 51-66). New York: McGraw-Hill.

Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics* 17, 3-45.

Stevens, K., Keyser, S. & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In *Invariance and Variability in Speech Processes* (J. S. Perkell & D. H. Klatt, eds.), PP. 426-449. Hillsdale NJ: Erlbaum.

Wang, W. & Li, K-P (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research* 10, 629-636.

Zee, E. (1980). A spectrographic investigation of Mandarin tone Sandhi. *UCLA Working Papers in Phonetics* 49, 98-116.