UCLA

Working Papers in Phonetics

Title

WPP, No. 86: Articulatory Timing in English Consonant Sequences

Permalink

<https://escholarship.org/uc/item/4sr0s3jw>

Author Byrd, Dani

Publication Date 1994-05-01

Articulatory Timing in **English Consonant Sequences**

Dani Byrd

UCLA WORKING PAPERS IN PHONETICS 86. MAY 1994

UCLA WORKING PAPERS IN PHONETICS 86

Articulatory Timing in English Consonant Sequences

 by

Dani Byrd

May, 1994

 \odot Copyright by

 $\bar{\mathcal{A}}$

Dani Margaret Byrd

1994

TABLE OF CONTENTS

 $\sim 10^{-10}$

LIST OF FIGURES

 $\mathcal{L}^{\text{max}}_{\text{max}}$

 $\frac{1}{2} \left(\frac{1}{2} \right)$, $\frac{1}{2} \left(\frac{1}{2} \right)$

LIST OF TABLES

 \sim \sim

ACKNOWLEDGMENTS

I extend my deepest gratitude to my doctoral committee--Patricia Keating, Peter Ladefoged, Ian Maddieson, and Donca Steriade whose rigorous scholarship I admire and will always try to emulate. Their teaching has been of fundamental importance to me from the inception of this thesis work, and their contributions have improved it greatly. In particular, one couldn't ask for a more supportive and enthusiastic advisor than Pat Keating. Cathe Browman and Louis Goldstein have my warmest respect and regards; without their seminal work in linguistic phonetics this dissertation would not have been undertaken. I am especially grateful to Louis Goldstein for introducing me to phonetics and fostering an unadulterated joy in the wonder of it all. Special thanks to Edward Flemming who contributed more to my education than he knows. His clarity of thought and generosity of intellect has made this thesis a significantly better piece of work. While I may not name them, the five speakers who participated in this research have my sincerest gratitude. I was so fortunate to have these participants who never wavered in their commitment from beginning to end. For comments, suggestions, and education, I am additionally grateful to Abeer Alwan, Ted Bell, James Byrd, Court Crowther, Bruce Hayes, Keith Johnson, Abigail Kaun, Dan Silverman, Caroline Smith and members of the UCLA Phonetics Laboratory. Many thanks to Andy Mueller for superb programming and great good will, to Cheng Cheng Saw for data processing for Chapter Six and for the graphics in Appendix A, to Dr. Stephen Prince and Sue Tom for making the palate casts and their generous contribution to a research effort, to John Crump, Michael Mulleady, and Rich Sico at Kay Elemetrics for generous technical support, and to Henry Teheranizadeh for technical assistance. Thanks to Kathy Byrd and Xandi Byrd for keeping my "real life" in order while academia reigned; and, especially, to my parents for gifting me with books and a Yale education. Special thanks and all my love to Oliver Foellmer whose love and patience enabled me to work with a good heart and with whom many of these thoughts were nurtured.

This research was supported by an NSF Graduate Fellowship, by an NSF Dissertation Improvement Grant #DBS 9213604 to myself and Pat Keating, and by the Department of Linguistics at UCLA. I am very grateful to the National Science Foundation and the UCLA Department of Linguistics for their financial assistance through the past four years.

ABSTRACT

Articulatory Timing in English Consonant Sequences

by

Dani Byrd

Doctor of Philosophy in Linguistics University of California, Los Angeles, 1994

Professor Patricia A. Keating, Chair

One of the most significant challenges in the study of speech production is to acquire a theoretical understanding of how speakers coordinate articulatory movements. Four experiments were conducted using electropalatography (EPG) to determine how articulatory timing in English consonant sequences is affected by consonant place and manner, by syllabic structure, by the number of consonants in the sequence, and by speech rate.

The results show that coda stops have less lingua-palatal contact than onsets. A tongue tip consonant is more overlapped by a following tongue body consonant than a tongue body consonant is by a following tongue tip consonant. Stop-stop sequences exhibit greater overlap than stop-fricative sequences. We find that an onset cluster is less overlapped and less variable in timing than coda clusters and heterosyllabic sequences. In considering the coordination of long consonant sequences and neighboring vowels, our findings suggest that the near edge of the consonant sequence is relevant in timing relationships across a word boundary. Within-word relationships appear defined by the C-Center of the tautosyllabic consonants. However, speakers and sequences differ in what timing arrangement is most stable. Lastly, we find that as speech rate increases, consonants spanning a word boundary shorten and undergo a relatively linear increase in articulatory overlap. However, rate has only a minimal effect on [d#g] which is almost completely overlapped at all rates.

We adopt the general approach of Articulatory Phonology to intergestural timing which uses phasing relationships to coordinate articulatory gestures. However, we propose a model of speech timing in which specific overlap between articulatory units will vary as a function of linguistic and non-linguistic factors. We argue that timing relationships are constrained languagespecifically to occur within permissible PHASE WINDOWS. Influencers which differ from utterance to utterance weight a PHASE WINDOW probabilistically, determining where in the range of permissible overlap relationships a token is likely to be realized. It is additionally suggested that the percept and functionality of what has traditionally been called a segment results from a characteristic stable timing, *i.e.* a narrow PHASE WINDOW which is lexically specified.

CHAPTER ONE: INTRODUCTION

1.0 **BACKGROUND**

Currently, one of the most significant challenges in the study of speech production is to gain a theoretical understanding of how speakers coordinate articulatory movements. The goal of this effort is to uncover principles of coordination rather than simply patterns of coordination. Many linguistic and extralinguistic factors are known to affect timing but rarely have very many of these effects been studied concurrently using a single set of speakers and experimental techniques. Here, I undertake such an investigation in considering the timing of English consonants in sequence.

Consonant sequences are of special interest in creating models of speech production, as many demands are concurrently placed on an individual articulatory structure, the tongue (in the case of non-labial consonants). Because the tongue must execute these demands in a short period of time, typically not every consonant is discretely articulated. This study investigates how these competing demands are resolved. Although consonant sequences are of special importance in understanding articulatory organization, very little articulatory data has been collected and published on such sequences. Most research on coproduction has considered movements involved in producing single intervocalic consonants. However, any assumption that consonant sequences have less extreme coarticulation than do adjacent vowels and consonants should be made cautiously and evaluated empirically. Much of the data in the recent work on consonant cluster production has been opportunistically collected, focusing on cases known to be of interest, rather than resulting from systematic data collection experiments. As a result, it is difficult to argue that the cases examined are characteristic of more comprehensive patterns of articulation.

Studies of acoustic durations in speech have found that many factors to influence duration in speech. Klatt's classic 1970 overview of such effects identified the following among factors shown to have an influence on the durational structure of the sentence: physical state, speaking rate, emphasis, novelty, segment, stress, neighboring segments, and word, phrase and discourse position. The acoustic effects observed by him and others must, of course, have physiological bases. The aims of the present study are similar in certain respects, but in the domain of speech production rather than acoustics. We wish to obtain a better understanding of the nature of certain influences on articulatory timing.

Considering the myriad of influences on speech timing, it is not surprising that speech timing should be highly variable and subject to complex interactions of these variables. We attempt to uncover how certain linguistic and extra-linguistic variables affects articulatory timing in consonant sequences. This study undertakes several investigations using a single experimental technique-electropalatography--and a consistent group of speakers. Specifically, I address the

following hypotheses. 1) The coproduction of consonantal gestures varies as a function of their articulatory places and manners. 2) The placement of syllable boundaries affects timing and displacement in consonant sequences. 3) As the number of consonants in a sequence is varied, the articulatory organization of the sequence with repect to its neighboring vowels changes. 4) Speakers adjust consonant sequence timing as a function of their speaking rate. As a whole, these hypotheses can be taken as a prediction that different types of factors-gestural, prosodic, and extra-linguistic--affect speech coordination and that they may interact in influencing timing. The following experiments provide significant data with which various models of timing in speech production can be further explicated to more accurately mirror articulatory coproduction.

Research in the area of speech production has long been hampered by the difficulty of obtaining accurate, quantified data on the movement of the articulators. Instrumental information on the movement of the tongue has been particularly difficult to collect. However, such data is of great importance as the tongue is the major contributor to vocal tract shape, and is involved in all vowels and almost all consonants in language. "The tongue is a boneless, jointless structure, yet it can elevate, depress, widen, narrow, extend, and retract. It also can create leverage, torsion, a midsagittal groove, a midsagittal arch, and move differentially, both laterally-to-medially and left-to-right" (Stone, 1991). For many sounds the tongue functions as the active (moving) articulator with the hard palate as the passive (non-moving) articulator. The hard palate is generally ignored in studies of speech production because it does not move; however, it is important in understanding tongue dynamics. As Stone explains, "The palate provides the tongue with a solid base of contact for sensory feedback, for light support during rapid or complex movements, and for resistance. When the tongue tip pushes against the palate, various tongue shapes and movements are facilitated." The approximation of the tongue to the palate is the immediate cause of many of the acoustic characterics associated with a large number of consonants. For this reason, understanding the mechanism and coordination of this approximation of the tongue to the palate is important in studying consonant production.

Information about tongue movement is crucial to understanding the variable nature of timing for consonant sequences, as most sequences include lingual consonants. Real-time three dimensional recording of tongue movement is not currently possible, but there are a number of ways to collect particular data about tongue movement. Ultrasound provides two dimensional images of the tongue surface through time at a rate of about 30 Hz (Stone, 1991) but is unable to penetrate bone and air (Foldvik, et al., 1991). Cinegraphic X-ray techniques and X-ray computerized tomography of articulatory movements have the harmful side effects of radiation. Magnetic resonance imaging (MRI) shows all the articulators clearly and provides threedimensional images but currently has poor time-resolution capabilities, although this is improving. MRI requires long acquisition times and requires phonetically trained subjects who can freeze their articulation of a particular sound (Foldvik, et al., 1991), making it impractical for large amounts of data collection and quantitative analysis. Data collection using X-ray microbeam facilities provides high quality information about articulator movements in the

midsagittal plane. However, the X-ray microbeam procedure provides no details about contact patterns, such as the magnitude or location of contact, or about movement outside the midsagittal plane. Finally, the electromagnetic articulograph or magnetometer technology is relatively new and produces data approximating microbeam data. All of these systems require specialized technical, and sometimes biomedical, expertise. Additionally, it is often difficult to acquire quiet acoustic data or to go back and re-collect slightly different data when necessary.

By comparison, dynamic electropalatography (EPG) (also known as dynamic palatography or palatometry) is a system for recording information about the tongue's contact with the hard palate over time. It is relatively inexpensive and technologically accessible. EPG is safe, collects movement data outside the midsagittal plane, provides spatial information on the shape of constriction contact, and allows the collection of reasonable acoustic data. Multiple sessions with a subject are possible and replicable. For these reasons, this method was chosen to collect articulatory data in the experiments below.

The majority of (non-clinical) phonetic research using EPG has concentrated on the extent of assimilations or the effect of vowels on a single consonant's place of articulation. Thus, much work has discussed spatial patterns of contact at a particular moment, without exploiting the dynamic capability of EPG. Conversely, instrumental techniques frequently used to explore questions of articulatory timing such as X-ray microbeam tracking of articulators are less conducive to examining certain details including tongue behavior outside the midsagittal plane, location of palatal contact, or tissue compression characteristics of tongue against palate. The research reported here uses standard techniques of EPG data collection, but our analysis focuses on dynamic patterns rather than static ones and considers consonant sequences rather than single intervocalic consonants. In addition to examining the temporal coordination of consonants in sequence, this study also reports on the spatial magnitude and duration linguapalatal contact for individual consonants. Works by Barry (1991), Gay (1981), Hardcastle, Gibbon, & Nicolaidis (1991), Hardcastle and Roach (1979), Marchal (1988), and Nolan (1992) serve as examples for the temporal analysis of EPG data.

Studies of speech production must address the development of theories of multimovement control and articulatory coordination (see Abbs, Gracco, and Cole, 1984). The linguist engaged in such an effort will rely in part on empirical data to determine the permissible timing relationships for certain movement patterns and the influence of linguistic variables on this coordination. The research described here will yield a detailed account of coproduction in selected sequences of English stops and fricatives. To produce a systematic study of the articulation of consonant sequences, a carefully selected corpus of articulatory data must be acquired. This work investigates the production of English alveolar and velar stops and alveolar fricatives in sequences. The consonant sequences studied are heterorganic or geminate and range from two to four consonants in length.

One must consider here precisely what phenomena are being evaluated in our experiments. We are interested in the relative coordination of two (or more) phonetic events. A point in one event may be measured relative to some other event, or two points in a single event, such as onset and offset, may be compared. Thus we are concerned here with both the coordination of linguapalatal contact between consonants and the duration of linguapalatal contact for a particular consonant. The time interval betweeen two articulatory events is often called the latency. Latency and duration may be evaluated relative to some other measure, in which case they are often evaluated as a percentage value.

Work has been published on the nature of coarticulation in consonant sequences, but the coverage of this work is not complete. In English, phonetic observation has revealed that the closure for the first consonant in a cluster generally is not released until after the closure for the second is formed (Jones, 1956; Catford, 1977; Hardcastle and Roach, 1979; and Marchal, 1988 (for French)). Catford (1977) estimates the degree of overlap between two adjacent consonants as being between 29% and 45% of the total sequence duration. Barry (1985), Nolan (1992), Browman and Goldstein (1990b), and others have shown that consonants in clusters that sound as if they have assimilated in place of articulation, or deleted altogether, are often in fact still articulated at their original point of articulation. The closure movements for these consonants overlap and may also weaken. The perceptual consequences of assimilation or deletion will depend on such factors as the combination of consonants involved, the speech rate, and the speaker's degree of casualness. Byrd (1992) used articulatory synthesis to show that a completely articulated alveolar stop is not perceived by listeners if it is substantially overlapped with a velar stop. This and other work indicate that perceptual and acoustic approaches to understanding these speech events may not always be revealing when consonant clusters are concerned. Detailed information is needed on articulatory movement to determine how such sequences are coordinated.

1.1 THEORETICAL FRAMEWORKS

Coproduction of consonant clusters has been analyzed differently in various theoretical frameworks. Articulatory Phonology is an innovative theory of speech production developed by Browman and Goldstein (1986, 1988, 1989, 1990a, 1990b, 1991, 1992, in press). In this model, dynamically specified articulatory gestures are the units of phonological representation. Under this account, the phonological primitives, *i.e.*, gestures, include temporal information yielding the duration of a unit. There is no external timing mechanism, i.e. no clock dictating the duration of each unit; the units are self-timed. Such a model is said to be an intrinsic timing or relative time model. Other intrinsic timing models include those of Fowler (1977, 1980), Bell-Berti and Harris (1981), and Saltzman and Munhall (1989). Articulatory Phonology posits no difference between the properties of canonical forms of phonological units and those units as instantiated in an articulatory plan (cf. Fowler, 1980). This model accounts for contextual variation by the simultaneous activation of different competing and non-competing gestures (Saltzman and

Munhall, 1989). In Articulatory Phonology, an utterance is described not only by what gestures are involved but by how the gestures are coordinated with each other. This theory captures coproduction by allowing gestures to overlap in time (Browman and Goldstein, 1989 and 1990b). The acoustic consequences of the coproduced units reflect their combined influence on the vocal tract.

In contrast, other theories assume a surface representation having nonoverlapping time slots superordinate to atemporal phonological features. In such models, the temporal characteristics of the segment are not part of its phonological representation; the units are externally timed. These are called extrinsic timing or absolute time models (e.g. Lindblom, 1983, Lindblom, Lubker, Gay, Lyberg, Branderud, and Holmgren, 1987). Many theorists within such frameworks propose multiply-linked phonological features and/or phonological underspecification (the lack of a feature specification) to account for coarticulatory processes. For example, Keating's (1990b) Window Model of coarticulation derives variation in the realization of segments by rules that alter the phonetic values of targets projected by that segment's distinctive features, and by acoustic/articulatory interpolation functions that operate across intervening segments not specified for a particular feature. Some extrinsic timing theories attribute articulatory overlap to the spreading without delinking of autosegments in the phonological representation, and weakening or target undershoot to rules in the phonetic implementation module of the grammar (Hayes, 1992; Daniloff and Hammarberg, 1973). The consequence of these operations is a smoothing of the articulatory transitions between adjacent segments at the expense of maintaining the canonical forms (Fowler, 1980). A further consequence is that speakers never (or rarely) actualize canonical forms (Fowler, 1980).

In this work, the focus will be on elaborating an intrinsic timing approach, rather than comparing theories of linguistic timing. As a starting point we adopt Browman and Goldstein's Articulatory Phonology framework because it offers an explicit approach to characterizing speech timing. However, certain concepts developed in a targets-and-interpolation approach will be relevant to our later discussions. Therefore, brief descriptions of these two approaches to linguistic timing--one intrinsic and one extrinsic--are presented below. It should be kept in mind that extrinsic versus intrinsic timing and direct versus mediated realization of primitives are separate issues, although not discussed as such here.

$1.1.1$ **KEATING'S WINDOW MODEL OF COARTICULATION**

Keating's Window Model of coarticulation (Keating, 1990a,b; cf. Manuel, 1987, Cohn, 1990) is a *targets-and-interpolation* approach in which articulatory and/or acoustic targets are projected temporally and spatially by the feature specification of the relevant segments. The interaction of adjacent segments is the output of formal rules of phonetic implementation that translate phonological (featural) autosegments into quantitative physical attributes by interpolating between the projected targets. Many of the commonly recognized overlap effects are attributed primarily to underspecification, offering an approach very different from that of Daniloff *et al.* (1973) and others noted above. Underspecification allows segments or sequences of segments to project no targets for a particular phonetic dimension. Interpolation of phonetic parameters takes place only between targets projected by segments specified for that parameter.

Crucially, in Keating's model a feature will project a target *window* which allows the realization of a featural specification to vary within a specified range (Keating, 1990a). She explains:

...this window is not a mean value with a range around that mean, or any other representation of a basic value and variation around that value. It is an undifferentiated range representing the contextual variability of a feature value. For some segments this window is very narrow, reflecting little contextual variation; for others it is very wide, reflecting extreme contextual variation. Window width thus gives a metric [of] variability. There is no other "target" associated with a segment; the target is no more than this entire contextual range...Windows are determined empirically on the basis of context, but once determined are not themselves contextually varied. That is, a feature value...does not have different widows for different contexts. Information about the possibilities for contextual variation is already built into that one window. (Keating, 1990b, p. 455, 456)

This concept will be fruitful in Chapter Seven when the nature of variability in interarticulator timing is explored, as the postulation of a window at the same time constrains and allows variability.

Since certain concepts from Keating's window model are taken up later, it is appropriate to remark briefly on how timing has been handled within this framework. While a complete account of timing has not been proposed within this model, it appears that timing relationships are conceived of as being specified outside the articulatory system, while information about successive spatial coordinates of an utterance is specified by rule from phonological autosegments. Clearly, linear precedence is a byproduct of the fact that phonetic representation is read off phonological representation in which such precedence relationships are encoded. Cohn (1990) offers one possible analysis of the temporal dimensions of a target. Keating (1990a) implicitly views target windows as having inherent duration. Cohn elaborates that "a feature specification maps to a target that takes up most of the duration [of the segment], since a feature specification is associated with the whole of an abstract segment or timing unit" (1990, p. 99). She suggests that the transition periods between targets can be "dominated" by either of the two adjacent segments depending on a hierarchy of priority among both the specific feature concerned and other co-occurring features. Cohn finds that these timing principles and the mechanisms of spatial windows and underspecification were not sufficient to generate her observed data on nasal coarticulation. She also proposes additional phonetic constraints on segment realization which follow from articulatory, aerodynamic, or perceptual requirements (Cohn, 1990, p. 125). Cohn discusses the example of a voiced oral stop between two [+Nasal] segments that is itself phonologically (and phonetically) unspecified for Nasal. To correctly derive the data, she concludes that a phonetic constraint must be implemented in the form of a [-Nasal] point target temporally anchored to the release of the stop. Other accounts of the temporal anchoring of projected targets to segment-internal points include those proposed by Huffman (1990), Kingston (1990), and Steriade (1993).

$1.1.2$ **BROWMAN AND GOLDSTEIN'S ARTICULATORY PHONOLOGY MODEL**

Fowler (1980) has suggested, "[I] nstead of treating coarticulation as an adjustment of the canonical properties of a segment in acquiescence to its neighbors, it may be viewed as the overlapping production of successive, continuous...segments. Thus feature spreading may be apparent but not actual" $(p.119)$.¹ Browman and Goldstein have undertaken such an approach.

The discussion of the experiments below will use as its framework the Articulatory Phonology model of Browman and Goldstein (Browman and Goldstein, 1986, 1988, 1989, 1990) a,b, 1992, in press). This innovative framework was chosen as a starting point in this thesis because it is one the few linguistic theories which has offered an explicit account of articulatory timing and because it does so within an intrinsic timing approach.

In Articulatory Phonology, articulatory gestures are modeled as an abstract 360°, critically damped, mass-spring oscillatory system. It will be important in the discussion below to understand how intergestural timing occurs in this framework. The coordination of gestures uses a relative phase description (Kelso and Tuller, 1987 cited by Browman and Goldstein, 1990b). In this approach, gestures are coordinated with respect to their dynamical states rather than an external clock (Browman and Goldstein, 1990b). Gestures are phased with one another such that a particular phase angle in one gesture corresponds to a particular phase angle in another gesture. (In implementing phasing, the abstract cycle for a gesture is treated as if it were the cycle of a virtual undamped system with the same stiffness (Browman and Goldstein, in press). The representation of an utterance, called a *gestural score*, must explicitly specify which gestures are phased with respect to each other (Browman and Goldstein, 1990b).

Significantly, Browman and Goldstein (1990a, in press) assume the synchronized phase angles to belong to a limited set of points in a gesture; specifically, the onset (0°) and target (240°) , and perhaps release (290°) . They liken the choice of a few invariant phase relationships to the choice of values for the dynamic parameters of the gestures (Browman and Goldstein, 1990b). In the work to follow, the status of invariant *(i.e.* a highly constrained, consistent set of), stable phasing relationships will be an important topic.

¹Note: in certain intrinsic timing models, 'acquiescence' also occurs among articulators associated with competing gestural primitives; its nature being determined by the (strength of the) activation variables associated with the gesture (see Saltzman and Munhall, 1989)

1.1.2.1 PHASING RULES

Phasing relationships form the basis for implementing timing in Articulatory Phonology. In its simplist incarnation, a phasing rule synchronizing two phase angles specifies the coordination between two gestures. In Articulatory Phonology "syllable-structure is a characteristic pattern of coordination among gestures" (Browman and Goldstein, in press, p.13). In more complex formulations it appears that groups of gestures, syllable-sized and smaller, may also be marshaled into an organization which may in turn be coordinated to another gesture (Browman and Goldstein, 1988). Phasing rules may also coordinate an arithmetically defined abstract index calculated from specific points in a set of contiguous gestures, *i.e.* the C-Center, with a point in another (vocalic) gesture (Browman and Goldstein, 1988).

There is, however, additional information to which phasing rules have access. The phasing rules proposed by Browman and Goldstein (1988 and 1990b) also may refer to consonants and vowels, *i.e.* whether a gesture is on the consonant tier or vowel tier. Also, phasing rules must consider what consonant gestures are *associated* with what vocalic gestures, and to gestural contiguity or lack thereof on a particular tier. For example:

A vocalic gesture and the leftmost consonantal gesture of an associated consonant sequence are phased with respect to each other. An associated consonant sequence is defined as a sequence of gestures on the C tier, all of which are associated with the same vocalic gesture, and all of which are contiguous when projected onto the one-dimensional oral tier. (Browman and Goldstein, 1990b, p. 354).

The general role of association is not yet clearly defined within Articulatory Phonology. Association lines encode precedence relations (overlap) in Browman and Goldstein 1990b; and they "connect gestures that are phased with respect to one another" in Browman and Goldstein, in press (p. 10). In addition to the association of consonants and vowels in the same syllable, a "statement" of ambisyllabicty "applies" which associates a coda consonant to the vowel of a following word (Browman and Goldstein, 1990b). This association procedure then forces the "reapplication" of the phasing rule which phases a vowel to the leftmost preceding *associated* consonant. There are then two types of operations-association and phasing. It is unclear to what degree these operations, *i.e.* association and phasing, occur in the lexicon and outside it. Both operations in both situations appear possible.

Browman and Goldstein (in press, citing Krakow, 1989 and Sproat and Fujimura, 1993) characterize the phasing relationships for lips and velum in nasals, and for tongue body and tongue tip in laterals making crucial reference to the *linking* of two gestures. Browman and Goldstein (in press) say that "[i]n the language of Articulatory Phonology....[1] *consists* of two gestures" (p. 13, emphasis added) and "that syllable-final *linked* gestures are phased so that the wider constriction degree comes earlier, whereas syllable-initially these gestures are phased

roughly synchronously" (p. 14, emphasis added). The relationship between linking and association is unclear.

Thus, in understanding phasing rules in Articulatory Phonology there appear to be two separate issues of interest: 1) what is to be coordinated; and 2) how it is to be coordinated. The work presented here focuses principally on the latter, but touches on the former. Specifically, the role of phasing rules will be an important topic of discussion.

1.2 **INVARIANCE IN TIMING**

The field of phonetics has a great deal invested in the search for articulatory and acoustic invariance. But, while a great deal of knowledge has been obtained in pursuing this research program, identification of invariance in most areas of speech has proven elusive. In recent years, the search for invariance has included hopes for stable acoustic characteristics of contrastive features. In the area of articulatory phonetics, the search for invariance has often proceeded hand-in-hand with the development of theories of intrinsic timing. However, there is no obvious *necessity* for intrinsic timing models to require invariant timing. Gracco states that "[t] he search for invariance has a long and generally unsuccessful history in investigations of speech production with the obvious conclusion that invariance is not a directly observable event..."(1992a, 20). While this may be *one* possible conclusion, it is surely not the only one. The presumed *existence* of invariance must be questioned with respect to each aspect of speech production. The discussion in this thesis will suggest that intersegmental timing is not an invariant in speech production.

Within a single articulatory movement, studies of articulatory kinematics have suggested that the relation of peak velocity to displacement (and in some studies, the relationship of this ratio to duration) is the dynamic *intra*-gestural property which remains stable across variation in linguistic and extra-linguistic contexts. Many such studies have been conducted: Kozhevnikov and Chistovich, 1965; Ohala, Hiki, Hubler, and Harshman, 1968; Mermelstein, 1973; Sussman, MacNeilage, and Hanson, 1973; Kuehn and Moll, 1976; Ostry and Munhall, 1985; Gracco, 1988; Gracco and Abbs, 1989; and Vatikiotis-Bateson and Kelso, 1993 (See also Ostry, Keller and Parush (1983), Munhall, Ostry and Parush (1985), and Kelso, Vatikiotis-Bateson, Saltzman, and Kay (1985); regarding systematic rate and stress effects on this relationship.)

In addition, examples of tight temporal relations between gestures associated with what, traditionally, would be considered a single segment have been found (Munhall, Löfqvist, and Kelso, 1986; Löfqvist and Yoshioka, 1980b, 1981c; Krakow, 1989). Saltzman and Munhall (1989) note that this argues for the existence of a higher multi-gesture unit in speech production. They conceive of implementing such a unit in terms of dynamical coupling.² However, Löfqvist

 2 Structure which is associated with multiple gestures in Articulatory Phonology include nodes on a rhythmic tier (Browman and Goldstein, 1990b) and gestural constellations. A constellation is represented using a gestural score corresponding to a particular utterance in Browman and Goldstein (1989: 201, 211; in press a: 5). Browman and

(1991), like others, found no evidence for invariant temporal intervals at the between segment level, as opposed to the *within* segment level. He argues that this may indicate a difference in gestural cohesion within and across segments.

In research on *inter-gestural timing*, the pursuit of invariance together with the intrinsic timing approach has yielded a conception of stable *(i.e.* invariant) phasing between gestures. Whereas this approach seems to have found some success, as described above, in considering timing between gestures belonging to the same segment or some multi-gesture structure (a question we will return to below), only limited success has been obtained in identifying invariant timing or phase relationships between segments or "constellations". Studies by Tuller, Kelso and Harris (1982) and Tuller and Kelso (1984) examine the articulatory coordination of intervocalic consonants with their adjacent vowels. (See Keller, 1990 for a concise review of this research program.) This work reported stable consonant latencies with respect to the vowel cycle. However, further work (W. Barry, 1983; Benoit, 1986, Munhall, 1985, and Sock and Jah, 1986) indicated that some of this correlation is due to a statistical artifact. Remaining effects reported in Munhall (1985) are suggested by Keller to be due to large speech rate variation affecting all coupled articulatory measures similarly and are not replicated within cells (see Keller, 1990, 1987). Keller (1990) also notes difficulties in replicating the results of the original study (citing Lubker, 1986; Nittrouer, Munhall, Kelso, Tuller, and Harris, 1988).

A similar effort to uncover invariant intergestural phasing (Nittrouer, Munhall, Kelso, Tuller, and Harris, 1988) concludes that while interarticulator timing may well be controlled in terms of phase relations, the specific amount of overlap may vary continuously as a function of linguistic and non-linguistic factors. This is a conception which we will develop here. Specifically, Nittrouer et al. state:

In contrast to the findings of Kelso *et al.* (1986), we found no support for the notion that the relative phasing of jaw vowel gestures and upper lip consonant gestures are stable across manipulations in linguistic and nonlinguistic factors. In fact, the evidence from the present experiment suggests that the intersegmental organization of gestures is a function of the utterance being produced. In other words, the phase relations between articulatory gestures used in the production of adjacent segments varies [sic] systematically based on linguistic and nonlinguistic structure, which includes speaking rate, stress pattern, syllable structure, and consonant identity. (Nittrouer, et al., 1988, p. 1659); emphasis added

Goldstein (1986) describe "the phonological structure of a lexical item as a 'constellation' of gestures, that is, a stable organization among gestures," (see also Browman and Goldstein 1990b). In other references a constellation is "syllable-sized" and associated with a stress node on the rhythmic tier, e.g. Browman and Goldstein (1990b, p. 351). In turning to a consideration of *inter-gestural timing*, the status of the timed elements, gestures and multi-gesture units, becomes relevant.

As outlined above, the search for invariant timing relationships has met with varying degrees of success within the gesture, between gestures in some larger multiple-gesture unit such as the segment or syllable, and between gestures not linked in such a unit. The first case, invariance within a gesture, has been confirmed in many efforts exploring task dynamics. The second case, between linked gestures, looks promising but prompts the question of what unit might be at work linking the relevant gestures. It is not any two gestures which display this cohesiveness but rather gestures which have long been considered to belong to the same segment--the lips and glottis in [p], the velum and lips in [m], and the tongue tip and body in [1]. As exemplified by Nittrouer et al. and Keller, the existence of invariant phasing relationships between segments is not obvious.

In discussing their proposed phasing rule for consonant clusters, Browman and Goldstein (1988) suggest that it may need to be refined to include syllabification effects and articulator specific effects. This seems to suggest concerns similar to those voiced here about the status of invariant phasing rules. Either the number of rules needs to be proliferated, e.g. specific to particular sequences; or the relationship must be implemented in such a way that variability is possible. We pursue the latter course here.

1.3 **IMPLEMENTING TIMING**

In Articulatory Phonology, lexical distinctions are made in a limited number of ways (Browman and Goldstein, in press). One, a gesture may be present or absent. (As a gesture is defined by what tract variable it operates on and its constriction location and degree; the presence or absence of a particular gesture is a contrast equivalent to contrasts in specification of tract variable, constriction location or constriction degree.) Two, the temporal coordination between gestures may differ. Lexical gestural scores may be altered in two ways after their creation. One, the gestural descriptors, *i.e.* the magnitude or stiffness of a gesture, may be changed, and two, changes in phasing may occur. Thus it appears that two instances of phasing may occur; one in the lexicon and one external to the lexicon. (Recall that association, *i.e.* an operation specifiying what is to be phased, may also operate on lexical entries.) Lastly, utterances must be coordinated above the word level such that there is some mechanism for phasing a word relative to the preceding word. This, by definition, must also take place outside the lexicon. A clearer understanding of the operation of association and phasing will require the further development of a theory of the lexicon. Once the lexical specifications are understood, the status of rules altering phasing relationships becomes clearer. Currently in Articulatory Phonology, some allowance must be made for rules changing phasing relationships, for example due to rate, stress and style. But in order to capture linguistic variability in this way, issues of rule ordering, the representation of rule input and output, and bleeding/feeding relationships must be considered. The explanatory usefulness of such an approach must also be considered. The proposal explored in Chapter Seven addresses these concerns by allowing variability in the assignment of phasing relationships rather than by creating a set of timing rules which operate on the phonetic/phonological representation.

One interpretation of the research described above in Section 1.2 with respect to interarticulator timing is that potentially contrastive coordination and non-contrastive coordination are subject to different constraints and requirements. We pursue this possibility. Because phasing relations create meaningful distinctions in the lexicon, it does not seem unreasonable to assume that they may be categorical here as evidenced in discrete, stable timing relationships. However, rather than pursuing the conception of a small number of invariant phasing relationships specified *outside* the lexicon, I suggest that such non-contrastive phasing relationships are variable and will depend upon a wide-range of linguistic and extra-linguistic factors. That is, while maintaining the use of phasing to capture intrinsic timing, we reject the assumption of invariant timing.

The discussion undertaken in Chapter Seven outlines a probabilistic, rather than deterministic, approach to intergestural phasing. I call this approach the PHASE WINDOW Model. I suggest that a particular phasing relationship (for example between two consonants in sequence) is constrained both physically (by human genetics) and language-specifically (by learning) to occur within a certain permissible window. I propose, in accordance with Nittrouer et al., that linguistic and non-linguistic variables which differ from utterance to utterance determine where in the range of permissible overlap relationships a token is likely to be realized. We call this range the PHASE WINDOW. We call the variables which weight the PHASE WINDOW influencers. (Note that Keller calls similar variables "co-determiners.") In sum, the concept of the PHASE WINDOW refers to the range of possible phasings. Like Keating's spatial Window Model of Coarticulation, the temporal PHASE WINDOW Model assumes that some kind of optimization occurs on line. Of course, this optimization is extremely complex because of the large number of simultaneously present *influencers*. Clearly, we do not adopt extrinsic timing or targets-interpolation approaches. We envision the proposed PHASE WINDOW Model as replacing the use of phasing rules within an intrinsic timing, Articulatory-Phonology-like, framework, and as a rejection of the notion of invariant intergestural timing.

Keller (1990) rejects invariant (and intrinsic) timing citing some of the same reasons which motivate our desire to allow for variability within an intrinsic timing framework. He comments that models such as that proposed by Kelso, Saltzman and Tuller (1986) "do not make provisions for non-motoric co-determinants of speech timing...[and] do not incorporate constraints inherent in task execution and system operation" (Keller, 1990) (Keller's conception of task constraints, e.g. specific segment durations, and of system constraints, e.g. "linguistic proficiency," is not adopted here. For Keller these are available because of his rejection of intrinsic timing in favor of direct influences from the central nervous system and reference to absolute time.) Maintaining an intrinsic timing framework, we suggest that there are upper and lower limits placed on a particular PHASE WINDOW which are determined by both system contraints (physical motor and cognitive capability) and task contraints (learned languagespecific, permissible coordination). Clearly the window defined by the latter constraints will be properly contained in that defined by the former. Other utterance-specific *influencers* weight the window but do not constrain the timing further. Keller's idea of mutual competition between codeterminers of speech timing and inclusion of perceptual and prosodic factors as influences on timing is in agreement with the theoretical proposal here of a mechanism to incorporate motoric and non-motoric *influencers* on speech timing a way as which allows variability in articulatory timing. We agree with Keller that

surface variability does not "hide' some ill-understood invariants under measurement error and articulator imprecision. On the contrary, surface variability is considered to be the inevitable and theoretically predicted concomitant of a communicative behavior capable of achieving its overall objectives by satisfying a large variety of competing demands... (Keller, 1990, p. $357)$

This approach suggests the pursuit of a research program to identify the factors influencing timing relationships and the nature of those effects. To this end, specific potential *influencers* on timing in consonant sequences are examined below. The empirical goal of this thesis is to pursue a series of studies to determine the effect of a variety of linguistic and extra-linguistic variables on the articulatory timing of a set of lingual consonant clusters. This thesis has as its theoretical goal the development of a model of speech timing which allows variability in the assignment of phasing relationships; and in which variability is constrained and, to some degree, predictable. We call this approach, outlined in Chapter Seven, the PHASE WINDOW Model.

CHAPTER TWO: AN OUTLINE OF EXPERIMENTS AND METHODOLOGY

2.0 **FOUR EXPERIMENTS**

The studies described below examine the spatial and temporal production of English alveolar and velar stops and alveolar fricatives in sequence. Heterorganic sequences of varying lengths, both within and across words, are considered. The following chapters investigate the effects of rate, syllabification, sequence length, and consonantal place and manner on the timing of consonant sequences. Interactions of factors with speakers are noted, but the main focus of the experiments is to describe these main effects independently.

Below are some hypotheses evaluated in this study. Two types of timing are considered: local and global. We use *local timing* to refer to the coordination of lingual movements for two adjacent consonants. In the examination of local effects both changes in a particular movement due to its context and changes in the relationship between two movements as a function of context are considered. We use *global timing* to refer to how a sequence of consonants is coordinated as a unit to the vowels on either side of it. The experiment addressing global timing roughly replicates Browman and Goldstein's (1988) experiment on C-Centers, but focuses on coda clusters rather than onset clusters. Experiments One and Two both consider local timing effects, and Experiment Three addresses the question of global timing and the question of what is phased together. Experiment Four is relevant in considering both local and global timing. Experiments One, Two, and Four concentrate on how the consonant cluster phasing is realized.

In Experiment One, effects of the place and manner of constriction and syllable affiliation are addressed. Sequences of two consonants in different order across a word boundary (for example "bag dab" and "bad gab") are evaluated to determine whether the relative timing of the consonants differs. Hypotheses one through four are considered. Hardcastle and Roach (1979) found that the time between the initiation of contact for an adjacent [t] and [k] was shorter for a [tk] cluster than for a [kt] cluster. This suggests the formulation of a hypothesis regarding the effect of place on overlap.

H1: A tongue tip gesture is more temporally overlapped with a following tongue body gesture than a tongue body gesture with a following tongue tip gesture.

This hypothesis is of relevance in phonology because of the exceptional behavior of coronals in assimilation processes and the often-made, controversial assumption that this asymmetry is due to the fact that coronals are underspecified for Place features. The above hypothesis, if supported, suggests an articulatory basis for this behavior (see also Byrd, 1992).

Because we had reason to believe there were place effects, we wanted to also investigate any effect of articulatory manner that might exist as well. The degree of overlap permitted in a particular condition may be a function of the constriction degree of the consonants involved. For example, the following hypothesis is tested:

H2: A closure (compression) movement is more temporally overlapped by a following consonant articulation than is a fricative (constriction) movement.

Experiment One also examines how the syllabic position of a consonant, *i.e.* word-final (coda) versus word-initial (onset), correlates with degee of lingua-palatal contact. A decrease in the gestural magnitude of codas is often assumed to be the motivation for lenition processes, for example. The following hypothesis is proposed.

H3: Movements in coda position are smaller than movements in onset position.

This has been shown for labial and coronal stops (Fromkin, 1965, Krakow, 1989; Browman and Goldstein, in press). We also consider the coronal fricative and the velar stop to determine if this is generalizable to other places and manners. Similarly, we evaluate how individual (lingual) consonants differ in displacement depending on the neighboring consonant. Differences in lenition are examined both in sequences with two lingual consonants $(e, g, [g \# d])$ and in sequences with a lingual and a labial $(e.g. [g#b])$ consonant. It might be expected that more reduction would occur if two consonants are making demands on the tongue simultaneously. The role of spatial and temporal variability in reduction is addressed to determine whether reduction *(i.e., a decrease in duration or displacement)* always co-occurs with increased variability. For example, assuming the validity of the preceeding hypothesis H3, is it true that:

H4: Movements in coda position will be more variable than movements in onset position.

An alternative, of course, is that greater displacement is associated with greater variability.

Understanding the interaction of prosodic structure such as syllable boundaries on segment level representations or phonetic implementation is a crucial question for the phonology-phonetics interface. Experiment Two considers the effects of the placement of syllable boundaries on magnitude and timing. The following two hypotheses are tested to determine how syllable affiliation affects timing between consonants in a sequence.

H5: Temporal coproduction of articulatory movements for sequential consonants is greater if the consonants are tautosyllabic, less if they are heterosyllabic.

H6: Coordination of articulatory movements for sequential consonants is less variable if the consonants are tautosyllabic than if they are heterosyllabic.

The latter is found to be the case by Browman and Goldstein (1988), Hardcastle and Roach (1979) and others. The status of a consonant sequence as a possible onset in a language has been suggested to be a factor in the assignment of phasing relations (Browman and Goldstein, 1990b; see also Kent and Moll, 1975). Browman and Goldstein's supposition was that potential onsets would be phased together by a rule which would not apply to other consonant sequences. They suggest that a sequence, regardless of its canonical syllable association, is phased according to their rule for consonant clusters if, and only if, that sequence is well-formed as a possible syllable onset or coda cluster. If this is the case, then their rule states that each consonant in such a sequence will be phased so as to synchronize its onset to the offset (specifically 290°) of the preceeding consonant. Browman and Goldstein propose that if the sequence is not a possible onset or coda, then the consonants are not phased with respect to one another. Browman and Goldstein thus predict that the status of being a well-formed possible cluster will correlate with the "tightness of [its] gestural organization" (Browman and Goldstein, 1990b: 369). They suggest, specifically, that non-well-formed sequences may have more variation in overlap and may be more likely to have greater amounts of overlap. Their approach is discussed in more depth in Chapter Four.

In Experiment Three, sequences of coda consonants and sequences of coda consonants followed by an onset consonant are used to evaluate the predictions of Browman and Goldstein (1988) regarding the global timing of consonant sequences. Browman and Goldstein examined one speaker's production of sequences like #sp, #pl, #spl, s#p, p#l, and s#pl. They found that onset consonants and sequences of onset consonants are timed to the following vowel by phasing their C-Center to the vowel. (For simplicity, we call the relationship between the consonants and the following vowel the CV relationship and that of a sequence to a preceding vowel the VC *relationship.* Of course more than one C may be included in this global timing relationship.) The C-Center was defined as the arithmetic mean of the centers of each plateau of peak articulatory displacement for each consonant in the onset sequence. The relationship of an onset sequence to a preceding vowel was found by Browman and Goldstein to be most stable for the time between the achievement of target of the first onset consonant and the vowel. Browman and Goldstein called this point the "left edge" and defined it as the initial edge of the plateau of peak articulatory displacement for the first consonant in the onset sequence. (We use the term first edge with an analogous definition.) Browman and Goldstein additionally examined the behavior of a single coda consonant followed by an onset sequence. Only the VC timing relationship was examined there, not the CV coordination across the syllable (C#V) boundary. This type of sequence was found to be organized in the VC relationship in the same way as an onset sequence without a coda consonant. Alignment of the achievement of target of the first

consonant, whether an underlying onset or coda, with the preceding yowel was found to be the more stable than either the C-center of the coda consonant or of the whole sequence. In sum, Browman and Goldstein's experiment showed that consonants in sequence are organized with respect to the preceding yowel by the left edge of the first consonant in the sequence regardless of the underlying syllable affiliation of the consonants. They also showed that onset sequences are organized with respect to the following vowel by their C-Centers. They conclude based on this data that syllable-initial consonants are "related to their words in terms of a single global metric for the entire cluster, the C-Center, [syllable-final consonants] appear to be related to their words in terms of the local metric of achievement of target" (Browman and Goldstein, 1988, p. 149). They make the additional prediction that the more consonants in a prevocalic cluster, the shorter the acoustic realization of the vowel, as these consonants, unlike coda consonants, overlap the interval of yowel activation. Conversely, there should be no decrease in the acoustic duration of the preceding vowel as coda consonants are added.

In light of this work and with the desire to explore coda sequences in addition to single coda consonants, Experiment Three evaluates hypotheses seven through twelve. Hypotheses seven and eight are suggested by the findings of Browman and Goldstein (1988).

H7: The most stable VC organization involves the initial edge of the first consonant in the sequence.

H8: In a coda sequence followed by a single onset consonant, the most stable CV organization involves the C-Center of the onset consonant.

Browman and Goldstein did not examine the CV timing relationship for coda consonants or report a comparison between the C-Center for the onset sequence versus the C-Center for an entire coda+onset sequence for the CV timing relationship. It could be that the CV relationship, like the VC, is unaffected by the canonical syllable affiliations of the consonants. This possibility suggests the following alternative hypothesis:

H9: For a consonant sequence, the most stable CV organization involves the C-Center of the entire consonant sequence.

The nature of the overlap between the consonant sequences and the preceding and following vowels is also investigated. These hypotheses bear directly on Browman and Goldstein's (1988) predictions described above regarding the acoustic durations of the vowels neighboring the consonantal sequences. Our hypotheses also relate to work on compensatory shortening. Munhall et al. (1992) found that acoustic vowel durations were shorter before clusters than before single consonants. They say that this shortening is compensatory in nature

such that vowel duration gets shorter as coda duration increases. We test the following hypotheses:

H10: The period from the acoustic vowel onset to the initial edge of a following consonant sequence is unaffected by the number of consonants in the sequence.

H11: The period from the final edge of a consonant sequence to acoustic offset of the following vowel is shorter if an onset consonant is present than if no onset consonant is present.

H12: The period from the final edge of a consonant sequence to acoustic offset of the following vowel is unaffected by the number of coda consonants in a preceding sequence.

These hypotheses represent the predictions of Browman and Goldstein (1988). Finally, in Experiment Three the articulatory realization of consonant sequences of varying length and composition is reported informally, concentrating on the characteristics of reduction in these sequences.

The last experiment manipulates speaking rate to determine what influence rate has on the articulation of heterosyllabic CC sequence. Differential effects on the specific sequences are also considered. The approach to studying rate is modeled in a general way on Gay's (1981) study of the EMG and X-ray data for evidence of mechanisms controlling speech rate in VCV sequences. Gay found that:

...speakers do not control their rate of speech by either a single mechanism or along a single dimension. The fact that the duration of segmental units, the displacement and velocity of articulatory movements, and the temporal overlap between individual segments undergo nonlinear transformations during changes in speaking rate, precludes the operation of a single mechanism for rate control. On the surface, we can attribute these nonlinearities to linguistic constraints such as sound types, language differences, and stress, or to individual differences in motor strategy. However, a more satisfactory understanding of the organization of speech rate cannot be realized until we have a better understanding of the organization of speech motor behavior in general. (p. 158)

Gay (1981) found that duration changes were not distributed proportionally across consonant and vowel segments. Gay interpreted this discovery as reflecting a restructuring of the temporal pattern of an utterance rather than a simple change in the spacing of motor commands (Gay, 1981, p. 151). We ask the parallel question of whether timing changes in a consonant sequence are distributed proportionally across the consonants. That is, do speakers control rate by decreasing the duration of both consonants in the sequence in the same manner? A possible

alternative is that they change the relative timing of successive gestures so that overlap increases as rate increases (see Gay, 1981; 157). He suggests, as have others, that speakers may employ different strategies to effect rate change. We consider the two mechanisms of shortening and overlap for changing speaking rate in two-consonant sequences. In our study, the following hypotheses are considered:

H13: As rate increases, the absolute time between the two consonants will decrease.

H14: As rate increases, the relative time (overlap) between the two consonants will increase.

We also want to know whether these timing changes are linear in nature.

H15: Stop-stop sequences will be subject to greater influences of rate than sequences involving a fricative.

Questions about spatial changes in articulatory contact are also addressed. Hypotheses regarding which consonants are lenited and under what conditions are evaluated. With respect to spatial displacement, Gay compares Lindblom's (1963, 1964) reasoning that the degree of undershoot is directly proportional to duration with his (Gay's) 1968 and 1974 work and Kent's (1970) work suggesting that rate may cause changes in articulatory effort or velocity signifying a reorganization in muscle forcing function in addition to a temporal reorganization (Gay, 1981: 152). Gay says, "while target undershoot commonly accompanies an increase in speaking rate, it is by no means ubiquitous" (1981: 152-3). Specifically, the following hypothesis is tested:

H16: An increase in speech rate causes all consonants in a sequence to be reduced, *i.e.* to be produced with less contact between the tongue and the palate.

If this is not the case, differences between consonants and syllable position of course become the interesting variables. Barry (1991), for instance, found coronal gestures to reduce in rapid speech, but not dorsal ones.

To summarize, this study considers the effect of speaking rate, syllabification, sequence length, and consonantal place and manner on the timing, magnitude, and variability of consonantal constrictions in heterorganic consonant sequences.

$2.I$ EQUIPMENT AND THE NATURE OF THE DATA

As described above, electropalatography will be used to record information about speakers' tongue-palate contact over time. The electropalatograph uses an artificial palate of thin

acrylic embedded with electrodes (usually over 60). The pseudopalate may be manufactured individually for each subject from a dental cast or be uniform for all speakers with similar palate sizes. The palate is scanned, and contact data are obtained at a sampling rate of 100 to 200 Hz (Hardcastle, Gibbons, and Nicolaidis, 1991). The data are typically recorded by computer and are qualitatively examined visually in the form of diagrams showing the arrangement of electrodes on the palate indicating contact or no contact at each electrode. The sensor information from the palate is converted into a real-time graphic moving image of tongue-palate contact.

We use the Kay Elemetrics Palatometer which has a thin acrylic palate that extends around the teeth. In order to create a palate for each speaker, a dental impression of each speaker's hard palate was made. Then a dental stone cast made from the impression was used by Kay Elemetrics to manufacture a custom-fitted artificial palate. These acrylic palates were tested for snugness of fit before being accepted for the experiments below. None were rejected. The Kay Palatometer uses a palate with 96 electrodes. It scans the palate at a 100 Hz sampling rate with a scan time of 1.7 milliseconds to acquire all 96 values in a sample. The data acquisition unit is isolated from the host computer with opto-isolation integrated circuits. Twelve volt DC is supplied to the unit from an external medical grade power supply with current limiting circuitry in the subject-contact circuitry. The Palatometer is designed to interface with the Kay Computer Speech Lab, an acoustic analysis system allowing the simultaneous examination and analysis of spectrograms, waveforms, and palatograms. The speech acoustic signal is acquired simultaneously with the linguapalatal information at a sampling rate of 12,500 Hz.

The data analysis methods for EPG data are comparatively straightforward, particularly in light of the specialized software provided with commercial systems. We will concentrate on analyses yielding dynamic information about the consonant sequences. Quantitative data is generally obtained using specifically designed software for data reduction. Such software uses numerical indices to describe the amount or frequency of contact. An index calculation might include contact across the whole palate or contact in a particular articulatory region or contact for a certain row or arc of electrodes. An index of linguapalatal contact might be given for a specific point in time or might be a description across time. This latter has been called a "totals" display (Hardcastle, 1991).

Very few researchers have examined dynamic temporal patterns in EPG data, as opposed to static spatial patterns. Here, we've explored innovative methods of data analysis. Marchal (1988) uses a qualitative approach to describe temporal characteristics of two-stop sequences in French, with four classes of articulatory and acoustical behavior as evidenced in closure and acoustic release patterns. Barry (1985) uses a similar qualitative classification of two-stop sequences, with three classes--'non-assimilated', 'assimilated plus residual articulation', and 'totally assimilated.' Nolan (1992) uses a similar categorization in examining coronal reduction: "full-alveolar, residual-alveolar, and zero-alveolar." This study will augment qualitative

descriptions of temporally dynamic patterns with quantitative analyses such as those described by Barry (1991), Hardcastle and Roach (1979), and Hardcastle (1991). Most temporal displays show the total number or percent of electrodes contacted at each frame over time. These curves have been called "trajectories" or "contact profiles". Measurements can be made off these displays. This method is used by Barry (1991) for temporal analysis of coronal#velar sequences separated by a word boundary.

The present work uses a display showing the percent of contacted electrodes in particular articulatory regions over time. Below is shown a sample contact profile for an [sg] sequence. Time is on the x-axis and the percent of a region contacted on the y-axis. First, we see front region contact increasing for the s, and then at frame eight we see contact in the back region for g initiating. This type of display will form the basis of the graphic data presentation in the following in Chapters Three and Four.

FIGURE 2.1 SAMPLE CONTACT PROFILE FOR [sg].

The novel use of speaker-specific articulatory regions, the criteria used in establishing these regions, and the exact data analysis methods are described further below. Software external to the Kay Elemetrics software was written for UCLA and used to calculate these regional contact profiles.

$2.1.1$ **LIMITATIONS OF EPG**

Like any instrument, EPG has limitations. Some of these have already been mentioned; others are specified in the relevant sections later in the chapter. At these points we attempt to address the specific relevance of these issues, if any, in this study. However, here we would like to enumerate more generally several concerns in using modern electropalatography. Many of these problems are commensurate with difficulties of other types of movement tracking systems. The reader is encouraged to see Hardcastle (1984) for a thorough overview.

First, the selection and number of experimental subjects is important in any empirical study. The use of EPG, due to the expense and difficulty of making well-fitted pseudopalates, may increase the likelihood of experimental subjects being few in number and unrepresentative of a larger population (Ladefoged, 1957). Most previous EPG studies include only one or two speakers of a language. Second, pseudopalates interfere with sensory feedback, potentially inducing non-typical articulation. However, researchers have concluded that it is unlikely that the simple tactile sensory resources of the palate play a significant role in sensory discriminations in the mouth (McDonald and Aungst, 1967). Fletcher (1992) states that lingual feedback is sufficient to compensate completely for the loss of tactile information from a thin pseudopalate. Third, the presence of an appliance in the mouth might interfere with normal articualtion. Research has shown no significant difference in patterns of tongue-palate contact between direct palatography and EPG (Hardcastle, 1972, see also Flege, 1976 and Fletcher, McCutcheon, Wolf, Sooudi, and Smith, 1975; but see Hamlet and Stone, 1978). No difference in intelligibility with versus without the palate was found by Kozhevnikov and Chistovich (1965). Note that the palates used in the experiments below are thin compared with other commercially available fitted pseudopalates. Fourth, the electrode coverage is limited mostly to the hard palate area making it possible that contact on the pseudopalate under-represents the full area of velar closure, specifically contact occurring well onto the soft palate. However this problem is not serious for front velars. Hardcastle and Roach (1979), using a smaller pseudopalate, observed complete velar closures on their pseudopalate in the phrase 'catkin.' In the present study using the Kay Palatometer with 96 electrodes, an examination of ten repetitions of the control utterance for velars, 'Say bag gab again,' showed that every token for every subject had a complete seal across the back of the palate for the velar closure. Some tokens showed up to five electrodes contacted along the mid-sagittal plane. The Kay Palatometer pseudopalates also have electrode coverage onto the dentition bordering the hard palate. Interdental contact is not captured anteriorly; however, a degree of dental contact is observed as the electrode coverage extends onto the top of the front teeth. Lastly, because the instrumention only measures contact, inferences about articulator velocity, complete trajectory, tongue shape, or time of innervation are hazardous and may be undertaken only when clearly indicated by the tongue-palate contact patterns.

2.2 **EXPERIMENTAL MATERIALS**

Experiment One considers the following two-member, heterosyllabic consonant sequences:

$C1 \Rightarrow C2 \parallel C2 \parallel$	IJ		ш	
		3U	ab	
	υs	SS		gs
	bd		aα	ga
		sg		gg

TABLE 2.1: SEOUENCES IN EXPERIMENT ONE

All of these sequences have a word boundary separating C1 and C2. Throughout this study word boundaries are used as a general diagnostic for syllable boundaries. This is justified in that Hardcastle and Roach (1979) found no effect of a word boundary (-VC#CV[']-) as opposed to a coda+onset boundary (-VCCV-) on any of their timing measures. The test words and carrier phrases in which these sequences were recorded were *Type 'baC Cab' again*. (No underlining appeared on the speakers' pages.) The vowel bordering the consonants was pronounced $[\mathbf{\hat{x}}]$; this in turn was bordered by bilabial stops so as to minimize any lingual coarticulation. The test words and carrier phrases in which these sequences were recorded were (underlining has been added):

The use of the adjacent low front vowel, $[\alpha]$, was intended to create a somewhat front velar constriction that would be most observable on the pseudopalate and, at the same time, to minimize linguapalatal contact during the vowel.

Experiment Two considers the following two-member consonant sequences:

where # stands for a word (syllable) boundary

‡Note that the tokens used for ks# were taken from the recordings made for Experiment Three which used a slightly different carrier sentence

TABLE 2.2: SEQUENCES IN EXPERIMENT TWO

The test words and carrier phrases in which these sequences were recorded were (underlining of the consonant cluster has beed added):

As in Experiment One, target sequences were bounded by /xe/ whenever possible, which in turn was bordered by bilabial stops so as to minimize any lingual coarticulation. As /æ/ cannot occur at the end of a word, the initial sequence [sk] was preceded by [ə].

Experiment Three considers the following consonant sequences:

TABLE 2.3: SEQUENCES IN EXPERIMENT THREE
The test words and carrier phrases in which these sequences were recorded were:

Experiment Four considers four heterosyllabic sequences: $d\#g$, $g\#d$, $s\#g$, and $g\#s$. They were recorded in the frame sentence "Say baC Cab again."

For Experiments One, Two, and Three the material was randomized in ten blocks; thus ten repetitions were recorded. Tokens from blocks two through eight are included in the data analysis. For Experiment Four the goal was to create as much variation in speaking rate as possible without allowing abnormally slow speech. To accomplish this, blocks of the four sequences were created with a rotating order (with the constraint that no two instances of the same C1 abutted). 32 such blocks were recorded with the speakers instructed to increase their speech rate through each sentence in a block of four. The first sentence was cued with the word "Normal", the second with "Medium", the third with "Faster" and the fourth with "Fastest." The speakers were instructed specifically to read the first sentence in each block at a normal speaking rate.

All material for the two-member sequences included in Experiments One and Two (and some additional material) were recorded in one two hour session. Experiment Three and Experiment Four were recorded at a second one and a half hour session two to four weeks later. All material from the first recording session was randomized and recorded together. Experiments Three and Four, recorded during the second session, were recorded separately.

Throughout the remainder of this work, the consonant sequences will be presented in bold-face, e.g. sp. Since all of the consonant clusters in Experiments One and Four span a word boundary, only Experiments Two and Three will include a # symbol marking the syllable boundaries, e.g. #sk, s#k, sk#.

2.3 **SUBJECTS**

Five speakers were recorded. These include two men and three women who have grown up and been educated in Southern/Central California. Speakers were paid at a standard compensation rate. Speakers will be referred to as Speaker A, Speaker B, Speaker K, Speaker M, and Speaker S. All speakers speak with a dialect characteristic of Southern or Central California. All speakers reported no speech or hearing pathology. Details of their language background and relevant personal characteristics are given below.

Speaker A was born in Pennsylvania and moved to Santa Cruz, CA at 8 months where she lived until she attended college in Santa Barbara, CA. One year was spent in Paris, France during this

period. Since this time, she has lived in the Los Angeles area. Speaker A was a graduate student in linguistics (not studying phonetics) and was in her twenties at the time of recording.

Speaker B grew up in San Diego and Los Angeles areas of California. She attended college in Pomona and Los Angeles. Since college she has lived in San Diego for five years, West Los Angeles for 13 years, and Germany for five years. Speaker B was a graduate student in phonetics and was in her fifties at the time of recording. Speaker B has had training as a singer which included pronunciation training.

Speaker K grew up in the San Bernadino area of Southern California. She attended college in the San Francisco area of California, after which she lived in Tonga for a year. She then returned to live in the Los Angeles area. Speaker K was a graduate student in phonetics and was in her twenties at the time of recording.

Speaker M grew up in the Santa Cruz area of California, and he moved to Los Angeles where he has lived since starting college. He was an undergraduate major in linguistics. He was in his twenties at the time of recording.

Speaker S was born in Indiana where he spent his first ten years. Subsequently, he lived in the San Francisco area of California through college. Since that time he has spent an additional two years in the San Fransisco area, two years in Germany, and thirteen years in Los Angeles. Speaker S was a graduate student in linguistics (not in phonetics) and in his thirties at the time of recording.

2.4 **RECORDING SET-UP**

Before each experimental recording, the speakers wore their artificial palates for an hour of normal activity to accommodate. For recording, the speaker was seated in the laboratory near the palatometer computer module, but facing away from the monitor. The palate electrodes were calibrated using the Palatometer software. For each recording, one practice page $(i.e.$ one block) of material was read by the speaker before the recording started. Nonsense words $(e.g.$ sab) were pointed out to the speaker at this time. An independent but simultaneous voice recording was made via a head mounted, directional microphone connected to the external Computer Speech Lab hardware which interfaces with the computer. Both the EPG and audio signals were recorded directly into a single computer file. Speakers were cued for each sentence by the word Go from the experimenter. Each sentence was cued individually, and there was a pause after each one. Recording sessions lasted from one to two hours. Speakers were invited to take a break half-way through, and were required to do so for the two-hour sessions. Subjects were instructed to maintain a constant, fluent reading rate with no unusual stresses, and deviations

from this were pointed out by the investigator during the training block. If a speaker paused, hesitated, or otherwise had a false start, he or she was prompted for a repetition of that sentence.

2.5 **REGION DEFINITION**

Much EPG work refers to articulatory regions on the pseudopalate. These are groups of electrodes in a definable subsection of the palate, often corresponding to a "place of articulation." All previous EPG work of which I am aware has used *predetermined* regions on the pseudopalate for data analysis. These regions are the same for all speakers in an experiment and may even be hardwired into the pseudopalate itself. Rather than either using hardwired articulatory regions or otherwise predetermined and identical pseudopalate regions for all speakers, I determined articulatory regions empirically for each speaker. This determination was based on control utterances having no lingual coarticulation with another consonant. These regions were empirically defined in order to allow for differences from speaker to speaker in the placement of the electrodes with respect to anatomical configurations or in speaker-specific articulatory patterns such as laminality or apicality. All ten repetitions of the ambisyllabic dd, ss, and gg sequences were used to establish front and back regions of the palate for every speaker. Crucially, for each subject, no electrodes that were contacted at the minimum for the $[\mathbf{\hat{x}}]$ vowel contact were included in the consonantal region. This ensures that the moment of initial contact measured in the sequence will in fact be the concomitant of the formation of a consonant constriction rather than normal vocalic contact. All electrodes contacted after the vowel minimum, until and including the frame of maximum contact during the consonant, were designated as belonging in the relevant region: front for s and d, back for g. Any electrodes which were marked in this way as members of both regions were also excluded. These cases were generally few and always adjacent to the excluded vocalic region. The result of this is that the measurements made are conservative in identifying the frame of initial contact for a consonant but that there is a high degree of confidence that the contact measured is actually attributable to the upcoming consonant in that region. All other (i.e. uncontacted) electrodes were also included in one of the two regions. It was important to make the region as large as possible to avoid saturation whereby the contact level remains at 100% over a period of time. If an electrode was never contacted during the control sequences, it was included in the region to which it was physically closest. This was determined by measurements made with a flexible ruler on the acrylic palates themselves. The resulting regions for each subject can be seen in Appendix A where an effort has been made to preserve the x, y, and z dimensionality of the pseudopalates, as palate size and depth differ substantially from person to person.

2.6 **MEASUREMENTS**

Because of the establishment of speaker-specific articulatory regions on the pseudopalates, this study uses a *percent* display (rather than a "totals" display) showing the percent of contacted electrodes in the front and back regions in each frame rounded to the nearest integer. The main basis of the quantitative analysis in this dissertation is these contact profiles.

Of course this method, like movement tracking of articulatory trajectories, offers no means of determining the moment of neural activation. However, it is roughly commensurate with studies of articulatory trajectories. It does miss the very beginning and end of articulatory movements before the sides of the rising tongue contact the palate. X-ray microbeam movement tracking, comparably, misses the beginning of a movement if it occurs outside the mid-sagital plane.

2.6.1 MEASUREMENTS FOR EXPERIMENTS ONE AND TWO

For Experiments One and Two the contact profiles for both regions were examined to determine four timepoints during each period of consonantal activity: first frame with any contact in the region, first frame at maximum contact, last frame at maximum contact, and last frame with any contact in the region. (Note that, between the first and last frame of peak contact, any dips away from maximum contact, although infrequent, were allowed.) The percent contact at the peak was also recorded. These measures were used to calculate the temporal latency between the two consonants, their individual durations, and the degree of lingual displacement as indicated by linguapalatal contact. The specific variables calculated for various tests include:

•Region duration (FRONT DURATION and BACK DURATION)--the duration of linguapalatal contact in a region in seconds; *(time of final contact - time of initial*) $contact$) $region$

•Sequence duration (SEQUENCE DURATION)--total duration of linguapalatal contact for a sequence in seconds; (time of final contact - time of initial $contact)$ _{sequence}

•Region duration relative to the sequence (FRONT DURATION (%) and BACK DURATION $(\%)$ -the duration of articulatory contact in a region as a function of the entire constriction duration of the sequence; REGION DURATION/SEQUENCE DURATION

•Peak contact (FRONT MAXIMUM and BACK MAXIMUM) the maximum percent contact in a region

• Sequence overlap (SEQUENCE OVERLAP $(\%)$)--the percent of the total sequence duration during which contact occurred in both regions

•C1 overlap (C1 OVERLAP $(\%)$)--the percent of C1 duration during which contact for C₂ also occurred

• $C2$ overlap (C2 OVERLAP $(\%)$)--the percent of C2 duration during which contact for C1 also occurred

•Time between onsets (Δ ONSETS)-time between initial contact in one region and initial contact in another region in seconds

•Time between peaks (Δ PEAKS)--time between peak contact in one region and peak contact in the other in seconds, where the time of peak contact is calculated as the temporal center of the plateau of maximum contact

•Time between C1 release and C2 onset (\triangle RELEASE OF C1 TO ONSET OF C2) $$ time between the first frame after the final frame of maximum contact for C1 and the initial contact for C2 in seconds

•C2 onset relative to C1 (C2 ONSET RELATIVE TO C1 $(\%)$)--the percent of the time into C1 at which the initial contact for C2 occurs

•C₂ peak relative to C₁ (C₂ PEAK RELATIVE TO C₁ $(\%)$)--the percent of the time into C1 at which the peak contact for C2 occurs, where the time of peak contact is calculated as the temporal center of the plateau of maximum contact

Again, not every dependent variable will be evaluated in every experiment.

Region duration is a measurement of the length of time required for a particular articulatory constriction from the onset to the offset of palate contact. Region duration relative to the sequence is a measurement of the temporal dominance of one constriction in a sequence of constrictions. Absolute and relative region duration may be correlated, but only the latter is affected by overlap. These are the measurements which will be used to infer how the independent variables affect the overall time taken for the linguapalated contact of a consonant. Peak or maximum linguapalatal contact is indicative of a consonant's degree of lingual displacement. Reduced consonants will have less contact. (Note that the measure of peak contact, MAXIMUM, doesn't take into account the degree of displacement before and after the peak. The variables described below of area and flatness of the contact profiles serve as additional indicators of reduction.) Sequence overlap is a measure of the time in the consonant sequence during which contact is being made in both the front (coronal) and back (dorsal) regions. It indicates the degree of coproduction occurring in the sequence. C1 overlap and C2 overlap indicate how overlapped a particular consonant may be by another constriction. The time between onsets, time between peaks, and time from C1 release to C2 onset are measures of absolute latency reflecting the temporal coordination of the tongue tip and tongue body movements. The time from C1 release to C2 onset was chosen as a timing measure as an approximation of Browman and Goldstein's (1990b) suggested phasing of consonants in sequence (see discussion above) such that one consonant begins at the offset (290°) of the preceeding one. All three of these timing measures roughly reflect the *latency*, *i.e.* temporal

offset, of the second consonant in a sequence with respect to the first. The variables of C2 onset and peak relative to C1 are also measures of latency, these taking into account differences in C1 duration. They reflect the relative latency of $C2$ with respect to $C1$. The variance of these measures roughly reflects the degree of stability of the phasing relationship in a particular sequence.

In addition to the variables outlined above, the contact profiles themselves were used to calculate four indices which depended on the overall shape of the profile, i.e. took into account the displacement at each frame. These are described below.

• AREA under the curve (contact profile) is calculated by summing the percent contact in the region at each frame over time. This is an index describing overall displacement and temporal extent and will be used as a measure of articulatory reduction, in addition to the variables of region duration and peak displacement.

•NON-OVERLAPPED AREA is the area as described above for only that portion of the contact profile during which the other region in the sequence showed no contact. This is also an index of reduction which attempts to take into account *hiding* effects of the other consonant in the sequence.

•SKEW of the contact profile is an index describing the degree of asymmetry between the onset and offset portion of the contact profile.

•FLATNESS (MEAN/MAX) is a measure of the mean contact divided by the maximum contact thereby yielding an index of the overall flatness of the contact profile.

All of these variables are indices calculated over the linguapolated contact interval for a single consonant. Thus they are calculated for both the front and back regions.

2.6.2 MEASUREMENTS FOR EXPERIMENT THREE

In Experiment Three the same points in the contact profiles were measured as for Experiments One and Two with the following difference. Dips between frames of maximum displacement, while infrequent were allowed in Experiments One and Two. In Experiment Three, however, only dips of a single electrode after the first maximum in the *front* region were allowed. This ensured that the normal bobbling occasionally seen for [s] or [d] would not count as separate consonants. Otherwise intervening minima marked a separate consonant beginning at the next frame in which contact increased. This metric never produced more than the canonical number of consonants. In the time-synchronous acoustic waveform, measurements were made paralleling Browman and Goldstein's (1988) choice of preceeding-vowel and following-vowel anchorpoints. In both their experiment and this experiment, the point of acoustic closure for the consonant following V2 was chosen as the following-vowel anchorpoint. We call this the V2

ANCHOR. For the preceeding-vowel anchor, Browman and Goldstein chose the midpoint of the labial movement for the bilabial consonant preceding V1. Here, not having lip movement data, we have chosen the acoustic midpoint of the bilabial closure as the following-vowel anchorpoint. We call this the V1 ANCHOR. These measurements are shown schematically in Figures 2.1 and $2.2.$

FIGURE 2.2 SCHEMA I FOR A [kdk] SEQUENCE SHOWS THE LEFT AND RIGHT ANCHORPOINTS FOR EXPERIMENT THREE AND THE CONSONANTAL EDGE AND C-CENTER MEASUREMENTS. IN THIS SCHEMA ALL CONSONANTS ARE REALIZED WITH SEPARATE MAXIMA.

In the contact profiles for some sequences there was no separate maximum for each consonant.

FIGURE 2.3 SCHEMA II FOR A [kdk] SEQUENCE SHOWS THE LEFT AND RIGHT ANCHORPOINTS FOR EXPERIMENT THREE AND THE CONSONANTAL EDGE AND C-CENTER MEASUREMENTS. IN THIS SCHEMA, THE CONTACT PROFILE FOR THE BACK REGION PROPERLY CONTAINS THAT FOR THE FRONT CONSONANT. IN THIS SITUATION ALL CONSONANTS WERE CONSIDERED AS REALIZED.

In cases where the contact profile for one consonant completely contained that of a consonant in the other region, it was counted as contact for two consonants. This is shown in Figure 2.2. Completely-contained is defined as having the first mark (i.e. first frame of contact) for a consonant in the same frame or before the first mark of the next consonant *and* having the last mark *(i.e.* last frame of contact) for a consonant in the same frame or after the last mark for the preceeding consonant. In the schema above the contact in the back region completely contains that in the front region. Therefore, the back region contact counts for two consonants, the k in first position and the k in third position in the $[kdk]$ sequence. This metric never produced more than the canonical number of consonants in a sequence.

Other variables for Experiment Three are the times from the anchor poins to the first and last edges of the sequence. Browman and Goldstein (1988) defined their left edge as the edge of the plateau (within approximately 1.3 mm) of peak displacement of an X-ray microbeam pellet attached to an articulator. Analogously, we define FIRST EDGE as being the first frame of maximal contact for a consonant. LAST EDGE is defined similarly here as being the last frame of maximal contact for a consonant. Browman and Goldstein (1988) derive the C-center point for a sequence by computing the temporal midpoints between the left and right edges of the peak displacement plateaus and then calculating the mean of all the plateau midpoints of the gestures in the sequence. We do likewise taking the mean of the centers of each plateau of maximum displacement of each consonant in the sequence.

For the measures in Experiment Three, different sets of tokens are considered. For those measurements evaluating the first edge of the consonant sequences, all tokens (except two in

which no contact occurred for the first consonant) are included. For all measurements evaluating the C-Center or last edge, those sequences in which all consonants are articulated (where "all" includes the cases where one profile completely contains another) are included. In sequences having an onset consonant, when the center or edge of the onset or coda cluster is considered separately, only those sequences in which every consonant has a separate contact maximum are included, in order to be sure of isolating contact for the onset or coda accurately.

$2.6.3$ **MEASUREMENTS FOR EXPERIMENTS FOUR**

In Experiment Four measurements of the consonant sequences are the same as those outlined above for Experiments One and Two. However, an additional measurement of speaking rate is necessary. This measurement was also made from the articulatory data and was specific to the carrier sentence in which these sequences occurred: "Say baC Cab again." The number of frames from the first frame *after* the last contact for the consonant sequence to the last frame of maximum contact in the back region for the [g] in *again* was used as an index of speaking rate.

2.7 **STATISTICAL MODELS**

Data analysis will focus on tendencies apparent across subjects; however, significant individual differences will also be reported. In Experiments One and Two, spatial and temporal measurements made from the contact profiles are used in a repeated-measures General Linear Model (GLM) Analysis of Variance (ANOVA) model. Except where noted for specific results, this model uses pooled data with [Speaker] added as a random independent variable. This computational method uses the [Speaker x Variable] interaction as the error term in the test for [Variable] as described by Winer (1971) to provide control over individuals between experimental units (see Choi, 1992). (Thus, the degrees of freedom reported for the error term is that of the [Speaker x Variable] interaction.) The SuperAnova and Statview packages (Abacus Concepts, 1989) are used to perform the statistical tests. The following sections specific to each experiment will describe the quantitative variables tested. Because variability is also of interest in this dissertation, the Levene statistic for testing equal variability will also be employed (Levene, 1960 cited in Dixon, 1988). This statistic uses the absolute values of the deviations from the group means as data. The deviations were calculated here separately for each subject using his or her mean. The Levene F statistic was then computed as a one-way ANOVA F using the computational method for repeated-measures outlined above. It is recognized that one of the assumptions underlying the ANOVA is that variance within each of the treatment groups is homogeneous. However, F-tests are in fact quite robust to departures from homogeneity of variance (Winer, 1971). In instances showing a significant main effect on variability as determined by the Levene statistic and in means as determined by ANOVA for the experiments below, the means for the treatment levels will be given for comparison purposes.

Experiment Three, modeled on Browman and Goldstein (1988), compares the standard deviations of the various intervals from the anchorpoints. Any other analyses are detailed in Chapter Five. Experiment Four regresses the independent variable of rate against various measures of latency, overlap, consonant duration, and amount of lingua-palatal contact. Results with speakers both pooled and separated are reported with r^2 and significance levels for linear fits.

2.8 **WHAT WILL NOT BE FOUND HERE**

Although this dissertation is designed in the hope of covering many aspects of consonant sequence production, unfortunately not all potentially interesting processes can be pursued. Stress will not be a subject of investigation. Likewise, the process of gestural blending, whereby two or more movements using the same active articulator interact to yield some compromise movement, while interesting, will not be included here. This proviso includes the post-alveolar affricates found in English. As noted above, main effects on spatial and temporal coproduction of the consonant sequences will be the core of the discussion because so little has been described about them in the literature to date.

CHAPTER THREE: EXPERIMENT ONE

3.0 SEQUENCES WITH ONE LINGUAL CONSONANT

The first sequences we consider are those with only one lingual consonant. These include sequences with geminated consonants--dd, ss, and gg--which are realized with a single raising and lowering of the tongue, and those in which one lingual consonant occurs with a labial consonant--db, bd, sb, bs, gb, and bg. In all these cases, only a single consonant articulation is recorded on the palate, and no other lingual constriction interferes with it. Seven tokens of each sequence from each of the five speakers are analyzed. The effects of syllable position and consonant identity on the displacement and duration of the lingual consonant are tested. These results will later be compared to the sequences of two lingual consonants, in Section 3.1.

The contact profiles for the front region for dd (geminated), db (coda), and bd (onset) are shown for each speaker separately in Figures 3.1-3.5. All are aligned at the onset of contact in the front region. Contact then increases as the tongue and jaw raise and decreases as they lower over time. In order to preserve information about variability both within and across speakers, the tokens are not averaged. The y-axis represents the percent of the region being contacted at a particular point in time. The x-axis represents time in frames of .01 seconds. The axes cover the same ranges for all figures in this chapter in order to facilitate comparisons. Important items to note in the contact profiles include the overall degree of tongue displacement for a consonant, its duration, and its variability in both the x-plane (time), and the y-plane (displacement). The null hypothesis, clearly not upheld, is that the three contact profiles--geminate, coda, and onset-for a consonant are basically the same.

FIGURE 3.1 SPEAKER A: CONTACT PROFILES IN THE FRONT REGION FOR dd, db AND bd

FIGURE 3.2 SPEAKER B: CONTACT PROFILES IN THE FRONT REGION FOR dd, db AND bd

 $\bar{\lambda}$

FIGURE 3.3 SPEAKER K: CONTACT PROFILES IN THE FRONT REGION FOR dd, db AND bd

FIGURE 3.4 SPEAKER M: CONTACT PROFILES IN THE FRONT REGION FOR dd, db AND bd

FIGURE 3.5 SPEAKER S: CONTACT PROFILES IN THE FRONT REGION FOR dd, db AND bd

$3.0.1$ **COMPARISON OF MEANS**

First let's consider differences in the amount of contact in the front region for **d**. Recall that this is the maximum or peak contact expressed as a percentage of the total possible contact in the front region. ANOVA determines there to be a significant effect of sequence on the maximum front contact for **d** (F(2,8)=5.75, p=.0283) such that the coda **d**'s have lower peaks than \mathbf{d}' 's in the other two sequences: 44% of the region contacted for \mathbf{d} versus 56% for \mathbf{d} and 55% for bd. This parallels results reported elsewhere on reduction of tongue tip gestures in word final position with both EPG and X-ray microbeam data (e.g. Barry, 1992 and Browman and Goldstein, in press a). There is also a significant interaction of Speaker and this effect. As can be seen from the contact profiles, all speakers except Speaker B follow this pattern. We will continue to see throughout Chapters Three and Four that in many cases where there is a single exceptional speaker, it is Speaker B. In this case, not only are Speaker B's codas fortified to match her and the other speaker's onsets, but also her onset **d**'s are lenited almost as much as the other speaker's codas--43% maximum contact for onset bd and 52% for coda db. The other speakers excluding Speaker B have a mean of 58% contact for the onset d's and 42% for codas.

Next consider duration of contact in the region. The duration of linguapalatal contact for **d** is longer in dd than for the other sequences ($F(2,8)=12.55$, $p=.0034$). For three Speakers, A, K, and M, codas were shorter than onsets as well. Let's consider the shape of the profile, or temporal distribution of contact. One measure of this is skew. Roughly speaking, a greater (positive) skew indicates a shorter time from contact onset to peak (closure formation) in the contact profile than from peak to the final contact (closure release). For these same three speakers, codas also had a greater positive skew than onsets. There was a significant interaction of sequence and speaker in affecting skew $(F(8,90)=19.460, p=.0001)$. This can be seen particularly in the contact profiles for Speaker K. The time taken in forming the contact was shorter than that needed for the release (i.e. positive skew) for both coda and onset d's with the asymmetry being greater in coda position. Another measure of shape is the flatness of the contact profile as indexed by the mean contact divided by the maximum contact, MEAN/MAX. There was a significant interaction $(F(8,90)=6.464, p=.0001)$ of speaker and sequence on flatness. All speakers had flatter dd profiles than they had for the other sequences. Three speakers, K, A, and B, also had flatter onsets than codas.

The next sequences we consider are ss, sb, and bs. The contact profiles for these are shown for all speakers in Figures 3.6-3.10.

FIGURE 3.6 SPEAKER A: CONTACT PROFILES IN THE FRONT REGION FOR ss, sb AND bs

FIGURE 3.7 SPEAKER B: CONTACT PROFILES IN THE FRONT REGION FOR ss, sb AND bs

FIGURE 3.8 SPEAKER K: CONTACT PROFILES IN THE FRONT REGION FOR ss, sb AND bs

FIGURE 3.9 SPEAKER M: CONTACT PROFILES IN THE FRONT REGION FOR ss, sb AND bs

FIGURE 3.10 SPEAKER S: CONTACT PROFILES IN THE FRONT REGION FOR ss, sb AND bs

First notice the amount of contact in the front region. ANOVA shows there to be no difference among the sequences ss, sb, and bs, in maximum contact in the front region. Next, duration of contact is of interest; there are significant differences in duration $(F(2,8)=18.35, p=.001)$. The geminated sequences have the longest durations of contact, and, excepting Speaker S, onsets are longer than codas. For all speakers, contact profiles for codas are less flat and have a greater skew than either the geminated or onset consonants, as determined by the main effect of sequence on MEAN/MAX (F(2,8)=8.125, p=.0118) and SKEW (F(2,8)=16, p=.0016). In fact the coda s was the only one of the three s's to have a positive skew for all subjects; only two subjects had a positive skew for the onset s. This parallels the findings for d, suggesting that while coda s's may not undergo spatial lenition, they, like coda d's, are shorter and have faster closure formation than the onsets.

Now we turn to the single dorsal consonant, the stop g . The sequences we consider are gg, gb, and bg. Here the relevant articulatory region is the back one. The contact profiles for these are shown for all speakers in Figures 3.11-3.15.

FIGURE 3.11 SPEAKER A: CONTACT PROFILES IN THE BACK REGION FOR gg, gb AND bg

FIGURE 3.12 SPEAKER B: CONTACT PROFILES IN THE BACK REGION FOR gg, gb AND bg

FIGURE 3.13 SPEAKER K: CONTACT PROFILES IN THE BACK REGION FOR gg, gb AND bg

FIGURE 3.14 SPEAKER M: CONTACT PROFILES IN THE BACK REGION FOR gg, gb AND bg

FIGURE 3.15 SPEAKER S: CONTACT PROFILES IN THE BACK REGION FOR gg, gb AND bg

ANOVA of these sequences shows there to be a significant effect of sequence on maximum displacement in the back region $(F(2,8)=5.476, p=.0318)$. For all speakers, except Speaker B,

displacements decrease from onset to geminated to coda **g**'s. For Speaker B, onsets rather than codas have the lowest maximum contact; codas are still less displaced than geminated **g**'s. There is also a significant difference between the duration of **g** in the three sequences $(F(2,8)=9.51)$, p=.0077). All speakers' geminated sequences are longest, and all, except Speaker B, have longer onset **g**'s than coda **g**'s. There is also a significant effect on the flatness of the contact profile, MEAN/MAX, $(F(2,8)=5.788, p=.0279)$. Contact profiles for all speakers except Speaker A are flatter for codas than onsets. There is no main effect on SKEW for the velar consonant, although there is a significant interaction of Speaker and sequence $(F(8.90)=8.511, p=.0001)$ with speakers K, A, and S having onsets more skewed to the right than codas, although most speakers' skews for all three sequences are negative.

The pooled-subject means for these six sequences are shown below:

TABLE 3.1: POOLED-SUBJECT MEANS FOR PEAK, DURATION, SKEW, AND MEAN/MAX FOR db, bd, db, sb, bs, $ss, gb, bg, gg.$

However, recall that Speaker B has codas and onsets which generally pattern opposite to those of the other speakers; this makes an examination of the group means less useful. Graphical comparisons are shown below to illustrate differences in reduction and shortening:

FIGURE 3.16 MAXIMUM PERCENT CONTACT IN REGION FOR dd, bd, db, ss, bs, sb, gg, bg, gb.

FIGURE 3.17 DURATION OF CONTACT IN REGION FOR dd, bd, db, sb, ss, bs, sb, gg, bg, gb.

In summary the data presented thus far in this chapter indicate that:

•For the stop consonants, onsets (bC) are generally more displaced (i.e. have greater maximum contact and summed area of contact during the consonant) than codas (Cb).

This supports hypothesis three (H3)--that movements in coda position are smaller than movements in onset position--in part. Its support extends only for the case of stop consonants.

•For all consonants considered, geminated sequences (CC) are longer than either onsets (bC) or codas (Cb) in sequences with a neighboring **b**.

•For s, contact in onset position (bC) is longer than in coda position (Cb) .

• For all consonants considered, geminated sequences (CC) have flatter contact profiles than codas or onsets (bC) . For front consonants, they are also less skewed.

•Contact profiles for front consonants are flatter and less (positively) skewed in onset position (bC) than in coda position (Cb). This is true across speakers for s and for a majority of speakers for **d**.

•For the back stop consonant, contact profiles are flatter in coda position (Cb) than in onset position (bC). The majority of speakers also demonstrate contact profiles which are less (positively) skewed in coda position than in onset position.

$3.0.2$ **DISCUSSION**

Munhall and Löfqvist (1992) examined the blending of two larnyngeal gestures as a function of speaking rate. This situation is somewhat analogous to our geminated sequences where two lingual gestures are canonically present. They observed that a single smooth movement occurred as two laryngeal gestures overlapped at fast speaking rates. Their laryngeal data are in accordance with our geminated lingual stop sequences. Both data sets show the coproduced movement in the geminated sequences to be longer than a single non-coproduced movement. Additionally, they found no consistent tendency for the combined single movement to be larger than an individual (non-coproduced) movement, although the simulated summation of two movements predicts such a difference. At medium speech rates one of their speakers showed larger "geminated" movements but this behavior reversed at fast rates. Their other speaker showed no consistent difference. Our data above shows no consistent increase in maximum contact for the geminated consonants, suggesting that a summation process is not at work. Similarly, Kelso and Tuller (1987) cited (in prepublication form) in Browman and Goldstein (1986) report that larger gestures typically have both increased amplitude and steeper onset and offset slopes. Partly on this basis, Browman and Goldstein (1986) argue for the presence of two overlapping bilabial closure gestures in a [mp] sequence because of the similarity of slope and amplitude to a single bilabial closure gesture. This of course results in a longer overall movement. We observe a similar pattern of matching slopes and amplitudes with longer duration for our geminated [C#C] sequences as compared to the single consonants.

The data described above is also relevant in considering Barry's (1992) model of lenition of coronal consonants in which the tongue tip is modeled as a massless articulatory subsystem while the tongue body as modeled as having mass greater than zero. (An additional difference of stiffness in gestural descriptors for tongue tip and tongue body consonants has been discussed by Byrd, 1992. A mass difference is not incompatible with a stiffness difference.) Barry's model predicts a tongue tip gesture to be capable of a more rapid change in the direction of movement after maximum displacement. The tongue body, in contrast, is predicted to respond more slowly to the cessation of activation "with something resembling a plateau in the gestural trajectory" (Barry, 1992, p. 398). Due to this sluggishness, the tongue body movement should therefore appear flatter than the tongue tip movement. Our EPG measure of flatness, MEAN/MAX, provides an approximation of this flatness. In this model, the tongue tip also approaches its maximum displacement in a shorter time that it takes to move away from it, even though it does change direction rapidly at its peak. Tongue body movement is predicted to be much more symmetric in this respect. The gestural trajectories presented by Barry have a greater positive skew for the tongue tip movement. Our EPG measures of SKEW is a metric for this difference.

If it were the case that tongue tip gestures have a lower mass than tongue body gestures as suggested by Barry, we would make two predictions. First, contact profiles for **d** and **s** should have similar skews and flatnesses. Second, g should be flatter than the coronal consonants.

Let's evaluate these predictions. First, compare **d** and **s**. This comparison illuminates differences due to manner, *i.e.* constriction degree. (Other differences in tongue shape exist too.) Analysis of the sequences db, bd, sb, and bs (coded as having a lingual stop or a lingual fricative) determines there to be no significant main effect of manner on MEAN/MAX. There is a significant interaction of speaker and manner, but this doesn't seem to be the result of any one speaker. There is a main effect of manner on skew, however $(F(1,4)=7.767, p=.0495)$. All speakers except Speaker B had a greater skew for **d** than for s. Next, to determine how the tongue tip consonants as a group compare to the back consonant, the sequences db, bd, sb, bs, gb, and bg were tested for main effects of place-front versus back-on the measures of skew and flatness. While not reaching significance, there is a trend $(p=.1059)$ exhibited by all speakers, and a significant interaction of effect and speaker $(F(4,200)=13.429, p=.0001)$, for the back consonant to be flatter than the front consonants. This is the result predicted by Barry's model. Lastly, in accordance with Barry's model front consonants were more skewed than back consonants ($F(1,4)=0.0141$, $p=.0141$). Thus it appears that Barry's model in which the tongue tip articulator subsystem, but not the tongue body subsystem, is modeled as massless does capture differences in the contact profiles for front and back consonants. But it cannot be extended to account for effects of manner, *i.e.*, why **d** should be more skewed than **s**. Nor can Barry's hypothesized difference in dynamics capture other influences we have observed on SKEW and flatness.

We have also seen that both SKEW and MEAN/MAX differ for the front consonants depending on their syllable position, with codas tending to be more skewed and less flat. For the back consonant, though, effects tend in the opposite direction-onsets are less flat and, for three speakers, have a skew closer to zero. All consonants have longer contact for onsets than codas, and the two stops have codas which are less spatially displaced than onsets. Barry (1992) suggests that the mass difference for tip and body may be a cause of coronal lenition. However, we have seen here that both $\mathbf{d}'s$ and $\mathbf{g}'s$ are less displaced in coda than in onset; so there must be something else at work here as well; *i.e.* an effect of syllable position apart from the effect of articulator. Similar results will be discussed in Section 3.1 and in Chapter Six as well. Compare also Browman and Goldstein's (in press a) similar findings for [1]. If we were to pursue mass differences to explain this behavior, we would have to hypothesize different masses for onsets and codas using the same articulatory subsystem. Presumably, however, physical mass is constant, and some other gestural descriptor such as stiffness must alter according to syllable Beckman, Edwards, and Fletcher (1992) account for intonation-phrase final position. lengthening by positing just such a stiffness difference for closing gestures, with decreased stiffness yielding longer durations. We suggest on the basis of the data in this section that in addition to a mass and stiffness difference between tip and body articulators, the gestural descriptor of stiffness also differs for codas and onsets; the shorter codas having a greater stiffness. In order to explain the *spatial* lenition of coda position, Browman and Goldstein (in press a) suggest that it might be related to decreasing effort toward the end of a word. The satisfaction offered by this explanation would depend on the results of a comparison between word-final codas, word-medial codas, and other phrase-final codas. The same explanation could be offered regarding the end of the syllable. In summary, **d** is special because its articulatorbased effect-e.g. low mass--interacts strongly with the separate coda effect (or effects)--e.g. increased stiffness and/or decreased effort.

We propose that differences in skew and flatness for the front consonants are due to coarticulatory effects with the adjacent bilabial consonant and vowel. Recall that more positive skews go hand-in-hand with less flat peaks. In coda position, the consonant is followed by another consonant. This pre-consonantal context (Cb) does not require as rapid a lowering of the jaw as for the prevocalic onsets (bC) . Because the jaw can take longer moving away from peak displacement in the codas, the contact profiles are less symmetrical than in onsets. This yields a greater positive skew. However, the change in direction at the peak can occur fairly rapidly, creating particularly peaked, as opposed to flat, profiles for the front coda consonants. For the velar consonants, the effects of this coarticulation differ due to the sluggishness of the tongue body in comparison to the tongue tip. Because the tongue body does not change direction very rapidly, effects of SKEW and MEAN/MAX do not go hand-in-hand for g. Additionally, changes in jaw height have a minimal influence on the tongue body as compared to the tongue tip due to the posterior hinge location of the jaw. Here, the following contexts don't have as strong an influence on skew. The **g** in both contexts had fairly symmetrical contact profiles with approach

taking about the same amount of time as release (i.e., skew near zero). Because the onset consonant is immediately followed by a vowel having a low tongue position, the tongue must move down yielding a contact profile for g in onset (b#gæ) that is somewhat less flat than in coda (g#b). The difference in flatness here would be a result of the gestural blending of the two competing demands on the tongue body for **g** and [æ].

3.1 HETEROSYLLABIC SEQUENCES WITH TWO LINGUAL CONSONANTS

The following section considers the sequences dg, gd, sg, and gs, sequences where the two consonants are formed on different parts of the palate with different parts of the tongue. Seven tokens of each sequence from each of the five speakers are analyzed. We evaluate effects of order, i.e. front-back versus back-front, and manner, e.g. stop-stop versus fricative-stop, on both the contact profiles for individual consonants and the latency between the two consonants. In the following results, main effects are reported. Additionally, if there is no significant main effect but four of the five speakers show the same direction of effect, the main effect is recalculated excluding the exceptional speaker and reported when significant. The excluded speaker is noted. This procedure was adopted to ensure that robust behaviors were not ignored simply because a single speaker's means deviated hugely from the others. Probabilities less than or equal to .05 are considered significant; probabilities less than or equal to .08 are noted as nonsignificant trends $(n.s. trend)$.

$3.1.1$ **COMPARISON OF MEANS**

In considering effects of order, two analyses are conducted. The first uses all four sequences dg, gd, sg, and gs coded as front-back (dg and sg) or back-front (gd and gs). The second looks at the specific differences between dg and gd and between sg and gs. (Where an effect obtained in the analysis of all four sequences is determined by the later analyses to be present for just one of the pairs, *i.e.*, is dependent on the manner of the front consonant, it is only reported in the discussion of specific pairs of sequences.) This approach is adopted because of the inappropriateness of post-hoc tests for repeated measures analyses.

First, let's consider significant differences between the articulation of the consonants and their coordination in the analysis of all four sequences. A consideration of effects of sequence order on dg, gd, sg, and gs shows differences in the amount of contact in each order, front-back and back-front. First, consider the back region. Maximum back region contact and area under the back region contact profile behave similarly. For all speakers except Speaker B, g has lower means in coda, *i.e.* more reduction in the back-front order than the front-back order (with Speaker B excluded, for back region MAXIMUM contact $F(1,3)=20.489$ p=.0202 and for total area of contact during $g F(1,3)=15.444$, p=.0293). The front-back order has a mean back maximum contact of 71%; the reverse order has a mean of 56%. Thus contact in the back region was less for $\underline{\mathbf{c}}$ than for $C\underline{\mathbf{c}}$. The back region maximum in front-back order is approximately that found in the geminated **gg** sequences.

Syllable position also has a main effect on the front consonants, d and s. The significant difference in area under the front contact profile $(F(1,4)=12.27, p=.0248)$ shows that front consonants lenite in coda position. This effect of order again shows the coda consonants in Cg to have less contact than onsets in $\underline{\mathbf{g}}\underline{\mathbf{C}}$. We see below that **s** does not contribute to the effect; here again, the main effect on the maximum front contact is obtained due to the difference between dg and **gd**. As seen above, the fricative s does not appear to be susceptible to lenition in peak displacement in the way that the alveolar stop is.

Now we can also look at coproduction because we can observe both lingual consonants in the different regions of the palate. In considering differences in coproduction and consonant latency we again observe effects of sequence order. SEQUENCE OVERLAP, measuring the proportion of the sequence during which contact in both regions occurred, is significantly more greater in front-back sequences than in back-front sequences $(F(1,4)=10.922, p=.0298)$. We see correspondingly that the duration of the sequence is significantly longer for the less overlapped back-front sequences ($F(1,4)=9.963$, $p=.0343$). The tongue tip consonants were significantly more overlapped by a following tongue body consonant than the reverse. This is shown by the main effect of sequence order on what percent of C1 contact was overlapped by C2 contact $(F(1,4)=22.244, p=.0092).$

Differences in contact profiles for an individual consonant and in latencies between consonants are examined in more detail for the paired sequences dg and gd and sg and gs. Let's consider first the stop-stop sequences. The contact profiles for these sequences are shown in Figures 3.18 through 3.27 and are presented in pairs for each speaker in order to facilitate comparisons of the sequences. The front-back sequence appears on top and the back-front on the bottom. Here, both consonants composing each sequence are exhibited in the contact profiles. They are shown as they actually occur over time, starting at the frame immediately before the first regional contact for the consonants to the frame immediately following the last regional contact for the consonants. There is a time latency shown in the figures such that contact for the second consonant generally starts after that for the first. If there were no overlap, we would expect in the front-back sequences (top figures) to see the profile with circles (front region contact) rise and fall followed by the profile with squares (back region contact) rise and fall. The bottom figures for the back-front sequence would show profiles with squares before those with circles. However, notice below the extent of the overlap actually observed and the variability in the time of onset for C2 relative to C1.

FIGURE 3.18: CONTACT PROFILES FOR dg FOR SPEAKER A

FIGURE 3.19: CONTACT PROFILES FOR gd FOR SPEAKER A

FIGURE 3.20: CONTACT PROFILES FOR dg FOR SPEAKER B

FIGURE 3.21: CONTACT PROFILES FOR gd FOR SPEAKER B

FIGURE 3.22: CONTACT PROFILES FOR dg FOR SPEAKER K

FIGURE 3.23: CONTACT PROFILES FOR gd FOR SPEAKER K

FIGURE 3.24: CONTACT PROFILES FOR dg FOR SPEAKER M

FIGURE 3.25: CONTACT PROFILES FOR gd for SPEAKER M

FIGURE 3.26: CONTACT PROFILES FOR dg FOR SPEAKER S

FIGURE 3.27: CONTACT PROFILES FOR gd FOR SPEAKER S

Generally we observe that the coda d's are relatively small and variable and the onset of C2 (relative to the onset of C1) is later for gd than for dg . In fact, for most of the speakers, the

consonants of dg are nearly completely overlapped with contact for g often starting synchronously with that for **d**.

The table below summarizes the results of the ANOVA comparing dg and gd.

TABLE 3.2: A SUMMARY OF THE SIGNIFICANT EFFECTS IN THE STATISTICAL ANALYSIS OF dg AND gd

These results demonstrate that both consonants have less displacement when in coda position. Additionally, as seen in the earlier **db** and **bd** sequences, the **d**'s are more skewed when in coda position. The peak displacements differ in the two sequences, having the following means: dg--40% d, 78% g and gd--61% g, 65% d.

Next, consider the coproduction relations between the two consonants: overlap of various points and total duration. As is evident from even a casual inspection of the contact profiles, dg sequences are substantially more overlapped than gd sequences--with some dg tokens even having initial contact for the back region precede that in the front region. This timing difference causes the effects detailed in the lower half of Table 3.2. The percentage of the sequence during which contact in both regions occurs (SEQUENCE OVERLAP) and the overlap of C1 are both much greater for dg than for gd. The mean SEQUENCE OVERLAP is 59% for dg and 46% for gd. The mean C1 OVERLAP is 87% for dg and 53% for gd. The difference in coproduction is also evidenced by the greater latency between onsets for **gd** and the fact that the C2 starts much later relative to C1 for gd . The mean value for the time between onsets was .01s (1 frame) for dg and .07s (7 frames) for gd. The measure of relative latency, C2 ONSET RELATIVE TO C1, has means of 8% for dg and 46% for gd. Because both consonants are being produced almost simultaneously in many dg cases, the duration of the d in dg is a greater percent of the total sequence duration. Likewise, the total sequence duration is shorter for $dg(0.15s)$ versus .17s). This is not surprising in light of the extensive overlap in this sequence. Consider the schematic in Figure 3.28 which is to scale for the *pooled* mean values of Δ ONSETS, FRONT DURATION, and BACK DURATION only.

FIGURE 3.28: A SCHEMATIC OF THE TIMING RELATIONSHIP FOR dg AND gd; TO SCALE FOR THE POOLED MEAN VALUES OF ΔONSETS, FRONT DURATION, AND BACK DURATION ONLY

While the durations for each consonant are not significantly different in the two sequences, the shortness of the **d** causes it to be proportionally more overlapped than **g** regardless of its syllable position. The absolute latencies also differ substantially. Virtually all of the contact for **d** occurs during that for **g**, except for a slight latency between the onsets of C1 and C2 contact which preserves canonical order. While gd has a greater latency between onsets, it has a slight offset between the moments of *final* contact.

The next pair of sequences include the coronal fricative and the dorsal stop: sg and gs. The contact profiles are presented in pairs below. As with those above, differences in contact profiles for an individual consonant and in latencies between consonants are examined. The contact profiles for these sequences are shown in Figures 3.29 through 3.38 and are presented in pairs for each speaker in order to facilitate comparisons of the sequences. The front-back sequence appears on top and the back-front on the bottom. Both consonants composing each sequence are exhibited in the contact profiles.

FIGURE 3.29: CONTACT PROFILES FOR sg FOR SPEAKER A

FIGURE 3.30: CONTACT PROFILES FOR gs FOR SPEAKER A

FIGURE 3.31: CONTACT PROFILES FOR SP FOR SPEAKER B

FIGURE 3.32: CONTACT PROFILES FOR gs FOR SPEAKER B

FIGURE 3.33: CONTACT PROFILES FOR SP FOR SPEAKER K

FIGURE 3.34: CONTACT PROFILES FOR gs FOR SPEAKER K

FIGURE 3.35: CONTACT PROFILES FOR sg FOR SPEAKER M

FIGURE 3.36: CONTACT PROFILES FOR gs FOR SPEAKER M

FIGURE 3.37: CONTACT PROFILES FOR SP FOR SPEAKER S

FIGURE 3.38: CONTACT PROFILES FOR gs FOR SPEAKER S

The analysis of variance comparing sg and gs yields many of the same fortition and timing effects seen for the stop-stop sequences. These results are summarized in Table 3.3.

Dependent Variable	Confidence Level $F(1,4)$; $F(1,3)$ when 1 Sp. excluded	Description		
	Individual Consonants			
BACK MAXIMUM	$F=16.18$, p=.0276	B sg > g s, Speaker excluded		
FRONT AREA	$F=29.597, p=.0055$	gs > sg		
BACK AREA	$F=8.149, p=.0462$	sg > gs		
NONOVERLAPPED FRONT AREA	$F=11.902$, p=.0261	gs > sg		
BACK DURATION	$F=8.142$, p=.0649 n.s. trend	sg > gs , Speaker S excluded		
	Coproduction/Latency			
BACK DURATION (%)	$F=14.96$, $p=.018$	sg > gs		
SEQUENCE DURATION	$F=13.073$, p=.0364	sg, Speaker K gs > excluded		
SEQUENCE OVERLAP $(\overline{\%})$ F=14.601, p=.0188		sg > gs		
$\overline{\text{CI OVERLAP}}$ (%)	$F=9.650, p=.0530$ n.s. trend	sg > gs , Speaker A excluded		
$C2$ OVERLAP $(\%)$	$F=67.755$, p=.0012	sg > gs		
\triangle RELEASE OF CI TO ONSET OF C ₂	$F=14.678$, p=.0313	sg > gs, Speaker M excluded		
C ₂ ONSET RELATIVE TO $C1(\%)$	$F=9.650$, p=.0530 n.s. trend	$sg > gs$, Speaker A excluded		
C2 PEAK RELATIVE TO $C1$ (%)	$F=16.573$, p=.0152	gs > sg		

TABLE 3.3: A SUMMARY OF THE SIGNIFICANT EFFECTS IN THE STATISTICAL ANALYSIS OF sg AND gs. n.s. trend DENOTES A NON-SIGNIFICANT TREND OF $\leq .08$.

All the effects in common between the stop-stop sequences and sg and gs show the same direction of asymmetries between front-back and back-front. The **g** contact is greater in onset position with a peak of 64% in sg and 50% in gs. There is a tendency for both s and g to shorten in coda position. s and g are shorter in coda than onset for all speakers except Speaker S. Unlike the bs and sb sequences, three speakers do show somewhat lower maximum contact for s in onset position as compared to coda position. This tendency to reduce s and g in coda is apparent in the significant differences between onset and coda position in contact area which is a cumulative summation of contact during a consonant thus affected by both duration and displacement.

Like the stop-stop sequences, these sequences are significantly more overlapped when the front consonant precedes the back than the reverse. Consistent differences in absolute latency of C2 are not so readily apparent. The pooled means for the timing measures are given below; the stop-stop means are included for comparison.

	sg		gs	dg		gd
SEQUENCE DURATION	.20s	¥	.22s	$\overline{.15s}$	tr.	.17s
SEQUENCE OVERLAP	44%	∗	27%	59%	tr.	$\overline{46\%}$
CI OVERLAP	55%	tr.	50%	87%	\ast	$\overline{53\%}$
C ₂ OVERLAP	66%	∗	37%	62%		80%
AONSETS	.07s		.06s	.01s	*	.07s
ARELEASE OF C1 TO ONSET OF C2	$-.01s$	\ast	$-.03s$	$-.04$		$-.03$
C ₂ ONSET RELATIVE TO	45%	tr.	50%	8%	\ast	46%
C ₂ PEAK RELATIVE TO	82%	\ast	99%	86%		77%

TABLE 3.4: A SUMMARY OF POOLED MEANS FOR gd, dg, gs, AND sg; * INDICATES A SIGNIFICANT DIFFERENCE, tr. INDICATES A TREND (P<.08).

Recall that the stable phasing relationship proposed by Browman and Goldstein (1990b) is the synchronization of the onset of $C2$ with the release of $C1$. Our analogous measure is the interval between the first frame after the last peak contact for C1 to the first contact for C2, ARELEASE OF C1 TO ONSET OF C2. Of course our observed onset will be later than the gestures' actual onset (as is true to a greater or lesser degree with all kinematic measurements, movement tracking or otherwise.) However, stability or lack of stability in this relationship should be observable regardless. This measure shows no consistent difference between dg and gd. (Although there is a significant interaction of main effect and speaker with three speakers having greater intervals for **gd**.) This measure however does distinguish s**g** and **gs** which are not distinguished by \triangle ONSETS as dg and gd are.

In addition to comparing differences in gestural magnitude and duration of the consonants and effects of sequence order on timing, we also presented a hypothesis regarding the effects of manner on overlap. This hypothesis was that a closure (compression) movement is more temporally overlapped by a following consonant articulation than is a fricative (constriction) movement. An ANOVA including the sequences dg and sg determines that the stop-stop sequence exhibits significantly more overlap of C1 by C2 ($F(1,4)=12.327$, $p=.0247$). The onset of C2 relative to C1 is earlier for dg (F(1,4)=14.141, p=.0198) and the time between onsets is less ($F=34.778$, $p=.0041$). There is also a trend for the interval from the RELEASE OF C1 TO ONSET OF C2 TO be greater for sg than dg $(F(1,3)=10.095, p=.0502,$ Speaker B excluded). These measures all indicate greater coproduction in the stop-stop sequence. If we consider the effect of the manner of C2 for the sequences in the reverse order-gd and gs--we do find somewhat parallel results. The proportion of the sequence during which contact for both consonants occur, SEQUENCE OVERLAP, and the percent of C2 OVERLAP both indicate greater overlap in gd than gs. The sequence duration is also longer for gs, again in accordance with the lesser overlap. Four of the five speakers show a non-significant trend for **gs** to have longer absolute latencies of C2 (\triangle PEAKS and \triangle RELEASE OF C1 TO ONSET OF C2) than gd, (Speaker B

being exceptional). The more general conclusion is that the sequences including a fricative were less overlapped than those having only stops.

In summary, the comparison of means for the within and between consonant effects for the heterosyllabic sequences gd, dg, gs, and sg demonstrate the following findings (comparisons with the bd, db, bs, sb, bg, and gb sequences are also noted):

Syllable Position Effects:

•In stop-stop sequences, like gd, dg, bd, db, bg, and gb, the stop articulation is smaller in coda position than in onset position.

This supports hypothesis three $(H3)$ --that movements in coda position are smaller than movements in onset position--in part. Its support extends only for the case of stop consonants.

•In sequences of two lingual stops, like **gd** and **dg**, stops in coda position do not significantly differ in duration from those in onset position. In sequences with a labial stop, like bg, gb, db, and bd, however, onsets are longer than codas.

•In stop-fricative versus fricative-stop sequences, like gs and sg, linguapalated contact for tongue body stops is both shorter and smaller in coda position than in onset position. This is also true in sequences where g occurs with a labial stop, like gb and bg.

This again supports hypothesis three $(H3)$ in part-- the case of stop consonants.

•Sequences including the alveolar fricative and a stop, like sg, gs, bs, and sb, show some tendency for the fricative to be shorter and smaller in coda position as compared to onset position. In the sequences with a labial stop, duration is affected; in sequences with a lingual stop, both displacement and duration are mildly affected.

This supports hypothesis three (H3) to a certain degree for the case of the fricative consonant, s.

•Linguapalatal contact profiles for **d** are more positively skewed in coda position than in onset position, both in sequences of two lingual consonants like dg and gd and in those with a labial and a lingual consonant like bd and db. s is also more skewed in coda position in bs and sb sequences.

•Linguapalatal contact profiles for the tongue tip consonants, both **d** and **s**, are flatter in onset position than in coda position, both in sequences of two lingual consonants and in those with a labial and a lingual consonant

Consonant Articulator Effects:

•Linguapalatal contact profiles show a tongue tip consonant to be more overlapped by a following tongue body consonant than a tongue body consonant is by a following tongue tip consonant.

•The above effect is found for both tongue tip gestures--stop and fricative.

These findings confirm hypothesis one (H1) which proposed that overlap varied in this way.

•Contact profiles show an alveolar fricative to be less overlapped than an alveolar stop by a following velar stop. More generally, sequences involving a fricative are less overlapped than those which only have stops.

This supports hypothesis two (H2) which proposed that overlap would differ in this direction.

$3.1.2$ **DISCUSSION OF CENTRAL TENDENCIES**

The differences between codas and onsets in contact duration are in accordance with acoustic closure duration differences for word initial versus word final stops (Lehiste, 1960). Word initial b's have also been found to have both longer and greater muscle activity associated with them as compared to word final b's (Fromkin, 1965). Catford (1977:222) estimates, using EPG, the degree of articulatory overlap in CC sequences as being between 29% and 45% of the total sequence duration. However, he does not report the number of subjects studied or the number of repetitions. We observed *means* of between 27% and 59% overlap of contact for two consonants in sequence. If Catford was using total contact duration to indicate sequence duration, which does not seem unreasonable, we conclude that we observe generally more overlap than Catford. We also observe a much wider *range* of overlap than that observed by Catford, with individual tokens ranging from 11% to 91% overlapped.

We also saw above that d in either C1 or C2 position was more overlapped than s in the same position. Why should this be so? There may be a perceptual motivation for this difference¹. The abrupt discontinuities at the edges of a stop consonant are important perceptual cues in recovering that consonant (Stevens and Keyser, 1989). For a fricative however, abrupt discontinuities or edges are not as important as the frequency distribution of the frication noise itself. It may be that there is a perceptual motivation for not obscurring too much of the s duration. (Stevens and Keyser (1989) note the importance of the acoustic consequences of a slower release.) However, for a d, only an edge need be discernible as the abrupt change in amplitude is the perceptually salient cue. This would create a situation in which it is important not to overlap too much of the s, while **d** can withstand being more overlapped. Ladefoged, DeClerk, Lindau and Papcun (1972) and Johnson, Ladefoged and Lindau (1993) have outlined an auditory theory of speech production in which speech movements are directed by auditory goals and have also gone on to suggest that "the acoustic product of speaking is the crucial determinant of the *organization* of speech articulation" (Johnson, *et al.*, 1993, emphasis added). In Chapter Seven we will expand on the possibility that acoustic/perceptual goals may influence inter-articulator phasing. I agree with Johnson et al. to the extent that a model of speech organization must allow for acoustic and perceptual goals to be *an* influence on articulatory

¹I thank Ian Maddieson for suggesting this issue to me.

timing. The difference in s and **d** overlap serves as an example of how this type of influence may operate. Further evidence for such an influence might be found in an understanding of why the front-back sequences were more ovelapped than the back-front sequences.

With respect to the timing asymmetries outlined for sequences of the same gestures in opposite orders, Hardcastle and Roach (1979) found that the time between the initiation of contact for an adjacent [t] and [k] was shorter for a [tk] cluster than for a [kt] cluster. They concluded that this was because the movement from [t] to [k] involves the contraction of a single intrinsic tongue muscle to raise the back of the tongue while the movement from [k] to [t] requires the use of two muscles including the extrinsic genioglossus to reposition the tongue upwards and forwards. Recasens et al. (1993) suggest that one reason for the difference found by Hardcastle and Roach could be that the tongue tip has a greater "degree of flexibility" and that this causes greater anticipatory coarticulation--although, the exact connection between these assertions is not clear. In our data, the rather large difference observed in our data between dg and gd, in which contact for the consonants often starts nearly simultaneously in dg, does not seem to support Hardcastle and Roach's explanation. The difference in time taken to contract one intrinsic muscle versus one extrinsic and one intrinsic muscle is likely to be incredibly short. In fact, it is not even clear that the assumption that using two muscles takes longer than using one is valid. The well-supported conception of muscle groups organized into coordinative structures suggests that it is unlikely that the exact number of muscles involved should create differences in intervals between gestural activations. In order to describe the large differences in temporal latency for **dg** and **gd**, some difference in interarticulator timing, not simply execution, is probably required.

So, why should dg be more overlapped that gd? We suggest that speakers will make less of an effort to preserve less robust perceptual cues. Because the cues for the unreleased d are so weak to start with, there is little motivation for the speaker to preserve them. Even a little adulteration from C2 can overwhelm them. The formant transitions for **d** show relatively small excursions (Öhman, 1967), don't have a large effect on F3, and have an articulation, and hence, formant movement, that is relatively rapid (Kuehn and Moll, 1976). We also know that final alveolars in VC contexts are perceptually more confusable than bilabials and velars (Winitz, Scheib, and Reeds, 1972, see Table V and VI). The slight benefit which would be gained by lessening the overlap for dg is not worth the loss of transmission speed. (Recall that one of the most important aspects of communicative efficiency is parallel transmission, whereby information about several linguistic units is transmitted simultaneously in tandem. See Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967 and Mattingly, 1980.) As the listener will presumably be able to tell that there were two consonants due to the closure duration and top-down knowledge, d may well be the default guess when there are no strong acoustic cues present suggesting an alternative. The other obvious possibility is that the listener would assume a geminated C2 thereby giving rise to the phenomena of place assimilation so common in dC clusters.

Finally, why is dg so overlapped? Note that sg is more overlapped than gs and that gd tends to have shorter latencies of C2 than gs. This, along with the results presented above comparing dg vs. gd and dg vs. sg, suggests that there may be two effects at work, at least for lingual consonants. Front-back sequences are more overlapped than back-front sequences and stop-stop sequences are more overlapped than fricative-stop sequences. These influences combine to yield a large amount of overlap in dg.

3.1.3 COMPARISON OF VARIABILITY

One consequence of examining the contact profiles visually is that one is left with the distinct impression that certain dimensions of particular sequences are much more variable than others. The Levene statistic (Levene, 1960 cited in Dixon, 1988) offers a relatively interpretable means of comparing variability in two conditions. (Pooled comparisons of variance such as Cochran's test or the Hartely test (Winer, 1962) could only be made separately for each speaker which would limit them to small pools of data.) The Levene statistic uses the absolute values of the deviations from the mean as data. For this type of repeated measures data, the absolute difference from the mean was calculated for each subject separately using that subject's mean for a particular sequence. Based on visual inspection of the contact profiles, the Levene statistic was calculated for the following variables: for both front and back regions MAXIMUM and DURATION, and for each sequences' C2 ONSET RELATIVE TO C1 and ABETWEEN RELEASE OF C1 AND ONSET OF C2. The Levene F statistic is computed as an ANOVA F (Dixon, 1988). The Levene statistics were submitted to a repeated measures ANOVA using the method described previously in order to take into account speaker-specific differences in variability.

In our formulation of experimental hypotheses, we outlined an interest in the role of variability in reduction. Specifically, given the spatial and temporal reduction seen above in coda position, we consider whether codas are also more variable than onsets. The Levene F for the above variables was calculated for each pair of sequences: dg and gd, and sg and gs. In all presentations of the Levene statistics below, significant results may indicate that the ANOVA comparisons of the means are not valid due to the violation of the assumption of homogeneity of variance. However, F-tests are in fact considered quite robust to departures from homogeneity of variance (Winer, 1971). In fact, in this experiment the significant main effects observed on the variables examined in the comparison of variability are paralleled by significant main effects in only one case (FRONT MAXIMUM CONTACT). The reader is invited to inspect the contact profiles and means in confirming our conclusions.

Let's consider the stop-stop sequences first. Recall that in these sequences both consonants showed spatial (MAXIMUM contact) reduction in coda position. In the Levene analysis, for four speakers d is also more variable in MAXIMUM contact in coda position than in onset position ($F(1,3)=13.099$, $p=.0363$, with Speaker B excluded). The exceptional behavior of Speaker B might be taken as support for the compatibility of reduction and variability.

Generally, reduction occured in conjunction with increased variability. Accordingly, Speaker B was didn't exhibit a lot of coda reduction of **d** in dg and also didn't have more variability in peak displacement in coda position; in fact, her onset d's had relatively little contact and were more variable. Perhaps, Speaker B's exceptional data is due to a more careful style of speech. This speaker is both older than the others and, due to her singing background, had undergone pronunciation training in her past. These may be factors encouraging a tendency to produce more accurate or cautious speech. (On such an age related effect, see, Amerman & Parnell, 1992; Ramig, 1983; Kent and Burkard, 1981; Ringel, Chodzko-Zajko and Offenback, 1986; and Welford, 1977). The figures below show peak contact and standard deviation for each speakers' consonants in the stop-stop sequences.

FIGURE 3.39: PEAK CONTACT AND STANDARD DEVIATION IN THE STOP-STOP SEQUENCES

In addition, there was also a significant effect in the Levene test for back region DURATION such that the duration of the velar closure in gd was more variable than in dg, although there was no significant main effect on the mean durations. There was no effect of syllable position on the means or variability for the duration of **d**, or the MAXIMUM for **g**.

In the Levene analysis of **gs** and **sg**, we also find parallels between reduction and variability, in this case a lack of both. Recall that we saw above that there was no significant difference in s magnitude or duration for these two sequences. The Levene analysis also finds no significant differences in variability on these dimensions.

In analysis of both pairs of sequences, we saw that the back maximum contact means indicated coda reduction for **g**. However, the Levene analysis of variability for both pairs showed no overall greater coda varability for **g** maximum. Three speakers do, however, have more variable back peak displacements in coda position.

These results suggest that variability may play an important role in reduction, at least for the alveolar consonants. The environments which are systematically found to show reduction in both displacement and duration are environments which yield variability in these dimensions. Also, the fricative which does not reduce in displacement as a coda also does not show more displacement variability in this position. This has been observed before. MacNeilage (1970) describes a suggestion originating with Stevens and House (1963): "It may...be that targets for fricatives are specified with more precision than targets for stop consonants because the acoustic result is more dependent on precise articulator position in the former case" (p. 193).

The relationship between spatial displacement and variability is one of the issues inspiring Keating's Window model of coarticulation. To account for our displacement data within this framework, onset stops would have more specific targets than coda stops. That is, the reduced codas would not have lower target tongue positions but simply have less specific targets or wider windows. Importantly, to account for our data the window would have to a lower bottom margin; not simply be displaced downward. Note, however, that Keating's window model loses some of its appeal if the target projection from a segment's features must take place in a context-dependent fashion. As the prosodic affiliation of a segment is not specified featurally, window projection by (certain) features would have to have access to prosodic information.

The sg and gs sequences also show one other significant interaction of sequence and speaker. The ONSET OF C2 RELATIVE TO C1 is also more variable for gs than sg, with Speaker K the only exceptional speaker. No reason for this difference is immediately apparent. However, it is relevant to a claim made by Browman and Goldstein (1990b), outlined earlier here, regarding the relationship between overlap and the status of the sequence as a possible onset or coda. They suggest that their phasing rule for consonant sequences applies only to sequences whose gestures on the C-tier *(i.e.* oral gestures) form possible onsets (or codas) when reassociated to the following vowel. In instances where this is not the case, they make two predictions. First, such non-resyllabifiable sequences allow massive, sometimes complete, gestural overlap which may obscure some consonants. Second, such sequences do not have the "same kind of tight organization available," and for such sequences "variation in the degree of overlap is possible." One can infer from these remarks that no variation in phasing is possible for well-formed sequences. One qualification should be pointed out here regarding the definition of a wellformed sequence. In their discussion Browman and Goldstein allow both possible onsets and codas to count as well-formed. However, their association and phasing rules are written so as to reassociate the *leftmost* consonant in a sequence to the *following* vowel and then phase the consonant sequence C-to-C if a well-formed (i.e. "conforming to the syllable structure constraints of the language") sequence results. The directionality of this reassociation operation suggests that it is only the status of being a possible *onset* that is relevant for the phasing rule examining an "associated" consonant sequence.

Let's examine Browman and Goldstein's predictions regarding overlap and variability, as a function of "resyllabifiability". If a well-formed sequence consists of oral gestures which could form either an allowable onset or an allowable coda, then of the set gd , dg , gs , and sg , only dg is excluded. In fact, given this criterion of well-formedness, it is difficult to find any twomember heterorganic sequence, other than those having an alveolar stop (t, d, n) in first position, whose oral gestures do not comprise a possible onset or coda. (See Prator and Robinett (1985) who list 38 two-member onset sequences and 65 two-member coda sequences.) Indeed we find that dg is the most overlapped for all speakers and the most variable in relative timing for four of the five speakers. This is in accordance with Browman and Goldstein's claims. However, the variability was not significantly different from gd.

However, it could be that this criterion of possible onset or coda acts to predict increased overlap and variability exactly and only when **d** or **t** is non-final in a cluster. It is possible that in English the increased coproduction of such sequences is due to the nature of C1 rather than the well-formedness of the sequence as a whole. (See Byrd (1990) and Barry (1992) for alternative explanations.) A relevant test case might be a $[**tr**$ sequence where the **t** is initial but in an acceptable cluster; although the articulation specific to the [r] might be difficult to measure.

Alternatively, if we consider Browman and Goldstein's criteria for well-formedness to refer to only a possible *onset*, as is suggested by the reassociation rule, only sg of these four sequences has oral gestures which form a possible onset. Indeed, we do find that sg is significantly less variable (for four speakers) than gs. It is, however, significantly more overlapped. Thus the Browman and Goldstein prediction succeeds on one count but fails on the other.

In summary, the comparison of variability in the extent and duration of an articulation and in interarticulator timing for the heterosyllabic sequences gd, dg, gs, and sg demonstrates the following:

• At least for alveolar consonants, the same dimensions that are subject to reduction in coda position, are also more variable in coda position than in onset position.

This is relevant to hypothesis four $(H4)$ --that movements in coda position are more variable than movements in onset position. However, we do not find support for this hypothesis specifically as stated.

•Generally the timing measures were not significantly different in variability for these heterosyllabic sequences; the one exception being the significantly greater variability of gs as compared to sg in the relative latency of C2.

CHAPTER FOUR: EXPERIMENT TWO

4.0 **SEQUENCES DIFFERING IN SYLLABLE A FFILIATION**

This chapter considers sequences of two consonants which are syllabified as onset clusters, coda clusters, and heterosyllabic (heteromorphemic) sequences. This experiment is relevant to the ongoing research effort to determine the effects of prosody on articulator timing. Recall the hypotheses five and six from Chapter Two:

H5: Temporal coproduction of articulatory movements for sequential consonants is greater if the consonants are tautosyllabic, less if they are heterosyllabic.

H6: Coordination of articulatory movements for sequential consonants is less variable if the consonants are tautosyllabic than if they are heterosyllabic.

The latter is found to be the case by Browman and Goldstein (1988), Hardcastle and Roach (1979) and others.

As discussed in the preceding chapter, Browman and Goldstein (1990b) differentiate "well-formed" sequences from "ill-formed" sequences, where a well-formed sequence is conceived of as a sequence of oral gestures which could comprise a possible onset or coda if reassociated. Sequences not satisfying this criterion are claimed to be more variable and more overlapped than ones satisfying it. This approach claims that sequences whose oral gestures form potential onsets or codas will be less overlapped and less variable than other sequences.

Additionally, work in traditional prosodic phonology (McCarthy and Prince, 1986; Hayes, 1989; Zec, 1989) has proposed certain constituent structures for the coda, onset and heterosyllabic sequences. Consonants in a coda cluster are superordinated by a single mora which in turn is dominated by the syllable node. Sequences of onset consonants, however, are represented as being individually directly dominated by the syllable node. Sequences of consonants spanning a word boundary would be dominated jointly only by phrasal constituents. The representations are shown below:

coda cluster:

 σ

μ Ï

heterosyllabic sequence:

 σ

Given certain assumptions about the relationship between prosodic constituency and phonetic timing, the above representations would accord with particular articulatory patterns. It seems not unreasonable to expect that those consonants exhaustively forming a constituent, like the coda cluster, to have more coproduction than those not comprising a consituent, like the onset and heterosyllabic sequences. This suggests a simple extension of hypothesis five, that overlap is be greater within words than across words, to the case of the lower level moraic constituents. Given these structures and the above assumption, the hypothesis is that the moraic coda cluster is more overlapped than the other clusters. Additionally, it might be assumed that those consonants forming a moraic constituent are more cohesive, *i.e.* more stable in their timing, than the other sequences. Likewise, this assumptions and the structures above suggest an extension of hypothesis six. Just as timing is more stable within the same word, it is hypothesized to be more stable within the moraic coda cluster. This chapter considers the degree of compatibility, given these assumptions, between the articulatory data collected here and the prosodic representations shown above, and additionally addresses hypotheses five and six and the claims of Browman and Goldstein (1990b).

41 **COMPARISON OF MEANS**

Articulatory timing, lenition, and variability are considered in the sequences: s#k, #sk, $sk\ddagger$, $gt\ddagger$, $gt\ddagger$, $gt\ddagger$, $gt\ddagger$, and $ks\ddagger$. The contact profiles for these sequences are shown below. Each subgroup, i.e. fricative-stop, stop-stop, and stop-fricative, will be examined separately. First, however, all seven sequences are considered together. Effects on individual consonants are left to the reports below on sequences composed of the same oral gestures.

An ANOVA was conducted in which the seven sequences were coded as heterosyllabic, onset, or coda. Note that #sk is the only onset sequence included. The variables analyzed for this experiment are the same as those in Experiment One except that skew and flatness are not considered here since our hypotheses just concerned timing, reduction, and variability. Significant main effects are reported. As in Experiment One, if there is no significant main effect but four of the five speakers' show the same direction of effect, a main effect was calculated excluding the exceptional speaker and reported when significant. Because of the questionable validity of a post-hoc test with a repeated-measures ANOVA, the description of the effects offered in Table 4.1 is the relative ranking of means, consistent across speakers, between each pair of the treatment levels.

The following main effects of syllable affiliation on sequence timing are observed.

TABLE 4.1: A SUMMARY OF THE SIGNIFICANT EFFECTS IN THE STATISTICAL ANALYSIS OF s#k, #sk, sk#, g#d, gd#, g#s, ks#. ONSET REFERS TO THE ONSET CLUSTER, CODAS REFERS TO THE CODA CLUSTERS, AND HETERO REFERS TO THE HETEROSYLLABIC SEQUENCES.

We can see from Table 4.1 that the overlap between the consonants does differ depending on the syllable affiliation. The first consonant in an onset cluster is less overlapped by a following consonant than in a coda cluster or heterosyllabic sequence. The second consonant in a coda cluster is also less overlapped than in an onset or heterosyllabic sequence. This may be due to some word-final lengthening which extends C2 thereby decreasing the proportion of it overlapped by C1. The absolute and relative latencies between the consonants also differ depending on the syllable affiliation. Latencies are longest in the onsets cluster, in accordance with their lesser degree of overlap. The onset of C2 relative to C1 occurs the latest for onset clusters and earlier for coda and heterosyllabic sequences. No consistent difference in timing occurs between coda clusters and heterosyllabic sequences.

To examine differences in contact profiles for individual consonants due to syllable affiliation, each of the subgroups is considered separately below, *i.e.* sk, #sk, sk# and gd, gd# and **gs**, ks#. We also wish to determine whether differences exist in the timing of coda clusters and heterosyllabic sequences. Specifically, how does varying the sequence composition affect the degree of coproduction in coda clusters versus heterosyllabic sequences?

$4.1.1$ **FRICATIVE-STOP SEQUENCES**

The set of fricative-stop sequences is comprised of s-k sequences with three different prosodic affiliations: onset cluster, coda cluster, and heterosyllabic sequence. The contact profiles for these sequences are shown below, with each speaker's sequences grouped together for more convenient comparison.

FIGURE 4.2: CONTACT PROFILES FOR s#k FOR SPEAKER A

FIGURE 4.3: CONTACT PROFILES FOR #sk FOR SPEAKER A

FIGURE 4.4: CONTACT PROFILES FOR sk# FOR SPEAKER A

FIGURE 4.5: CONTACT PROFILES FOR s#k FOR SPEAKER B

FIGURE 4.6: CONTACT PROFILES FOR #sk FOR SPEAKER B

FIGURE 4.7 CONTACT PROFILES FOR SK# FOR SPEAKER B

FIGURE 4.8: CONTACT PROFILES FOR s#k FOR SPEAKER K

FIGURE 4.9: CONTACT PROFILES FOR #sk FOR SPEAKER K

FIGURE 4.10: CONTACT PROFILES FOR sk# FOR SPEAKER K; NOTE THAT THE ABSCISSA SCALE HAS BEEN CHANGED FOR THIS PLOT TO A MAXIMUM OF 35 FRAMES FROM 30 FRAMES AS IN THE OTHER FIGURES.

FIGURE 4.11: CONTACT PROFILES FOR s#k FOR SPEAKER M

FIGURE 4.12: CONTACT PROFILES FOR #sk FOR SPEAKER M; NOTE THAT THE ABSCISSA SCALE HAS BEEN CHANGED FOR THIS PLOT TO A MAXIMUM OF 35 FRAMES FROM 30 FRAMES AS IN THE OTHER FIGURES.

FIGURE 4.13: CONTACT PROFILES FOR sk# FOR SPEAKER M

FIGURE 4.14: CONTACT PROFILES FOR s#k FOR SPEAKER S

FIGURE 4.15: CONTACT PROFILES FOR #sk FOR SPEAKER S; NOTE THAT THE ABSCISSA SCALE HAS BEEN CHANGED FOR THIS PLOT TO A MAXIMUM OF 35 FRAMES FROM 30 FRAMES AS IN THE OTHER FIGURES.

FIGURE 4.16: CONTACT PROFILES FOR sk# FOR SPEAKER S

We want to consider reduction and coproduction of the s-k consonants in these sequences. If the relevant variable is whether the consonants occur in the same syllable, we expect #sk and sk# to group together in coproduction. If the relevant variable is whether the consonants are in a syllable onset or syllable coda, then we expect, for example, reduction of C1 to pattern alike in sk# and s#k; but timing possibly to differ for all three. If the relevant variable is whether the consonant is word final, then we expect, for example, s to reduce in s#k and k to reduce in sk#. Below in Table 4.2 significant main effects from the ANOVA for these three sequences are reported. The description of the effects offered in Table 4.2 is the relative ranking of the sequences when the ranking is the same for four of the five speakers.

Dependent Variable	Confidence Level	Description		
	$F(2,8)$, $F(2,6)$ when 1			
	Sp. excluded			
	Individual Consonants			
FRONT AREA	$F=5.069$, p=.0378	#sk > sk# > s#k		
FRONT NONOVERLAPPED	\overline{F} =19.967, p=.0008	$#$ sk > sk# and s#k		
AREA				
FRONT DURATION	$F=4.565$, p=.0476	#sk > sk# > s#k		
BACK DURATION	$F=6.491$, p=.0316	$#s\mathbf{k} < s\mathbf{k}$ and $s\#k$, Speaker M		
		excluded		
	Coproduction/Latency			
SEQUENCE DURATION	$F=9.023$, p=.0155	$#s\mathbf{k} > s\mathbf{k}$ and $s\#k$, Speaker K		
		excluded		
BACK DURATION (%)	$F=11.713$, p=.0042	$#$ sk \lt sk# and s#k		
SEQUENCE OVERLAP $(\%)$	$F=4.747, p=.0437$	$#s\mathbf{k} < s\mathbf{k}$ # and $s\#s\mathbf{k}$		
$C1$ OVERLAP $(\%)$	$\overline{F=11.717, p=.0042}$	$#$ sk \lt sk# and s#k		
AONSETS	$F=11.928, p=.004$	$#$ sk > sk# and s#k		
APEAKS	$F=4.872$, $p=.0413$	$#s\mathbf{k} > s\mathbf{k}$ # and $s\#k$		
Δ C1 RELEASE TO C2	$\overline{F=6.263, p=.0231}$	$\frac{1}{4}$ sk > sk# and s#k		
ONSET				
C ₂ ONSET RELATIVE TO	$F=11.563$, p=.0044	$#$ sk > sk# and s#k		
$C1$ (%)				

TABLE 4.2: A SUMMARY OF THE SIGNIFICANT EFFECTS IN THE STATISTICAL ANALYSIS OF s#k, #sk, sk#.

The differences in front region contact duration shows that [s] shortens most as a single consonant in coda position and is longest when part of an onset cluster. Being the first consonant in a coda cluster results in a duration intermediate between that of the single coda consonant and the first consonant in an onset cluster. This reduction is also reflected in the measure of front region cumulative area of contact. The back consonant was shortest as the second member of the onset cluster

We see substantial effects of sequence type on the timing measures. Onset clusters are significantly less overlapped and have longer absolute and relative latencies. Coda clusters and heterosyllabic sequences, however, appear to behave in the same way with respect to the timing variables measured here. The onset cluster is also longer in total duration than heterosyllabic sequences and coda clusters. This is in accordance with the decreased overlap in these clusters. The pooled means for the timing measures are given in Table 4.3 below.

MEASURE	#sk	s#k	sk#
FRONT DURATION	.18s	$\overline{.15s}$.16s
BACK DURATION	.12s	.13s	.14s
SEQUENCE DURATION	.22s	$\overline{.20s}$.21s
SEQUENCE OVERLAP (%)	36%	$\overline{41\%}$	$\overline{44\%}$
Cl OVERLAP $(\%)$	43%	54%	55%
AONSETS	.10s	.07s	.07s
APEAKS	.07s	.05s	.06s
Δ C ₁ RELEASE TO C ₂ ONSET	0s	$-.02s$	$-.02s$
C2 ONSET RELATIVE TO (%)	56%	46%	44%

TABLE 4.3: MEAN TIMING MEASURES FOR s#k, #sk, AND sk#.

We see that, although not consistent across speakers, there is a general tendency for the coda cluster to be more overlapped than the heterosyllabic sequence. This can be seen both for overlap and in latency of C2 relative to C1. (Recall that "latency" refers to the temporal offset between the two consonants.) A consideration of the other heterosyllabic and coda pairs may illuminate differences between them which did not reach significance here.

In sum, for these sequences it does not seem to be the case that consonants in the same syllable are more overlapped than consonants spanning a syllable boundary. Rather, sequences in coda position or initiated by a coda consonant are more overlapped than those in onset position. Reduction of the front consonant s, here C1, occurred in the form of shortening. Onset s's were the longest and coda s's the shortest. Of the coda s's, word-final coda s's were the most reduced in duration. C2, here k, was shorter as the second member of a consonant cluster but did not reduce in maximum contact in any position.

$4.1.2$ **STOP-STOP SEQUENCES**

The two stop-stop sequences, $g\#d$ and $g\#d$, are considered next. The contact profiles for the heterosyllabic sequence **g#d** have already been presented in Chapter 3--Figures 3.19, 3.21, 3.23, 2.25, and 3.27, and the reader is referred to that chapter for these profiles. The contact profiles for the coda cluster $g d#$ are presented below.

FIGURE 4.17: CONTACT PROFILES FOR gd# FOR SPEAKER A

FIGURE 4.18: CONTACT PROFILES FOR gd# FOR SPEAKER B

FIGURE 4.19: CONTACT PROFILES FOR gd# FOR SPEAKER K

FIGURE 4.20: CONTACT PROFILES FOR gd# FOR SPEAKER M

FIGURE 4.21: CONTACT PROFILES FOR gd# FOR SPEAKER S

The ANOVA results comparing the means for gd# and g#d are shown in Table 4.4 below.

Dependent Variable	Confidence Level	Description			
	$F(1,4)$; $F(1,3)$ when 1 Sp.				
	excluded				
	Individual Consonants				
FRONT AREA	$F=7.553$, p=.0708	$g\#d > gdf$, Speaker S excluded			
	n.s. trend				
NON OVERLAPPED	$F=5.613$, p=.0769	$g\#d > g d\#$			
BACK AREA	n.s. trend				
BACK DURATION	$F=7.622$, p=.0701	$g\ddagger d > g d\ddagger$, Speaker S excluded			
	n.s. trend				
	Coproduction/Latency				
FRONT DURATION $(\%)$	$F=15.045$, p=.0303	$gd# > g#d$, Speaker B excluded			
CI OVERLAP $(\%)$	$\overline{F} = 8.414$, p=.0625	$gd# > g#d$, Speaker B excluded			
	n.s. trend				
<i>AONSETS</i>	$F=11.040$, p=.0293	$g\#d > g d\#$			
ARELEASE OF C1 TO	$\overline{F=15.211}$, p=.0175	gd# > g#d			
ONSET OF C2					
C2 ONSET RELATIVE	$F=8.551, p=.0613$	$g \# d > g d \#$, Speaker B excluded			
TO C1 $(\%)$	n.s. trend				
C2 PEAK RELATIVE	$F = 246.741$, p=.0001	$g\#d > g d\#$			
TO C1 $(\%)$					

TABLE 4.4: A SUMMARY OF THE SIGNIFICANT EFFECTS IN STATISTICAL ANALYSIS OF g#d and gd#. $(n.s. trend$ DENOTES A NON-SIGNIFICANT TREND OF $\leq .08.$)

No significant differences in reduction are observed, although there are trends for both consonants to reduce somewhat in the coda cluster. In accordance with the non-significant tendency observed for s-k above, the coda cluster appears to be more overlapped than the heterosyllabic sequence. Generally, the coda cluster has a higher degree of overlap and shorter absolute and relative latencies than the heterosyllabic sequence. (The opposite effect on the interval between release of C1 and onset of C2 can be explained by noting the difference in the skew of both consonants in the two cases. The C1, *i.e.* \mathbf{g} , in a coda clusters is more positively skewed (pooled mean=.067), while as a single coda consonant in a heterosyllabic sequence it is negatively skewed (pooled mean= -0.035). (Compare the similar skew of the coda g in $g#b$, pooled mean=-.032.) This difference in skew brings the release closer to the onset of C2 in the heterosyllabic condition than in the coda cluster. This might yield the difference effect on the interval between release and onset.) The pooled mean timing measures are shown below in **Table 4.5.**

MEASURE	g#d	gd#
FRONT DURATION	.1s	.1s
BACK DURATION	.15s	.14s
$CI OVERLAP(\%)$	53%	$\overline{56\%}$
AONSETS	.07s	.06s
AC1 RELEASE TO C2	$-.03s$	$-.03s$
ONSET		
C2 ONSET RELATIVE TO	$\overline{46\%}$	42%
C1(%)		
C2 PEAK RELATIVE TO	77%	67%

TABLE 4.5: MEAN TIMING MEASURES FOR g#d AND gd#.

One other result worth noting in the stop-stop sequences is the *lack* of any significant effect on measures of spatial and temporal fortition. Specifically, the word-final d in the coda cluster is little different from the **d** in onset position, perhaps slightly shorter and smaller. This differs from the massive reduction of **d** as a single coda consonant that we observed in Experiment One. The difference here between \bf{d} in \bf{g} # \bf{d} and in \bf{g} \bf{d} # is that, in the latter, the \bf{d} is *preceded* by another consonant and is prevocalic. In order to determine whether the lack of great reduction is due to the existence of a heavy coda or due to the prevocalic context of d, we would need to compare a sequence having **d** as the *first* member of a coda cluster with a **d** in onset position. The only such sequence possible in English is the marginal cluster /nf/ as in Banff (Prator and Robinett, 1985) which would probably be realized with a final [mf] for most speakers. Recall, however, that we did observe that s was shorter as the first member of a coda cluster in sk# than it was initially in an onset cluster. (A single coda s is shorter than both of these.) This suggests that it is the fact that \bf{d} is prevocalic in \bf{g} d# that inhibits reduction, not that it is a member of a coda cluster.

$4.1.3$ **STOP-FRICATIVE SEQUENCES**

We next consider the stop-fricative sequences to determine if there are differences between the heterosyllabic g#s sequences and a ks# coda cluster having similar (oral) gestures. (As $gs#$ is not possible in English, the voicing of one of the consonants had to be changed to create an eligible coda cluster.) The contact profiles for the heterosyllabic sequence **g#s** have already been presented in Chapter 3--Figures 3.30, 3.32, 3.34, 3.36, and 3.38, and the reader is referred to that chapter for these profiles. The contact profiles for the coda cluster ks# are presented below.

FIGURE 4.22: CONTACT PROFILES FOR Ks# FOR SPEAKER A

FIGURE 4.23: CONTACT PROFILES FOR Ks# FOR SPEAKER B

FIGURE 4.24: CONTACT PROFILES FOR Ks# FOR SPEAKER K

FIGURE 4.25: CONTACT PROFILES FOR ks# FOR SPEAKER M

FIGURE 4.26: CONTACT PROFILES FOR ks# FOR SPEAKER S.

As before, an ANOVA was conducted to test for differences in reduction and coproduction between the two sequences, g#s and ks#. We find that the stop-fricative sequences behave differently from the stop-stop sequences. The ANOVA results comparing the means for g#s and ks# are shown in Table 4.2 below.

Dependent Variable	Confidence Level	Description				
	$F(1,4)$; $F(1,3)$ when 1 Sp.					
	excluded					
	Individual Consonants					
BACK MAXIMUM	$F=33.765$, p=.0044	g#s > ks#				
BACK AREA	$\overline{F=33.155}$, p=.0045	g#s > ks#				
NON OVERLAPPED	$\overline{F=14.833}$, p=.0309	$\overline{\mathbf{g}}$ #s > ks#, Speaker B excluded				
BACK AREA						
FRONT DURATION	$\overline{F=76.959}$, p=.0009	$g#s >$ ks#				
FRONT AREA	$\overline{F=34.422}$, p=.0042	$g#s >$ ks#				
Coproduction/Latency						
SEQUENCE	$\overline{F=27.506}$, p=.0063	$g#s >$ ks#				
DURATION						
C2 PEAK RELATIVE	$\overline{F=21.556}$, p=.0097	$\overline{\text{kst}} > \text{g#s}$				
TO C1 $(\%)$						

TABLE 4.6: A SUMMARY OF THE SIGNIFICANT EFFECTS IN THE STATISTICAL ANALYSIS OF g#s and ks#.

First, consider the differences in each consonant. We observe that **g** is smaller when it is the first member of a coda cluster as compared to when it is a single coda consonant. This result
is opposite to the findings for the front consonant in s#k and sk#. Recall that for these sequences we found that s was shorter as single coda consonant than as part of a coda cluster. We also find for **g#s** and ks# a shorter s in the coda cluster as compared to the onset position of the heterosyllabic sequence. In this sequence, it appears that both elements of the coda cluster are subject to reduction. The shortening of the marginal s in the coda cluster is different from the behavior observed for the marginal **d**, which did not reduce significantly in the coda cluster **gd#**.

The timing effects in the stop-fricative sequences also differ from those observed for the stop-stop sequences. The pooled means for the significantly different measures are shown in Table 4.7.

MEASURE	₽#S	ks#
FRONT DURATION	.16s	.13s
BACK MAXIMUM	50%	34%
SEQUENCE DURATION	22	19
C ₂ PEAK RELATIVE TO	99%	116\%

TABLE 4.6: MEAN TIMING MEASURES FOR g#s AND ks#.

Recall that $g dH$ showed more overlap and shorter latencies than gHd . However, gHs and ks# show minimal timing differences, and the differences that exist go in the opposite direction. In g#s, [s] achieves maximum contact earlier relative to C1 than it does in ks#. Additionally, although not a significant difference, all speakers have a higher sequence overlap for the heterosyllabic sequence, and four of five speakers have more C1 and C2 overlap in g#s and a shorter latency of C2 onset relative to C1 for $g#s$. These measures all indicate that there is more coproduction in the heterosyllabic fricative-stop cluster than in the coda cluster. However, the entire sequence duration is longer for the heterosyllabic sequence. The increase in duration of [s] relative to the whole sequence duration in the heterosyllabic sequence, taken with the overlap facts, indicates that the change in duration between the coda cluster [s] and the onset [s] is robust--.13s versus .16s. (Note that there is no interaction of [SPEAKER X FRONT DURATION] indicating that this s duration behaved similarly for all speakers.) This increased length for the onset consonant is accompanied by increased overlap, although the change in overlap is not enough to compensate for the overall greater sequence duration for the heterosyllabic sequence. The longer [s] and the lesser overlap in ks# suggest that [s] in this sequence is behaving more like an onset consonant than a coda. In fact, of the seven sequences in Experiment Three, only ks# and the onset cluster #sk have their first consonant less than 50% overlapped. See Table 4.8 below for a summary of pooled mean overlap values.

	SEQUENCE OVERLAP	C1 OVERLAP
g#s - heterosyllabic	27%	50%
ks# - coda cluster	20%	38%
g#d - heterosyllabic	46%	53%
gd# - coda cluster	$\overline{48\%}$	56%
s#k - heterosyllabic	$\overline{41\%}$	54%
$sk# - coda$ cluster	44%	$\overline{55\%}$
#sk - onset cluster	36%	43%

TABLE 4.8: MEAN OVERLAP FOR EXPERIMENT THREE.

We have seen that the stop-stop, stop-fricative, and fricative-stop sequences behave somewhat differently in terms of their coproduction. Specifically, the coda clusters appear to be more overlapped than the heterosyllabic sequences for the stop-stop sequences, but the reverses appears to be true for the stop-fricative sequences. The fricative-stop sequences appear to have no consistent difference in overlap between the coda clusters and the heterosyllabic sequences.

4.2 **DISCUSSION OF THE CODA SEQUENCES**

Steriade's (1993) Aperture Theory allows for different surface representations for the sequences studied. Specifically, while the stop-stop and fricative-stop sequences require three aperture positions (namely closure-closure-release and frication-closure-release, respectively), in the stop-fricative sequence, the frication can occur during the stop release yielding an affricatelike structure (namely closure-frication).¹ We would like to consider whether this difference in aperture structure is reflected in articulatory timing. Specifically, is the ks# cluster more overlapped and/or does it have a shorter total duration than the coda clusters $g d\#$ and $sk\#$?

However, theories in which marginal coronals are considered extrasyllabic or as syllable affixes² (Fujimura and Lovins, 1978) suggest that the phonetic implementation of such a sequence will have less overlap than coda clusters without such an element due to a looser affiliation of the appendix with its syllable. In this case, ks# and gd# would be likely to be less overlapped and longer than sk# due to the marginal coronal consonants.

We considered timing differences between these three coda clusters. The sequences decrease in total duration from sk# to ks# to gd# $(F(2,8)=6.558, p=.0206)$. The ks# sequence is less overlapped than the other two sequences ($F(2,8)=9.007$, $p=.0089$). The sequence gd# has a shorter latency (as measured by the temporal time between maximum displacement for each consonant) than both of the other sequences $(F(2,8)=17.364, p=.0012)$. The relative latency, (as determined by the percent of the way through C1 contact that the maximum displacement of C2

¹I thank Abigail Kaun for suggesting this point to me.

²Donca Steriade points out to me however that the distributional argument which suggests that k s# contains an appendix may well apply to sk# as well. Neither occurs as word-medial coda in English; in fact, ks# occurs marginally, eg. [*cks.tra*], but sk# does not. (Steriade, p.c.)

occurred), decreases, *i.e.* shows C2 peaking earlier in C1, from ks# to sk# to gd# for all speakers $(F(2,8)=29.778, p=.0002).$

These results suggest that coproduction *decreases* from **gd#** to sk# to ks#. This is compatible with a prosodic representation postulating marginal s's which are not as tightly bound to the rest of the syllable as other consonants. There is no evidence for the assumption that marginal d's behave similarly, as would be expected given a theory in which all coronal obstruents are considered to be appendices. The gd# sequence is shorter and more overlapped than both ks# and sk#. These results do not offer evidence that the realization of ks# in Aperture Theory as a two-position sequence differs from that of $gd#$ and $sk#$ as three-position sequences. This suggests that either the number of aperture positions has no direct reflection in the articulatory timing of the coda cluster or that the number of aperture positions does not differ in the postulated way.

Interestingly, these results also differ from the preference observed in Chapter 3 for frontback sequences to be more overlapped than back-front sequences. Prosodic differences in these three coda clusters may be more influential on timing than the place-order effect observed earlier. The results do however accord with the finding in Chapter Three that stop-stop sequences are more overlapped than sequences including a fricative.

Finally recall our first analysis in this chapter which demonstrated that the #sk onset cluster was less overlapped that the coda and heterosyllabic clusters. If this result were replicated for other sC clusters, we might have the beginnings of an explanation for an interesting morpheme structure constraint in English. In English, homorganic noncoronals may not flank both sides of the vowel in sCVC sequences (Davis, 1990). Tongue tip consonants of any type may appear in this position however. If sC onset clusters regularly have little overlap. this could yield a situation in which the C's in a $sCVC$ sequence are "pushed" close together, requiring too rapid an opening and re-closing of the articulator. In the case of $sCVC$ sequences where C is made with the same articulator as s , *i.e.* the tongue tip, we find a situation in which gestural *blending* occurs rather than overlap between different articulator subsystems. This might allow greater coarticulation in such sequences than in sk or sp sequences for example. Additionally, if the tongue tip is significantly faster than the other articulators, this would also facilitate the articulation of $sCVC$ sequences in which C is coronal. These factors might be the source of the exceptional status of coronals in this morpheme structure constraint. The constraint itself might exist in response to the relatively small amount of overlap in sC onsets.

4.3 **COMPARISON OF VARIABILITY**

At the beginning of this chapter, we outlined certain predictions regarding the cohesiveness of different types of sequences. Specifically, standard prosodic representation suggest increased overlap and decreased variability for consonants sharing a mora, *i.e.* coda clusters. Articulatory Phonology (Browman and Goldstein, 1990b), claims that "well-formed" sequences, *i.e.* those comprised of oral gestures forming possible onsets or, perhaps codas, will be least overlapped and least variable. Now that we have looked at the timing differences in these sequences, let's consider their relative variability in timiliare Levene F statistic was calculated as described in Chapter 2 in order to evaluate differences in timing variability between onset clusters, coda clusters, and heterosyllabic sequences. Two variables are considered: when C2 onset occurred relative to C1 (relative latency) and the interval between the release of C1 and the onset of C2 (absolute latency). All seven sequences, s#k, #sk, sk#, g#d, gd#, g#s, ks#, were included and coded as onset cluster, coda cluster, or heterosyllabic sequence.

The results show a significant main effect on the relative latency of C2. The onset cluster is less variable in overlap than either coda clusters or heterosyllabic sequences $(F(2,8)=4.923,$ $p = 0.0404$). The difference in variability between coda clusters and heterosyllabic sequences is negligible. We saw above that onset clusters were the least overlapped of the three sequence types.

Next, we examined each set of sequences separately to test for differences in timing variability between the coda and heterosyllabic sequences. g#d and gd# and g#s and ks# have no significant differences in timing variability. However, the analysis of s#k, #sk, and sk# determines that all speakers except Speaker B have more variable relative latencies of C2 in the heterosyllabic sequence s#k than in the coda cluster sk# $(F(2,6)=13.762, p=.0057,$ Speaker B excluded). Recall that we observed no consistent difference in the amount of overlap between consonants in these two sequences.

In summary, for these sequences we find only marginal evidence (4 speakers, 1 sequence type) for decreased variability in coda clusters due to greater gestural cohesion. In most cases, timing variability observed in coda clusters is like that in heterosyllabic sequences not having any word-level superordinate constituent. Onset clusters on the other hand exhibit less variability and less overlap than coda clusters or heterosyllabic sequences. We state this as a tentative conclusion because we only have #sk as a representative onset sequence. The effect here was robust however and visible by inspection of the contact profiles. This result is in accordance with Browman and Goldstein's evaluation of the importance of the well-formedness status of a *potential* onset cluster, as the other sequences of oral gestures in this experiment do not form possible onsets. The relatively greater variability and greater overlap in the non-onset sequence supports the Articulatory Phonology suggestion that increased overlap is accompanied by increased variability. The relatively stability of the onset cluster is not predicted by the standard constituent structure representation in which consonants in coda clusters form a single moraic constituent unlike those in onset clusters. For one of the three sequence types, s-k, we also see evidence for increased timing variability in a heterosyllabic sequence as compared to a coda cluster. Within the Articulatory Phonology framework, this might suggest that the status of potential *coda*, as well as potential onset, may also be a relevant factor. Two other coda clusters,

however, showed no lesser variability in their temporal organization than their analogous heterosyllabic sequences.

44 **SUMMARY OF SYLLABLE AFFILIATION EFFECTS**

In summary, we have found evidence that onset clusters, coda clusters, and heterosyllabic sequences differ in their interarticulator timing and in reduction. However, the precise nature of these effects depends on the consonants in the sequence. The onset cluster is less overlapped and less variable in its timing than the coda clusters and heterosyllabic sequences. For the fricative-stop sequences, minimal differences in the timing of coda clusters and heterosyllabic sequences are found. In the stop-stop sequences, coproduction is greater in the coda clusters than in the heterosyllabic sequences. In the stop-fricative sequences, however, there is some evidence that the coda cluster is less overlapped than the heterosyllabic sequence.

With respect to spatial and temporal reduction, there is little evidence for a magnitude difference of consonants in coda clusters as compared to single codas in heterosyllabic sequences. The one exception to this is the reduction of the velar consonant in the coda cluster ks#, as compared to $g#s$. There is also little evidence for the reduction of the final member of a coda cluster, *i.e.* C2, as compared to when it occurs as a single onset in the heterosyllabic sequences. That is, the large reduction of single codas compared to single onsets that we observed in Chapter Three, does not appear robust for the word final consonants in coda clusters.

Because of our intention to limit this experiment to two member sequences, we have failed to directly control experimentally for the variable of resyllabification. This restriction on the design yields stimuli in which the coda clusters are prevocalic. However, the clusters do not precede an unstressed syllable but rather a syllable receiving about the same level of stress as the syllable in which the cluster occurs. This should lessen the likelihood of resyllabification.³ Even so, the effect of possible resyllabification of the cluster must be considered, especially in light of the small degree of difference observed between the coda clusters and the heterosyllabic sequences.

Let's evaluate this possibility in terms of Browman and Goldstein's (1990b) statements for reassociation and timing in consonant sequences. Recall that their statement of association or ambisyllabicity associates the first consonantal gesture in a sequence to both neighboring vowels. They say:

 3 Note especially that the coda sequence with the least overlap, ks#, was also the sequence that was followed by the largest phonological phrase boundary: "Say backs Abigail."

The leftmost consonantal gesture of a consonant sequence intervening between two vocalic gestures is associated with both vocalic gestures. A consonant sequence is defined as intervening iff the entire sequence lies between the two vocalic gestures when projected onto the one-dimensional oral tier. (Browman and Goldstein, 1990b:257)

They use a heterosyllabic sequence in an example of this statement's application, associating the coda C to the vowel of the following word. After reassociation, the rule for consonant cluster phasing then operates if the sequence is a well-formed one; i.e. possible onset or (perhaps coda). (Recall that, as it is the *first* consonant which is associated to both vowels, the status of possible onset seems to be the crucial one, but no explicit differentiation between possible onset and possible coda is made by Browman and Goldstein.) If it is not well-formed, the rule does not operate. Such a sequence is predicted to be variable and more overlapped than a well-formed one.

With these experimental stimuli, the association process would operate on the sequences to yield the associations shown below. The associations between the C and V tiers added by the application of this statement is shown by a heavy line; underlying associations are shown by a lighter line. (Vacuous application of the association statement is not shown.) Note that, within a tier, cannonical sequencing relations are preserved visually, and between tiers, the sequencing is conveyed by the angle of the association lines.

onset cluster association:

coda cluster association: (prelim.)

heterosyllabic sequence association:

Browman and Goldstein suggest that a final s gesture does not co-occur, *i.e.* is not associated, with the vowel preceeding it. Following this suggestion, the reassociation of the ks# cluster yields:

Browman and Goldstein do not elaborate on the association of other consonants in this position. The consonant cluster phasing rule, aligning the onset of $C2$ to the release of $C1$, should apply only to consonants that are associated to the same vowel. Coda clusters with a marginal s should not undergo the rule.

As we can see, identical association relationships are yielded for onset and heterosyllabic clusters. Therefore, only the application, or lack thereof, of the phasing rule should engender timing differences. Whether the rule applies or not also depends on the well-formedness of the sequence. If status as possible onset or coda is the significant determiner of its application, then the onset and heterosyllabic sequences examined here are predicted to be timed similarly (and consistently). This prediction is not supported in the experiment above. If status as possible onset alone is the significant determiner of the application of this phasing rule, then the timing of the #sk onset should be different from all the other sequences. To a certain extent this was the case. It was less overlapped and more stable, as predicted. However, consider the behavior of the other sequences. If status as a well-formed potential onset is crucial to application of the CC phasing rule, **g#s** and k#s should behave similarly. Neither should undergo the phasing rule, the former because it is not a potential onset and the latter because it is not properly associated. This was not the case; timing differences were observed here. Also, these sequences were the least overlapped, not in accordance with Browman and Goldstein's prediction of large overlap when the CC Phasing Rule fails to apply. It is not clear if differences are predicted to exist between **g#d** and **gd#** and between s#k and sk# since we aren't certain about the predicted association of the coda cluster consonants. The association and phasing statements do not illuminate the timing differences *between* the various coda clusters or the differences in timing and variability that were found between the coda clusters and the heterosyllabic sequences. They do account nicely for the behavior of the onset sequence in comparison to the other sequence types.

Finally, let's return to the phonological representations for clusters shown at the beginning of this chapter. In addition to the preceding discussion, two points should be noted. First consider Selkirk's (1982) proposal that s+obstruent onset and coda clusters be considered

single segments (see also Steriade, 1982 and Lamontagne, 1993).⁴ Browman and Goldstein suggest that the presence of two oral constrictions and one glottal constriction for s-stop sequences causes the possibility of them acting as single units or as two units. Our observations of relatively less overlap in the onset #sk seem to run counter to other cases where other pairs of (non-laryngeal) constrictions form a single segment. Additionally within the examination of the coda clusters, the sk# coda cluster was less overlapped than the **gd#** cluster. These findings suggest that the supposition of s-obstruent clusters being a single segment in contrast to other clusters does not relate very transparently to intuitive notions regarding the greater extent of overlap within segments than across them. However, the stability of the #sk timing should be noted, as it is suggested in Chapter Seven that timing stability plays a role in segmenthood.⁵

In summary, we have seen no consistent evidence of overlap or variability differences between tautosyllabic and heterosyllabic sequences. Thus, hypotheses five and six are not supported. Additionally, the data examined did not appear compatible with the prosodic representations shown in Figure 4.1 given the assumptions outlined about the relationship between prosodic constituency and phonetic timing. That is, no consistent evidence of greater overlap or lesser variability was found for the coda cluster as compared to the heterosyllabic sequence. There seems to be no direct evidence that the coda cluster consonants are in closer association with one another than the heterosyllabic consonants.

Finally, let's reconsider the representation of the coda cluster presented for discussion in Section 4.0 in which the coda consonants share a single mora. There is no principle of moraic theory which requires this to be the case. We should also examine the possible representation in which only the first of the coda consonants is moraic.

⁴I thank Donca Steriade for signaling the interest of this point to me.

⁵One of the reasons why s-stop clusters are sometimes considered to form a single segment is the disinclination of such sequences to allow epenthesis. One reason for this may be that an unusually abducted glottis is found for /s/'s in comparison to voiceless stops (Hirose and Gay, 1972; Hirose, Lisker and Abramson, 1972; Collier, Lisker, Hirose, and Ushijima, 1979; and Yoshioka, Löfqvist, and Hirose, 1982) due to the necessity for a high rate of airflow. Additionally, in a number of languages only one glottal gesture is found for words beginning with s-stop clusters (Browman and Goldstein, 1986 and Goldstein 1990 citing Yoshioka, Löfqvist, and Hirose, 1981; Löfqvist and Yoshioka 1980 a, 1981; Petursson, 1977; and Fukui and Hirose, 1983). Because the following stop in s+stop sequences generally agrees in voicing, the insertion of an epenthetic vowel would require a rapid change from a very abducted glottis to a brief moment of voicing for the epenthetic vowel, to, again, an abducted glottic for the voiceless stop. (I thank Richard Wright for pointing out this possibility to me.) Other clusters of voiceless obstruents may have multiple openings of the glottis (Kingston, 1990 citing Löfqvist and Yoshioka 1981 a, b and Yoshioka, Löfqvist, and Hirose, 1981, 1982). These openings may also be less extreme voiceless fricatives are not involved. A case in which one could differentiate between the effect of degree of glottal opening and of single versus multiple openings would be if different epenthesis behavior was observed for sk versus [k onsets.

This representation is possible in principle, is more compatible with the findings above, and is suggested by independent evidence.⁶ Such a representation presents a greater the parallelism between the coda cluster structure and that of the heterosyllabic sequence. This is in line with our data which showing no consistent timing differences between these two types of sequences. It also seems in keeping with the lack of magnitude reduction observed for C2 in the coda cluster as compared to that consonant in onset position; suggesting parallelism between the moraic status of C2 in a coda cluster and in an onset. Furthermore, consonant duration facts in Swedish (Löfstedt, 1992) lend independent support for such a structure. Löfstedt describes the first consonant of a (monomorphemic) coda cluster, and only this consonant, as long under stress. (See also Prince, 1980 on a similar phenomena in Estonian). This suggests that C1 is different from C2 in a coda cluster with respect to its rhythmic, *i.e.* moraic, properties in the language. The above facts regarding overlap, spatial reduction, and rhythmic behavior are all compatible with the alternative representation of coda clusters shown above in which the first consonant of the complex coda is moraic and the following consonant(s) is directly affilliated with the syllable.

 $6I$ thank Donca Steriade for suggesting the issue of coda representation to me.

CHAPTER FIVE: EXPERIMENT THREE--GLOBAL TIMING, C-CENTERS, AND COMPENSATORY **SHORTENING**

5.0 **REALIZATION OF LONGER SEQUENCES**

This chapter investigates two aspects of the production of longer consonant sequences. First, it addresses the phenomenon known as compensatory shortening. Compensatory shortening is the term used to describe the acoustic shortening of adjacent vowels as the number and length of intervening consonants increases. Compensatory shortening is hypothesized to arise from the overlap of consonants onto adjacent vowels. The vowels may then appear shorter even if their gestures are unchanged. We can test this directly because we have both acoustic and articulatory data.

Secondly, this experiment also investigates the evidence for a global timing organization for consonants in sequence. We use global timing to refer to how a sequence of consonants is coordinated as a unit to the vowels on either side of it. Alternatively, local timing is used here to refer to the coordination of a single consonant in a sequence to the neighboring vowells. This experiment is based on the 1988 work of Browman and Goldstein. They concluded experimentally that sequences of consonants are timed locally to the preceding vowel by phasing the vowel and the achievement of target of the first consonant, and that onset sequences are timed globally to the following vowel by phasing the vowel and the sequence C-Center, the mean of all the midpoints of peak displacement for each gesture in a sequence. The experiment presented below focuses on sequences of coda consonants and expands on Browman and Goldstein's effort, which considered only single coda consonants. The limited structure of codas in their experiment may have influenced their findings especially with respect to the VC timing. The present experiment does not attempt to replicate their findings with respect to onset sequences, although single onset consonants are examined. Finally, Browman and Goldstein's experiment found no effect of canonical syllable affiliation in the choice between local and global timing organizations. The only effects of consonant affiliation they found were that between-word measures were more variable than within-word measures (see also Hardcastle and Roach, 1979), and an asymmetry due to vowel quality which will not be relevant here.

Recall from Chapter Two that this experiment uses sequences of one to four consonants forming a coda cluster or a coda cluster plus a single onset consonant. The sequences considered are: s#, k#, d#, sk#, kt#, ks#, sks#, skt#, ks#k, kt#k, sks#k, and skt#k. The sequences are flanked by [bæ] and [æb] on either side. Seven tokens of each sequence from each of the five speakers are analyzed. (The Browman and Goldstein experiment on which this effort is modeled used four sets of six utterances each, where timing for each set was analyzed separately and one token of each utterance was collected for a single speaker. The sequences they examined included: six of the form #C, four #CC, two #CCC, six C#C, two C#CC, and four C#.) In our experiment, each contact maximum in a region was counted as the realization of a consonant in that region. (See Section 2.6.2 for the details of measuring the linguapalatal trajectories.) A single maximum counted as two consonants in that region when the temporal period from initial to final contact in the region properly contained or was coextensive with contact for a consonant in the other region.

The produced tokens were inspected using these metrics and the "missing" consonants (i.e. those for which no lingua-palatal contact occurred in the defined regions) were determined. The sequences were realized as follows. For all speakers, all seven tokens of d#, k#, s#, ks#, kt#, sk#, and kt#k had all consonants realized on the pseudo-palate. This includes five tokens of kt#k for Speaker A, two for Speaker K, seven for Speaker M, and three for Speaker S in which the k contact was coextensive with or properly contained the d contact. For ks#k, Speakers A, B, and K had separate maxima for every consonant. Speaker M had separate maxima for three tokens and four tokens in which the k contact was coextensive with or properly contained the s contact. Speaker S had separate maxima for six tokens and one token missing C1. In the three member coda sequences and four member sequences consonants were often missing. Of the 35 tokens evaluated for each sequence, skt# had three with separate maxima for each consonant and 21 in which contact in the front region was coextensive with or properly contained that in the back. For sks# there were 14 tokens with separate maxima for each consonant and 19 in which contact in the front region was coextensive with or properly contained that in the back. For skt#k 30 of the 35 tokens lacked contact for C3 and C4. The other tokens had all consonants realized on the palate either with individual maxima or by the "properly contained" metric. Finally, for sks#k, 18 tokens had all consonants realized, three lacked contact for C4, 12 lacked contact for C3 and C4, one for C1 and one for C2.

As stated in Chapter Two, the duration variables reported below consider different numbers of tokens. This was necessary to ensure that the measurements are not skewed due to reduction (yielding no linguapalatal contact) of consonants in the sequence. In the timesynchronous acoustic waveform, measurements were made paralleling Browman and Goldstein's (1988) choice of preceeding-vowel and following-vowel anchorpoints. In both their experiment and this experiment, the point of acoustic closure for the consonant following V2 was chosen as the following-vowel anchorpoint. We call this the V2 ANCHOR. For the preceeding-vowel anchor, Browman and Goldstein chose the midpoint of the labial movement for the bilabial consonant preceding V1. Here, not having lip movement data, we have chosen the acoustic midpoint of the bilabial closure as the following-vowel anchorpoint. We call this the V1 ANCHOR. Other variables for Experiment Three are the times from the anchor points to the first and last edges of the sequence. Browman and Goldstein (1988) defined their FIRST EDGE as the edge of the plateau (within approximately 1.3 mm) of peak displacement of an X-ray microbeam pellet attached to an articulator. Analogously, we define FIRST EDGE as being the first frame of maximal contact for a consonant. LAST EDGE is defined similarly here as being the last frame of maximal contact for a consonant. Browman and Goldstein (1988) derive the C-center point for a sequence by computing the temporal midpoints between the left and right edges of the peak displacement plateaus and then calculating the mean of all the plateau midpoints of the gestures in the sequence. We do likewise taking the mean of the centers of each plateau of maximum displacement of each consonant in the sequence.

For those variables evaluating the FIRST EDGE of the consonant sequence all tokens are considered, excepting two for which C1 was missing. These two tokens are excluded from all results below. For those variables evaluating the C-CENTER or LAST EDGE of the sequence, sequences in which all consonants were articulated according to the above metrics are considered. For those variables in which the center or edge of an onset consonant is considered separately, only those sequences with a *separate* maximum for each consonant are considered. This is also true for variables measuring a point in the coda cluster of a coda+onset sequence. These criteria for inclusion result in the following number of tokens shown for subject by sequence type by variable below.

TABLE 5.1: NUMBER OF TOKENS USED IN EXPERIMENT THREE FOR CALCULATING THE MEANS AND STANDARD DEVIATIONS OF EACH VARIABLE BY SEQUENCE TYPE BY SPEAKER.

5.1 NEIGHBORING VOWEL LENGTH--COMPENSATORY SHORTENING?

Let's leave aside for a moment the evaluation of the global timing anchor point predictions made by Browman and Goldstein (1988), and look first at the length of the vowels neighboring the consonant sequence. We do not use a strictly acoustic measure. For the preceding vowel, we consider the period from the acoustic release of the bilabial consonant, V1 ONSET, to the first frame of maximum contact for C1, FIRST EDGE. For the following vowel, we consider the period from the last frame of maximum contact for the sequence, LAST EDGE, to the acoustic closure for the bilabial consonant, V2 END. It is assumed that measures of sequence edges are closely correlated with the acoustic cessation and initiation of the vowel.

$5.1.1$ **RESULTS**

The graph of the pooled means for V1 followed by different sequence lengths and types is shown below. All tokens (except the two missing C1 contact) are included.

(total # of $C's$ / # of $C's$ in coda) FIGURE 5.1: V1 DURATION MEASURED BY THE TIME IN SECONDS FROM V1 ONSET TO THE SEQUENCE FIRST EDGE, SPLIT BY UNDERLYING NUMBER OF CONSONANTS AND SEQUENCE TYPE.

In this figure, compensatory shortening would be seen as decreasing height of the bars from left to right, as a function of either the first number (total number of consonants) or the second number (number of coda consonants) under each bar. Such a decrease is not seen. We do see a dependence of vowel duration on the number of consonants in the coda (the second number under each bar), but the relation is positive instead of negative. Furthermore, there is little difference between one and two coda consonants; the effect is almost entirely that three-member codas greatly increase the duration of the preceding vowel. We see that the vowel preceding a three member coda sequence has approximately the same duration as a four member codacluster+onset-sequence, and a two member coda has similar preceding vowel duration to a three consonant coda+onset-sequence. However, yowel duration is *longer* before three consonant codas than before one and two member coda clusters. Indeed, a repeated measures ANOVA determines there to be a significant effect of number of coda consonants on this measure of vowel length $(F(2,8)=28.008, p=.0002)$ mostly due to differences between sequences with three coda consonants and those with one or two.

The graphs of the pooled means for V2 duration in contexts of different preceding sequence types are shown below. The top graph shows separate V2 duration means for each sequence type and length, and the bottom graph compares the same data separating different sequence lengths and collapsing across sequence type.

FIGURE 5.2: V2 DURATION MEASURED BY THE TIME FROM SEQUENCE LAST EDGE TO V2 END IN SECONDS SPLIT BY UNDERLYING NUMBER OF CONSONANTS AND SEQUENCE TYPE. ONLY THOSE SEQUENCES IN WHICH NO CONSONANTS LACKED CONTACT ARE INCLUDED.

Like preceding vowel length, following vowel length increases with increasing numbers of consonants. However, following vowel length, in contrast to preceding vowel length, appears to depend on the number of underlying consonants in the sequence, not on the number of coda consonants. A repeated measures ANOVA shows a significant effect of underlying sequence length on following vowel duration $F(3,12)=9.967$, p=.0014. The difference between the one and two member sequences is small but in the same direction for all speakers.

5.1.2 DISCUSSION

These suprising results for neighboring vowel durations conflict with the predictions made by Browman and Goldstein (1988) and conclusions drawn by Munhall, Fowler, Hawkins and Saltzman (1992).¹ Let's consider V1 duration first. Browman and Goldstein suggested that consonant sequences, regardless of their underlying syllable affiliation, are aligned to the preceding vowel by the attainment of target displacement for the first consonant. They state "[t]he first consonant in the sequence will be linked to, and partially overlap, the vowel, but the following consonants will not link to, or overlap, the vowel"; (p. 152) and "there is no automatic increase in the amount of overlap [of consonants and preceding vowel] as the number of consonants in the final cluster increases" (p. 150). They noted that this prediction is not in accordance with Munhall, Fowler, Hawkins and Saltzman (1992) who found earlier achievement of consonantal target with two coda consonants as compared to one. Munhall et al. (1992) found that vowel duration shortened as the duration of a coda cluster increased. In our data the duration of the coda cluster, as indexed by the time between sequence FIRST EDGE and sequence LAST EDGE, also increases as sequence length increases. It is also longer for three consonants with a word boundary than for three consonants without. However, preceding vowel duration doesn't decrease as predicted by Munhall *et al.* They found that the jaw lowering gestures were shorter in duration before a coda cluster than before a singleton coda consonant. They state that this vowel shortening seems to be achieved by an earlier onset of the raising gesture for the following consonant in a cluster than as a singleton. The blending of these requirements creates a shorter lowering movement. Munhall *et al.* state that an explanation for why raising gestures for clusters begin earlier than for singletons must await further research. The results shown above for the V1 interval are not in accordance with either of these findings. While no difference is found between one- and two-member coda clusters or between these and three-member sequences CC#C, when the number of *coda* consonants increases to three, the vowel interval lengthens. No difference is observed between CCC# and CCC#C, however.

It may be that these long coda sequences are more difficult to articulate and cause the speaker to slow down. Note however, that this proposed complexity is determined by the coda

¹Note that Coleman (1992) presents a temporal interpretation of syllable constituents in which coda sequences and onset sequences both temporally overlap their syllabic nucleus; but his approach makes no predictions about the degree of overlap, differences in overlap between onset and coda clusters, or how the number of consonants may influence overlap.

cluster, not by the length of the sequence as a whole. The clusters with only one or two elements in coda are not difficult enough to cause a slowing in speaking rate. Some small amount of compensatory shortening may even occur here. An important difference between Munhall et al.'s experiment and this experiment is that they tested only one and two member coda sequences. In our means, there is a tendency for a vowel before CC# to be shorter than before $C#$. However, when longer sequences are examined, the vowel duration does not continue to shorten. Thus the inferences made about compensatory vowel shortening from experiments using one and two member sequences may not necessarily be extended to longer squences. DeJong (1991) found some shortening of vowel duration for two member sequences as compared to singleton codas. However, his regression plots (Fig. 4, p. 9) show no shortening for three member [sts] codas as compared to two member [st] codas. It seems therefore that it is important to consider longer sequences of consonants before extending general timing conclusions based on one and two member sequences.

In the case of V2 duration, Browman and Goldstein (1988) did not explicitly test the CVtiming (rightward) of coda clusters, *i.e.* the between word timing. They did find that the CVtiming of onset clusters is organized to create a stable C-center. This means that, unlike VCtiming (leftward), as more consonants are added in an onset sequence the following vowel will be more overlapped. They also hypothesized that the vocalic gesture starts at the achievement of target of the first consonant in the sequence. They found that their observed VC timing relationship existed regardless of the canonical syllable affiliation of the consonants. If this result were paralleled in CV-timing, then we might expect Browman and Goldstein to extrapolate that CV-timing, like VC-timing, is anchored to the C-center regardless of which word the underlying sequence is affiliated with. If this were the case, coda sequences would be predicted to overlap a following vowel just like onset sequences.

In fact we find effects in the opposite direction. In $V_1C_nV_2$ sequences, rather than having V2 shorten as more preceding consonants are added, we see that it lengthens as the preceding consonant sequence lengthens. However, unlike V1, it does not seem to matter what the canonical syllable affiliation of the consonants is; CCC# gives a similar value for V2 as CC#C, a value which lies between that for sequences with two and four consonants. Yet the period between the two vowels, *i.e.* here, from the first contact maximum to the last contact maximum for the consonant sequence, does increase as sequence length increases from one to four consonants. Also, CC#C is substantially longer than CCC#, but their V2 durations remain the same. It seems that no invariant phasing relationship between the consonants and V2 will yield a

²Note that one-third of the single consonant tokens are final d's, while the two-consonant codas all begin with a voiceless consonant. We wanted to make sure that this imbalance wasn't obscuring a compensatory shortening effect so we also calculated the V1 duration means excluding the d# tokens. This changed the means only slightly yielding an 8ms longer V1 for single consonants as compared to a 3ms longer V1 with the d# tokens included. When the end of the V1 interval is defined more conservatively as the θ regional contact for C1 and d# tokens are excluded, the differenence again becomes slightly greater with C# 's V1 being 14ms longer than CC#'s. However, CC#'s V1 is still 23ms shorter than the CCC# tokens.

longer V2 duration as the number of consonants in the prevocalic sequence is increased. If the vowel were phased to the LAST EDGE of the sequence, we would expect no effect of the number of consonants on vowel length. If it were phased to the C-Center or FIRST EDGE, we would expect the vowel to shorten as sequence length increased. The increase in vowel length must be due to a rate, e.g. stiffness, change in the vowel itself such that the rate decreases as sequence complexity increases. This result is comparable to the increased duration observed for V1; however, the interval for V1 depended on the tautosyllabic consonants only. The V2 interval appears to be a function of the total number of preceding consonants regardless of their canonical syllable affiliation. Thus we find the following influences noted by arcs:

5.2 **MOST STABLE ANCHOR POINTS--C-CENTERS?**

The second focus of this experiment is modeled on Browman and Goldstein's 1988 C-Center experiment. Our experiment makes two significant alterations. First, it uses data outside the midsagittal plane and it emphasizes coda clusters rather than onset clusters. Recall their findings that onset consonants and sequences of onset consonants are timed to the following vowel by phasing their C-Center to the vowel. (See Section 5.0 for Browman and Goldstein's and our arithmetic definitions of the C-CENTER, sequence edges and vowel anchor points.) We call the relationship between the consonants and the following vowel the CV-timing relationship and that with the preceding vowel the VC-timing relationship. The VC-timing relationship of an onset sequence was found by Browman and Goldstein to be achievement of target of the first onset consonant with this vowel. Browman and Goldstein additionally examined the behavior of a single coda consonant followed by an onset sequence. Only the VC-timing relationship was examined here; not the coordination across the syllable (word) boundary. These $C +$ onset cluster sequences were found to be organized to the preceding vowel in the same way as an onset sequence without a coda consonant. Alignment of the achievement of the FIRST EDGE of the sequence, whether onset or coda, with the preceding vowel was found to be more stable than either the C-CENTER of the coda consonant or the whole sequence. Browman and Goldstein do not report on the CV timing relationship for sequences having coda consonants. It might be the case that the CV-timing relationship, like the VC-timing relationship, is unaffected by the canonical syllable affiliations of the consonants. If this is so the coda clusters should be aligned to V2 by their C-CENTERS.

Like Browman and Goldstein, we use below the standard deviations of the variables shown in Table 5.1 to compare the relative stability of the several anchorpoints to the adjacent consonants.

5.2.1 CV TIMING

The following table shows the standard deviations for the CV-timing relationships for each speaker. Recall that the anchorpoint (V2 ANCHOR) chosen for this relationship to be analogous to Browman and Goldstein's experiment was the point of acoustic closure for the bilabial consonant after V2. We present the values separately for each sequence length, sequence type, as well as pooled across sequence type, so that the reader may evaluate differences in variability for each subgroup. The most stable variable for each set for each speaker and for the mean for all speakers is shaded.

standard deviations in (s)	$Speaker \rightarrow$	mean s.d. across Speakers	A	B	K	M	S
C-CENTER to V2 ANCHOR	C#	.0223	.0085	.0164	.0576	.0152	.0141
	CC#	.0164	.0164	.0175	.0176	.0151	.0151
	CCC#	.0415	.0467	.0394	.0527	.0405	.0283
	CC#C	$\sqrt{0215}$.0200	.0208	.0205	.0259	.0200
	CCC#C	.0417	.0615	.0403	.0361	.0246	.0462
	coda only sequences	.0373	.0342	.0347	.0641	.0283	.0254
	$\overline{\text{coda} + \text{onset}}$ sequences	.0302	.0256	.0403	.0300	.0257	.0293
	all sequences	.0567	.0471	.0803	.0739	.0390	.0431
SEQ FIRST EDGE to V2 ANCHOR	C#	.0253	.0105	.0189	.0582	.0204	.0184
	$\overline{CC#}$.0251	.0253	.0261	.0324	.0174	.0244
	CCC#	.0638	.0744	.0642	.0677	.0493	.0632
	CC#C	.0475	.0622	.0302	.0295	.0502	.0652
	CCC#C	.0600	.0632	.0965	.0632	.0240	.0533
	coda only sequences	.0521	.0527	.0487	.0779	.0331	.0482
	coda + onset sequences	.0575	.0627	.0755	.0517	.0394	.0582
	all sequences	.0798	.0705	.1159	.0958	.0474	.0695
SEQ LASTEDGE to V2 ANCHOR	C#	.0240	.0109	.0172	.0586	.0202	.0130
	$\overline{CC#}$	$\overline{.0160}$.0204	.0168	.0188	.0126	.0115
	\overline{CCC}	.0321	.0382	.0255	.0455	.0339	.0173
	CC#C	.0205	.0197	.0121	.0386	.0151	.0170
CCC#C		.0371	.0069	.0517	.0208	.0460	.0604
	coda only sequences	.0281	.0242	.0236	.0534	.0248	.0144
	coda + onset sequences	.0352	.0193	.0590	.0371	.0255	.0348
	all sequences	.0379	.0232	.0571	.0527	.0300	.0265

TABLE 5.2: STANDARD DEVIATIONS FOR CV-TIMING RELATIONSHIPS SHOWN FOR EACH SPEAKER AND SEQUENCE TYPE IN SECONDS. THE CELL FOR THE VARIABLE HAVING THE LOWEST STANDARD DEVIATION FOR EACH SEQUENCE LENGTH, SEQUENCE TYPE, AND ALL SEQUENCES IS SHADED.

We also want to check measurements of the timing stability of the V2 ANCHOR to the onset consonant alone and the coda sequence alone in the coda+onset sequences. In particular, we consider the stability of the center of the *onset* consonant to the V2 ANCHOR as this was the most stable CV-timing relationship observed by Browman and Goldstein. We also want to check the CV -timing of the coda sequence in the $CC(C)$ #C cases, a measure not detailed by Browman and Goldstein. Browman and Goldstein do not find any stable CV- or VC-timing relationships to individual consonants in a sequence; they identify stable relationships with the sequence edge or

center. However, where possible, we want not to discount the possibility of the CV- or VCtiming relationship being sensitive to whether the sequence is tautosyllabic with the adjacent vowel--particularly as we saw above that V1 length is sensitive to the number of tautosyllabic consonants. The table below shows the standard deviations for the following intervals: ONSET C-CENTER to V2 ANCHOR, ONSET FIRST EDGE to V2 ANCHOR, and CODA C-CENTER to V2 ANCHOR for the coda+onset sequences. (Recall that the measurement of SEQUENCE LAST EDGE given above is the same as ONSET LAST EDGE for these sequences and that SEQUENCE FIRST EDGE is the same as CODA FIRST EDGE in these sequences.) Where these measures are more stable than any of the measures shown above for coda+onset sequences, they are shaded.

standard deviations in (s)	$Speaker \rightarrow$	mean s.d. across Speakers	A	B	K	М	S
ONSET C-CENTER to V2 ANCHOR	$\mathrm{coda} + \mathrm{onset}$ sequences	.0164	.0133	that financial .0192	.0206	.0157	.0133
ONSET FIRST EDGE to V2 ANCHOR	coda + onset sequences	.0162	.0157	.0239	.0231	.0044	.0139
CODA C-CENTER to V2 ANCHOR	coda + onset sequences	.0298	.0281	.0434	.0258	.0219	.0297

TABLE 5.3: STANDARD DEVIATIONS FOR CV-TIMING RELATIONSHIPS FOR ONSET CONSONANT AND CODA CLUSTER IN CODA+ONSET SEQUENCES SHOWN FOR EACH SPEAKER IN SECONDS. THE CELL FOR THE VARIABLE HAVING THE LOWEST STANDARD DEVIATION IN TABLES 5.2 AND 5.3 FOR CODA+ONSET IS SHADED.

Based on Tables 5.2 and 5.3 for the CV organization of the sequences, we can conclude that, of the variables measured here, the interval from the LAST EDGE of the sequence to the V2 anchorpoint is the most stable interval for coda only sequences of the form $C(C)(C)$ #. This is the case for all speakers and for the mean across speakers. That is, when all the consonants are in a different word (and syllable) from V2, then it's the last, closest, point in the consonant sequence that is most stably timed with respect to the vowel. For the coda+onset sequences, however, the most stable CV-timing relationship is from the C-CENTER of the *onset* consonant to the V2 anchorpoint. This is the case for four of the five speakers. The most stable relationship for the other speaker is also to the onset consonant, to its FIRST EDGE. That is, if any consonant in the consonant sequence is in the same word (and syllable) as V2, then that consonant is more strictly timed to the vowel, but from its center (or onset), not from its closest point.

This result is in accordance with the findings of Browman and Goldstein. However, the underlying syllable affiliation does appear relevant in determining CV global timing of a sequence. It is the interval from the coda-only clusters' LAST EDGES that is most stable with respect to the V2 anchor point in the following word, not from their C-CENTERS. In addition, for three speakers and the mean across speakers, the CV-timing interval bounded by the LAST EDGE for the coda-only sequences is also less variable than the CODA C-CENTER interval for the coda+onset sequences. This is a further indication that while the onsets may be timed to their word by their C-CENTERS, the codas of a preceding word are timed rightward by their LAST EDGE.

Recall that Browman and Goldstein also found for the VC-timing relationship that the across-word interval, e.g. V#C, was more variable than the within-word interval, e.g. VC#. We can make a similar comparison for the CV relationships here. A comparison between the stability of the LAST EDGE of the sequence for the across word (coda+onset) and within word (coda only) tokens shows no tendency for the former to be more variable than the latter. However, a more reasonable comparison might be to consider the intervals determined above to be the most stable for each type--i.e. the LAST EDGE for the across-word sequences and the ONSET C-CENTER for the within-word sequences. In this comparison, we do observe lower variability in the within-word cases for every subject.

$5.2.2$ **VC TIMING**

The following table shows the standard deviations for the VC-timing relationships for each speaker. Recall that the anchorpoint (V1 ANCHOR) chosen for this relationship to be analogous to Browman and Goldstein's experiment was the midpoint of the acoustic closure for the preceding bilabial consonant. The most stable variable for each set for each speaker and for the mean for all speakers is shaded.

standard deviations in (s)	$Speaker \rightarrow$	mean s.d. across Speakers	A	В	K	\mathbf{M}	$\mathbf S$
V1 ANCHOR to C-CENTER	C#	.022	.0307	.0224	.0194	.0202	.0184
	$\overline{CC#}$.0236	.0306	.0240	.0289	.0134	$\overline{.0210}$
	CCC#	.0297	.0312	.0220	.0364	.0261	.0331
	$\overline{CC\#C}$.0183	.0228	.0200	.0364	.0138	.0125
	CCC#C	.0338	.0166	.0354	.0552	.0277	.0342
	coda only sequences	.0471	.0442	.0440	.0681	.0397	.0394
	$\overline{\text{coda} + \text{onset}}$ sequences	.0510	.0430	.0609	.0709	.0204	.0596
	all sequences	.0511	.0463	.0558	.0698	.0370	.0469
V1 ANCHOR to SEQ FIRST EDGE	C#	.0238	.0285	.0258	.0227	.0235	.0185
	$\overline{CC#}$.0253	.0247	.0242	.0418	.0129	.0227
	$\overline{CCC#}$.0340	.0366	.0366	.0207	.0234	.0508
	$\overline{CC\#C}$.0333	.0396	.0101	.0309	.0362	.0499
	CCC#C	.0424	.0465	.0293	.0631	.0180	.0551
	coda only sequences	.0464	.0384	.0429	.0673	.0428	.0408
	$\overline{\text{coda} + \text{onset}}$ sequences	.0550	.0536	.0420	.0799	.0281	.0714
	all sequences	.0502	.0449	.0425	.0714	.0390	.0530
V1 ANCHOR to SEQ LASTEDGE	C#	.0244	.0341	.0213	.0207	.0250	.0211
	$\overline{CC#}$.0297	.0444	.0316	.0256	.0189	.0282
	CCC#	.0388	.0311	.0308	.0589	.0370	.0362
	$\overline{CC\#C}$.0312	.0345	.0278	.0464	.0282	.0192
	CCC#C	.0660	.0849	.0631	.0398	.0645	.0777
	coda only sequences	.0593	.0597	.0564	.0833	.0464	.0508
	$\overline{\text{coda} + \text{onset}}$ sequences	.0614	.0546	.0558	.0919	.0410	.0636
	all sequences	.0721	.0750	.0783	.0953	.0494	.0624

TABLE 5.4: STANDARD DEVIATIONS FOR VC-TIMING RELATIONSHIPS SHOWN FOR EACH SPEAKER AND SEQUENCE TYPE IN SECONDS. THE CELL FOR THE VARIABLE HAVING THE LOWEST STANDARD DEVIATION FOR EACH SEQUENCE LENGTH, SEQUENCE TYPE, AND ALL SEQUENCES IS SHADED.

As in the case of the CV-timing relationships, we also want to consider the possibility that something other than the *sequence* edges or center might form the most stable interval to the VCanchor point. Below Table 5.5 shows the standard deviations for the variables V1 ANCHOR to CODA C-CENTER, V1 ANCHOR to ONSET C-CENTER, and V1 ANCHOR to ONSET FIRST EDGE for the coda+onset sequences. When these measures are more stable than those presented in Table 5.4 for coda+onset sequences, they are shaded.

standard deviations in (s)	$Speaker \rightarrow$	mean s.d. across Speakers	A	B	K	M	S
V1 ANCHOR to CODA C-CENTER	coda + onset sequences	.0281	.0192	.0623	.0252	.0094	.0244
V1 ANCHOR to ONSET C-CENTER	$\text{coda} + \text{onset}$ sequences	.0421	.0389	.0837	.0193	.0192	.0495
V1 ANCHOR to ONSET FIRST EDGE	coda + onset sequences	.0433	.0412	.0824	.0187	.0238	.0504

TABLE 5.5: STANDARD DEVIATIONS FOR VC-TIMING RELATIONSHIPS FOR ONSET CONSONANT AND CODA CLUSTER IN ONSET+CODA SEQUENCESSHOWN FOR EACH SPEAKER IN SECONDS. THE CELL FOR THE VARIABLE HAVING THE LOWEST STANDARD DEVIATION IN TABLES 5.4 AND 5.5 FOR CODA+ONSET IS SHADED.

The first point to note about these results is that *none* of these measures define a timing relationship used by all speakers. Recall that Browman and Goldstein found the VC-timing relationship to be anchored to the FIRST EDGE of the sequence regardless of the underlying syllable affiliation of the consonants. We see in Tables 5.4 and 5.5 that the VC-timing interval for the coda+onset sequences is most stable from the V1 ANCHOR POINT to the CODA C-CENTER, not the sequence FIRST EDGE, for three speakers and for the mean across speakers. (The other two speakers favor the onset FIRST EDGE and the sequence FIRST EDGE.) For the coda-only sequences, the standard deviation for three speakers and the mean standard deviation across speakers is lowest for the interval from the FIRST EDGE of the sequence to the V1 ANCHOR POINT. However, the other two speakers do prefer the CODA C-CENTER here as well. (N.B. these are not the same two speakers in each case.) There seems to be some support here for Browman and Goldstein's predictions of the importance of the FIRST EDGE of a sequence in timing it to a preceding vowel. For one speaker, the sequence FIRST EDGE is also the most stable point for a coda cluster followed by an onset consonant, as found by Browman and Goldstein for sequences of a single coda consonant followed by an onset cluster. However, two speakers prefer the C-CENTER of the coda cluster as the most stable timing point for coda clusters followed by an onset consonant. Additionally, for four of the speakers in sequences of a coda cluster followed by a single onset consonant, the stability of the interval from the V1 ANCHOR POINT to FIRST EDGE of the sequence is not as great as that to the CODA C-CENTER, where very low standard deviations are observed. For the speaker who prefers the ONSET FIRST EDGE over the CODA C-CENTER, the standard deviation for the CODA C-CENTER interval is still substantially less than that of the sequence FIRST EDGE inteval.

5.2.3 CONCLUSIONS

Browman and Goldstein's (1988) study on which this experiment was based showed consonants in sequence to be organized with respect to the preceding vowel by the left edge (our initial edge) of the first consonant in the sequence, regardless of its syllabic affiliation. It also showed onset sequences to be organized with respect to the following vowel by their C-Centers. We attempted to expand this work by considering more repetitions, more sequence types-especially coda clusters, more timing metrics, and more speakers. Additionally, the CV acrossword timing relationship not addressed in Browman and Goldstein's study was considered here.

The following hypotheses were suggested in Chapter Two (maintaining their original numerical designations).

H7: The most stable VC organization involves the initial edge of the first consonant in the sequence.

H8: In a coda sequence followed by a single onset consonant, the most stable CV organization involves the C-Center of the onset consonant sequence.

Browman and Goldstein did not examine the CV timing relationship for coda consonants or report a comparison between the C-Center for the onset sequence versus the C-Center for an entire coda+onset sequence for the CV timing relationship. It could be that the CV relationship, like the VC, is unaffected by the canonical syllable affiliations of the consonants. This possibility suggested the following alternative hypothesis:

H9: For a consonant sequence, the most stable CV organization involves the C-Center of the entire consonant sequence.

None of these hypotheses received unequivocal support in this experiment. Hypothesis eight is fairly well supported and, in turn, hypothesis nine is not. Hypothesis seven receives mixed support for the coda clusters but is not supported for the coda+onset sequences, in which the interval from the V1 anchor to the coda C-Center was most stable.

Our findings suggest that timing relationships which cross a word boundary, such as that of an onset to the preceding vowel and of a coda sequence to a following vowel, attend to the near edge of the consonant sequence. That is, the edge of maximum displacement of the consonant nearest the word boundary appears to mark the edge of a relatively stable interval between that consonant and the vowel in the following/preceding word. However, the withinword timing relationships appear to attend more to the center of the tautosyllabic consonants; although, this is less clear in the VC-timing relationship. That is, the temporal mean of maximum displacement(s) for the consonant(s) in that word appears to mark the edge of a relatively stable interval between the marginal consonants of that word and the following/preceding vocalic nucleus. Additionally, the experiment above does not support a theory of timing for English consonant sequences in which the underlying syllable affiliations are irrelevant. The stability and nature of the timing relationships were found to differ depending on whether the consonants in the sequence were tautosyllabic with the anchorpoint. **The** intersection of these two findings--the support for global timing relationships and the relevance of underlying syllable affiliation-suggests that an avenue of future research might be the identification of gestural complexes through more elaborate manipulation of prosodic structure.

However, we have also seen that each speaker and each sequence is not necessarily consistent in what timing arrangement is most stable. These individual differences suggest that a single implementation of intergestural timing will not successfully predict timing for all instances or for all speakers. In order to have predictive value, a model of timing must have a mechanism which allows for variability. Johnson, Ladefoged and Lindau (1993) suggest that there is a "range of possible patterns of articulatory organization for speech production defined by functional utility and social convention..." (p. 712) and go on to emphasize that individuals will vary, in rather unpredictable ways, in their articulatory strategies. The concept of a range of permissible phasing relationships and the admissibility of variation from speaker to speaker and token to token will be elaborated in Chapter Seven.

CHAPTER SIX: EXPERIMENT FOUR--RATE EFFECTS ON CONSONANT OVERLAP, SHORTENING, AND **REDUCTION**

6.0 **OVERVIEW**

What do we do when we talk faster? That is, which aspects of articulation vary as speech rate increases? We know that faster speech rates cause a succession of phonological units to occupy less total time--presumably, that is what we mean by "faster." There are at least two conceivable ways in which this could happen. 1) The *duration* of each component articulatory unit might shorten (with some units being shortened more than others). As a consequence of decreasing duration, the spatial magnitudes of these articulations might also reduce (Lindblom, 1963, 1964). However, researchers have also suggested (Fowler, 1977; Gay 1981; Browman and Goldstein, 1990b) another view of fast speech proposing the following. 2) An increase in the *relative overlap* of units might yield the overall shorter duration of a sequence. Such changes have been shown for a single articulatory subsytem by Löfqvist and Yoshioka (1981b) and Munhall and Löfqvist (1992). Clearly, increased overlap and decreased component duration are not mutually exclusive. This experiment compares these two possible mechanisms for talking faster.

FIGURE 6.1: SCHEMA OF POSSIBLE WAYS IN WHICH SPEECH RATE MAY BE INCREASED.

Three issues are addressed experimentally. First, we determine whether the relative timing, *i.e.* overlap, of a C#C sequence changes as a function of rate. Increased temporal overlap in fast speech has been demonstrated between vowels and consonants (Gay, 1981) and has recently been hypothesized more generally to account for certain casual speech processes (Browman and Goldstein, 1990b). Gay found that "the duration of segmental units, the displacement and velocity of articulatory movements, and the temporal overlap between individual segments undergo nonlinear transformations during changes in speaking rate" (1981, p. 158). Gay interpreted this discovery as reflecting a restructuring of the temporal pattern of an utterance rather than a simple change in the spacing of motor commands. Additionally, we want to know whether rate affects articulatory timing in the same way for *different* consonant sequences.

Secondly, we also consider whether individual consonants in a C#C sequence *shorten* as rate increases. Recall that Gay (1981), like Kozhevnikov and Chistovich (1965), determined that duration changes were not distributed proportionally across consonant and vowel segments. I ask the parallel question of whether timing changes in a consonant sequence are instantiated in the same way across the consonants regardless of their place, manner, and syllabic position. That is, do speakers control rate by decreasing the duration of both consonants in the sequence in the same way?

Thirdly, as speaking rate may affect spatial reduction, we test whether the consonants in a C#C sequence reduce in lingua-palatal contact as speech rate increases. Concerning spatial displacement, Gay (1981) compares Lindblom's (1963, 1964) reasoning that the degree of undershoot is directly proportional to duration with his (Gay's) 1968 and 1974 work and Kent's (1970) work suggesting that rate may cause changes in articulatory effort or velocity signifying a reorganization in muscle forcing function along with a temporal reorganization. Gay says, "while target undershoot commonly accompanies an increase in speaking rate, it is by no means ubiquitous" (1981, p. 152-3). We consider specifically whether an increase in speech rate causes all the consonants in a sequence to reduce or whether this reduction is limited to certain consonants or certain syllabic positions. In particular, I evaluate a recent proposal (Barry, 1992) within the task dynamics framework (Saltzman and Munhall, 1989) to explain why coronals show more spatial reduction than velars. This proposal models the tongue tip as a massless articulatory subsystem, unlike the tongue body. It predicts a facilitation of coronal lenition due to the tongue tip's capacity for more rapid changes in direction of movement when confronted with competing demands. Recall that we have already evaluated this proposal with respect to the findings in Chapter Three. Alternatively, the very fact that the tongue tip can move faster might suggest that coronals are likely to be less lenited, as they would have little difficulty in making the articulation even in constrained circumstances.¹

6.1 **DATA**

In this experiment, four sequences-sg, gs, gd, and $d\mathbf{g}$ -were recorded for five speakers in a carrier sentence with reasonably large variations in speaking rate. The sequences were recorded in randomized blocks of four carrier sentences of the type "Say baC Cab again." Rate variation was created by asking the speaker to speed up for each sentence in the block of four. The experimenter cued the speakers with the words "Normal," "Medium," "Faster," and "Fastest" successively. It should be emphasized that these rate levels do not serve as variables in a categorical analysis, but rather, served only to engender a wide range of rate variation. Rate was considered as a continuous variable. A total of 640 utterances was recorded. (Of these, 11) are not included in the analysis, 8 due to complete lack of contact for a consonant and 2 due to data processing error.) A measurement of speaking rate from the carrier sentence was made by calculating the time from the frame after the end of any contact for the CC sequence through the peak contact for the **g** in "again," the last word of the carrier sentence. (See Chapter Two for a

 1 I thank Peter Ladefoged for pointing out this possibility to me.

complete discussion of the experimental method.) This rate measure is regressed against measures of the temporal interval between C1 and C2 onsets and C1 and C2 peaks (Δ ONSETS, Δ PEAKS), a measure of relative overlap (C1 OVERLAP (%)), and individual consonant duration and displacement (PEAK REGIONAL CONTACT) for the sequences. The following range of rate variation is observed for each sequence and each speaker. Remember that a higher value for the rate measure indicates a longer interval in the carrier sentence; hence, a *slower* speaking rate.

S peaker \Rightarrow			D	M		Sequence \Rightarrow	sg	gs	£Ο	
Mean Rate (c	\sim ر ے.	ر_ر	ر__	-			.	.	ر	ر
Range	h	ر…			O		ر بے ا	.	. .	

TABLE 6.1: MEAN OF RATE MEASUREMENTS AND RANGES SPLIT BY SPEAKER AND SEQUENCE. (THE LOWER THE RATE MEASURE, THE FASTER THE SPEECH)

6.2 **RESULTS**

Decreases in latency and increases in overlap are generally observed as rate increases, although not every combination of speaker and sequence shows this pattern.

6.2.1 EFFECTS OF SPEAKING RATE ON CC TIMING

Table 6.2 details the goodness of a first order fit in the regression analyses and the slope of the fitted line for each pair of speaker and sequence; only the significant $(p<.05)$ fits and slopes are shown. The significant fits and slopes are also shown for each sequence with the speakers pooled. A good fit means that as rate increased, the latency or overlap also increased in a relatively linear fashion. The higher the r^2 , the more linear the relationship.

Seq \Rightarrow Speaker µ	Measure	$\underset{r^2(m)}{\text{sg}}$	\mathbf{g} s $r^2(m)$	gd $r^2(m)$	dg $r^2(m)$
	ΔONSETS	.603(.357)	.354(.19)		
$\mathbf A$	Δ Peaks	.519(.299)	.213(.24)		.087(.07)
	C1 OVERLAP			.258(1.355)	
	ΔONSETS				.18(.129)
${\bf B}$	Δ PEAKS	.117(.112)	.182(.236)		
	C1 OVERLAP				
	ΔONSETS	.429(.253)			
$\mathbf K$	APEAKS	.335(.207)	.507(.267)	.233(.166)	
	C1 OVERLAP				
	ΔONSETS	.258(.23)	.124(.102)	.166(.15)	
M	APEAKS	.179(.165)			
	C1 OVERLAP	$.247(-1.212)$			
	AONSETS	.284(.111)	.29(.095)	.443(.122)	
S	APEAKS				.138(.182)
	C1 OVERLAP				
	pooled: AONSETS	.265(.241)	.255(.174)	.32(.306)	.152(.109)
	Δ Peaks	.285(.167)	.3(.189)	.077(.073)	.203(.175)
	C1 OVERLAP	101(-1.096)	$.07(-.826)$	$134(-1.116)$.061(-.904)	

TABLE 6.2: SIGNIFICANT (P<.05) PEARSON r^2 and slope values (m) FOR A LINEAR FIT FOR AONSETS. **APEAKS, AND THE PERCENTAGE OF C1 OVERLAP REGRESSED AGAINST RATE MEASURE.**

Perhaps the first thing one notes is that the correlation of these measures with rate is not very high. In part this is due to the limited resolution of the data (100 Hz sampling rate), but it's also the case that rate is only one of many factors influencing articulatory timing in these sequences.

Rate has a significant effect on the articulation of each consonant sequence such that absolute latency between consonants decreased and overlap increased as speaking rate increased.¹

However, not all speakers and sequences showed equivalent effects. By comparing the columns in Table 6.2, we can see that rate has the least influence on the dg sequences. This presumably is due to a ceiling effect, whereby dg is so overlapped even at slow rates that only a very minimal additional increase due to the rate change is evidenced. Speaker S also differed from the other speakers in that he tended to *decrease* overlap for the back-front sequences as rate increased. Among the other speakers, when the sequences (except dg) are pooled, Speakers B

 1 An examination of the 95% confidence bands for slope in the regression plots for each sequence (speakers pooled) exhibits no change in the sign of the slopes.

and M show significant effects of rate on both overlap and absolute latency in the predicted direction, while for Speakers A and K the effect of rate reaches significance for the measures of absolute latency.

The result when data from all speakers is pooled is that there is a relatively linear increase in coproduction, i.e. overlap, as rate increases. This can be seen in the consistent significant influence of rate on timing in the pooled data, which considers many more tokens than each speaker-by-sequence cell considered separately. Given a linear relationship, speaking rate accounts for up to 32% of the timing variation with the speakers pooled, as demonstrated by the r^2 values shown in Table 6.2.

The regression plots for the absolute (\triangle ONSETS and \triangle PEAKS) and relative (C1 OVERLAP) temporal offset between the consonants with speakers pooled are shown below in Figure 6.1 to illustrate differences among the four sequences.

 \mathcal{A}

FIGURE 6.2 A-L: REGRESSION PLOTS FOR ΔONSETS, ΔPEAKS, C1 OVERLAP (%) REGRESSED AGAINST RATE.

Several aspects of these plots are noteworthy. First, as rate decreases, generally overlap increases and latency decreases. The increase in overlap appears to be relatively linear. However, some differences among the sequences are apparent. One can see for dg, particularly in the case of C1 OVERLAP (Figure 6.2c), the ceiling effect described above. Regardless of speaking rate, there is a strong tendency for this sequence to be produced in a completely overlapped fashion. The data points tend to cluster around 100% overlap. In this case, the particular sequence appears to have a greater influence on timing than the rate at which it is spoken.

We noted above that Speaker S differed from the other speakers. In a closer examination of Speaker S, it seems that his front-back sequences show little effect of rate on relative latency; although absolute latency always shortens with increased rate. Like the other speakers, dg has almost complete overlap in most instances; se also shows relatively little rate effect (but does have a first order fit with a positive slope for C1 OVERLAP, $m = 257$, $r^2 = 0.031$). His back-front sequences, however, tend to show less overlap at faster rates. These differences are shown in Figure 6.3.

FIGURE 6.3: REGRESSION PLOTS FOR C1 OVERLAP (%) REGRESSED AGAINST RATE FOR SPEAKER S

Thus it appears for these sequences that something else about the articulation of the back-front was more influential on their timing than any advantage gained by increasing overlap at faster speech rates.

In summary, this experiment does provide articulatory evidence that the relative overlap of two units across a word boundary increases as rate increases. The exceptional behavior of dg and Speaker S, as outlined above, have been omitted from Figure 6.3 below to present a picture of the overall change in relative timing with respect to rate. Figure 6.3 plots the percentage of the way through C1 that C2 contact was initiated, against speaking rate. We see that as speaking rate increases, C2 starts earlier in C1, the r^2 for the linear fit being .24.

FIGURE 6.4: REGRESSION PLOTS FOR THE PERCENTAGE OF C1 AT WHICH C2 INITIATES REGRESSED **AGAINST RATE.**

The evidence shows that overlap does increase with rate and that this increase is generally linear.

$6.2.2$ **EFFECTS OF SPEAKING RATE ON INDIVIDUAL CONSONANTS**

Next recall the questions regarding shortening and reduction at faster speaking rates. The maximum displacements in the front and back regions for both consonants in each sequence were examined to determine if speaking rate affected displacement. The linear regressions (with speakers pooled) were calculated for each of the three consonants in each syllabic position. The only significant effect of rate was found to be for the front consonants in coda, i.e., C1, position. The effect was also greater for **d** than for **s**. **g** didn't reduce in either syllabic position.

FIGURE 6.5: REGRESSION PLOTS FOR PEAK DISPLACEMENT REGRESSED AGAINST RATE

Experiments One and Two found that both front and back stop consonants were subject to reduction in coda as opposed to onset position. The fricative s was not found to be subject to spatial reduction in coda position in the way that **d** was. However it appears here that both **d** and s, but not g, become reduced at increased rates when in coda position. Thus it appears that in the case of s the combined effects of speaking rate and syllable position have a more profound effect on spatial displacement than syllable position alone. On the other hand, rate, unlike syllable position, does not have any discernible effect on the displacement of the back consonant in either syllabic or on the front consonants in onset position.

 \sim

In general, all the consonants also became shorter as speech rate increased. This relationship is shown in Figure 6.5. Even the extremely overlapped **d** shortened with rate. The one exception to this pattern was the **d** in onset position which showed no effect of rate on its duration. When this exception is examined more closely, however, we find that three of the five speakers also show significant shortening here with r^2 's for a linear fit as high as .573. Two speakers, Speaker B and Speaker M, have no significant shortening.

Thus, overall, we do find support for the shortening of individual consonants as speaking rate increases. This shortening seems to take place regardless of the place and manner of the individual consonant or its syllabic affiliation.

FIGURE 6.6: REGRESSION PLOTS FOR ONSET AND CODA DURATIONS AGAINST RATE

6.3 **DISCUSSION**

The results evidence both mechanisms of faster speech discussed in the introduction to this experiment. As speaking rate increases, individual consonants shorten in duration and a relatively linear increase occurs in the overlap of the articulations. The particular sequences studied, however, do not behave identically. With respect to overlap, rate has only a minimal effect on d#g, which remains almost completely overlapped at all rates.

With respect to duration, we find that individual consonants generally shorten in duration as speaking rate increases. But with respect to spatial reduction, we find differences based on a consonant's place of articulation and syllable position. Target undershoot has been suggested to be proportional to duration (Lindblom, 1963, 1964). Others argue (Gay, 1981) that while undershoot commonly accompanies increased speaking rate, it is not always present. Indeed, we find that the consonants are not equally susceptible to reduction, although all shorten in duration. Reductions in lingua-palatal contact are the least consistent concomitant of fast speech, being found only for a subset of the faster articulations, namely, coda coronals. At faster rates, coda d reduces most, coda s less, and onset consonants and g not at all.

These results bear on a hypothesis about why coronals reduce more. Barry recently (Barry, 1992) proposed that the tongue tip be modeled in a task dynamics approach (Saltzman and Munhall, 1989) as a massless articulatory subsystem, unlike the tongue body. This predicts a facilitation of lenition in coronals because of a capacity for more rapid changes in direction of movement when confronted with competing demands. However, as noted in Chapter Three, this approach does not account for the observed importance of prosodic affiliation in reduction or the difference between coronals in amount of reduction. The different behavior of coronal stop versus fricative and coronal onset versus coda shows that this reduction cannot be due only to the low mass of the tongue tip. Such an approach would require different masses depending on manner of articulation and prosodic affiliation. Such an approach would keep "mass" from having any obvious physical interpretation.

In sum, this experiment demonstrates that for a set of obstruent consonants in consonant sequences spanning a word boundary, talking faster means decreasing articulatory durations and increasing the overlap between successive articulations. At the same time, it shows that how we adjust our speech rate depends on linguistic factors, both featural and prosodic. This experiment provides articulatory evidence that relative timing across word boundaries changes with speaking rate. This timing is functionally dependent on the units involved. We conclude that in implementing speaking rate the linguistic phonetic system must be sensitive to prosodic constituency, differences among consonants, and the interaction of rate and reduction.

CHAPTER SEVEN: THE PHASE WINDOW MODEL AND VARIABILITY IN SPEECH TIMING

"...to what extent [can] the notion of an 'ill-formed' word be reduced to that of a 'statistically improbable' word"

Pierrehumbert (in press)

7.0 **WHERE'S INVARIANCE?**

As we may recall from Chapter One, the search for invariant timing relationships in speech has had mixed success. In examining movements composing a *single* gesture, many stable aspects of temporal coordination have been found. Within a single articulatory movement, studies of articulatory kinematics have suggested that the relation of peak velocity to displacement (and in some studies the relationship of this ratio to duration) is the dynamic *intra*gestural property which remains stable across variation in linguistic and extra-linguistic contexts. (Kozhevnikov and Chistovich, 1965; Ohala, Hiki, Hubler, and Harshman, 1968; Mermelstein, 1973; Sussman, MacNeilage, and Hanson, 1973; Kuehn and Moll, 1976; Ostry and Munhall, 1985; Gracco, 1988; Gracco and Abbs, 1989; and Vatikiotis-Bateson and Kelso, 1993; see also Ostry, Keller and Parush (1983), Munhall, Ostry and Parush (1985), and Kelso, Vatikiotis-Bateson, Saltzman, and Kay (1985) regarding systematic rate and stress effects on this relationship.) Between gestures which are traditionally considered to constitute a segment, researchers have found instances of invariant and bimodal timing relationships (Munhall, Löfqvist, and Kelso, 1986; Löfqvist, 1991 citing Löfqvist and Yoshioka, 1984; Löfqvist, 1981c; Krakow, 1989; Browman and Goldstein (in press) citing Sproat and Fujimura (1993)). We emphasize that it is not just *any* two gestures which display this tight timing relationship but rather gestures which have long been considered to belong to the *same* segment.

However, the existence of invariant phasing relationships between gestures composing what would traditionally be considered *different* segments is not evidenced. The few studies of such timing relationships have had methodological flaws or have not found evidence of stable, *i.e.* invariant, timing. See Löfqvist (1991) and Keller (1990) for overviews. The experimental evidence of Nittrouer et al. suggests that:

...the intersegmental organization of gestures is a function of the utterance being produced. In other words, the phase relations between articulatory gestures used in the production of adjacent segments varies systematically based on linguistic and nonlinguistic structure, which includes speaking rate, stress pattern, syllable structure, and consonant identity. (Nittrouer, et al., 1988, p. 1659)

(But see Kelso, Saltzman, and Tuller (1986) whose results differ from the attempted replication experiment of Nittrouer et al., 1988.) Löfqvist (1991; examining data from Löfqvist, 1984) reports results parallel to those of Nittrouer et al., *i.e.* no evidence of invariant intersegmental timing. Fujimura (1986) says "the timing relation observed among a pair of articulators varies in an intricate manner" (p.229).

Kent (1986), although still maintaining the existence of relative invariance and a set of only a few acceptable timing relationships, does note that "variability in timing is not continuous nor unprincipled" (p. 240). Butcher's (1989) remarks on contextual variation take this sentiment futher. He says, "[t]his [contextual] variability is itself variable... in that in some instances its occurrence and extent is more predictable than in others" (p.39). Like myself and others, Butcher feels that one of the goals of the speech scientist is to define the nature and range of this variability.

Two separate issues are relevant in understanding speech timing: 1) what is coordinated; and 2) how is it coordinated. It is the latter question with which this discussion is primarily concerned. Any theory of timing must account for the variability observed in intergestural coordination. The preceding chapters have identified timing variability between speakers, within speakers, between gestures, and within gestures as a function of prosodic position. This chapter will outline a new framework for speech timing which allows timing variability to be understood within an intrinsic timing approach. The mechanism for timing proposed here, called the PHASE WINDOW Model, adopts phasing relationships such as those pursued in Browman and Goldstein's Articulatory Phonology. We argue for crucial differences, however.

Phasing relationships synchronize phase angles in two gestures (or perhaps also points in multi-gesture units) and have formed the basis for implementing timing in this framework. However, unlike Browman and Goldstein, we do not pursue the use of phasing rules to implement inter-gestural coordination. Articulatory Phonology currently implements each particular phasing relationship with a rule or rules specifying an invariant coordination. Browman and Goldstein (1991) state that "[t]here is a potential continuum [of overlap] ranging from complete synchrony...through partial overlap...to minimal overlap" and that "there are no a *priori* constraints on intergestural organization within the gestural framework. The relative 'tightness' of cohesion among particular constellations of gestures is a matter for continuing research" $(p. 319)$. In principle, any point in a gesture could be phased to any point in another gesture thereby yielding an infinite number of possible phasing relationships. The lack of principled constraint on possible phasings makes this approach over powerful. However, the postulation of phasing rules which only have access to three points in a gesture--the onset, target, and perhaps release (Browman and Goldstein, 1990a and in press)--is empirically overly constraining and theoretically unprincipled. While no one can doubt that certain timing relationships give rise to qualitative differences (see Goldstein, 1989, 1990 and Ohala, 1990), why would exactly these three phase angles and no others, e.g. 0° not 1° , exist for timing rules?

This use of phasing rules, at least in this form, therefore appears unconstrained in theory and overly constrained and unprincipled in practice. These facts lead us to believe that the instantiation of linguistic timing in terms of a set of phasing rules is of limited predictive value and is mostly useful as restatements of observed empirical facts. We formulate an alternative below which allows, but constrains, variability in phasing relationships.

7.1 **OPTIONS FOR PHASE RELATIONS**

A clearer understanding of the operations of association and phasing in Articulatory Phonology appears to be warranted in order to consider how variability in speech timing might be captured. There are two general approaches to phasing that currently seem to be formulatable within this framework.

The first is rather simplistic. It postulates a single set of phasing rules in a language which operate on every utterance to phase associated gestures, and allows any pair of associated gestures to be subject to only a single instance of phasing (although each may also be associated to other gestures as well). This is the rather rigid invariance approach. It states, for example, that the onset of the second consonant in a CC sequence is phased to be synchronous to 290° of the preceding consonant. Furthermore, it requires that no other rule later alter the timing relationship between this pair. Thus every CC cluster, regardless of speaker, rate, style, context, place, manner, stress, phrasing, etc. would be phased such that the onset of the second consonant is synchronized invariantly to the release of the first. This is clearly empirically inadequate.

The second approach is to say that there is a single set of phasing rules in a language which operate on every utterance to phase associated gestures, but any pair of associated gestures may be subject to a variety of phasing rules-i.e. may or may not be subject to one or more of a finite set of phasing rules. These rules would be specific to linguistic and extra-linguistic conditions, such as the well-formedness of the sequence. Such an approach might, for example, include a rule that phases the onset of the second consonant in a CC sequence following a stressed syllable to be synchronous to 290° of the first consonant in the sequence, but have a second rule phasing consonants in a CC sequence following an *unstressed* syllable such that the onset of C2 is synchronous with 270° of C1. A third rule might say that across a phrase boundary the phasing relationship is $C_2(0^{\circ})=C_1(300^{\circ})$. This approach allows variability in timing, *i.e.* different temporal organization in different contexts, but raises difficult questions of rule ordering/hierarchy. Additionally, its explanatory power is limited; why would these particular contexts call for these phasing rules and no others.

In order to capture linguistic variability in this way, issues of rule ordering, the representation of rule input and output, and potential rule interactions must be considered. Assuming that some pair of associated gestures meets the conditions specified by more than one rule, some hierarchy must be defined whereby the most "important" rule would be the last to apply and thereby determine the ultimate phasing relationship. Pursuing the above example, what is the final rule output in the case of a CC sequence following a unstressed vowel but across a phrase boundary? Supposing the phrasing-sensitive rule was more important, it would determine the ultimate phasing. In such a case the phasing should be exactly and invariantly $C2(0^\circ)=C1(300^\circ)$.

Furthermore, neither of these approaches predicts any within-condition variability. Any token of an utterance meeting conditions for the same set of rules should be phased invariantly. That is, repetitions of the same input should yield the same, invariant output timing organization.

With respect to consonant sequences, a single rule is proposed in Articulatory Phonology for the organization of consonant sequences. It, however, only applies to sequences meeting a certain condition *(i.e.* possible onset or coda). Currently in Articulatory Phonology, later rules changing phasing relationships would have to be admitted to allow for timing differences due, for example, to rate, stress, or style. In discussing their proposed phasing rule for consonant clusters, Browman and Goldstein (1988) suggest that it may need to be refined to include syllabification effects and articulator specific effects. It is not clear what they foresee as the nature of this rule refinement.

Neither of the two general approaches outlined above--context-insensitive (absolute) invariance or context-sensitive invariance--is entirely satisfactory. The first does not admit any variability in the timing of two associated gestures. Real biological systems are unlikely to operate in this manner. The second, multi-rule, approach also encounters the problem of variability, specifically in repetitions of the same utterance under the same conditions. Additionally, it offers little explant ory insight to the linguistic process of articulatory timing, not facilitating generalizability across similar rules.

We suggest that an alternative is possible--a probabilistic view of intrinsic timing. The various factors that influence coordination can be seen as competing simultaneously, each contributing effects on the ultimate probability distribution of the final outcome.

7.2 THE PHASE WINDOW MODEL

Our proposal is to allow variability in a single assignment of a phasing relationship rather than using a set of timing rules which operate sequentially to coordinate the articulatory units. We suggest that coordination is subject to different constraints and requirements when lexically contrastive versus when not. Temporal organization creates meaningful distinctions in the lexicon, e.g. voice onset time. In light of this, it seems reasonable to assume that temporal relations specified lexically are discrete and/or stable. This will be discussed further in section 7.6. However, *outside* the lexicon, we suggested that inter-articulator phasing relationships are not invariant. We propose here a probabilistic approach to intergestural phasing (and leave aside the question of what points stand in a phasing relationship). Our general approach is that a particular phasing relationship is constrained both physically (by biological inheritance) and language-specifically (by learning) to occur within a certain permissible window. We call this window the PHASE WINDOW. Maintaining an intrinsic timing framework, we suggest that there are upper and lower limits placed on a particular PHASE WINDOW which are determined by both system constraints (motor and cognitive) and language constraints (language-specific, learned permissible phasing relationships). Clearly the window defined by the latter constraints will be properly contained in that defined by the former. Utterance-specific (task-specific) influencers then act to weight the window but do not constrain it further.¹ The weighting of this PHASE WINDOW takes place in a probabilistic manner. We suggest, in accordance with Nittrouer et al., that linguistic and non-linguistic variables which differ from utterance to utterance determine where in the range of permissible overlap relationships a specific phasing is likely to be implemented.

Let's consider what the PHASE WINDOW for the oral constriction gestures in a CC sequence might look like. First, given the postulation that the two consonants are timed with respect to one another, we need to define what points stand in a phasing relationship. This question is not a focus of this thesis, although it was considered in Experiment Three. Let's assume that the onset of C2 is phased to some point in C1. Thus the relevant phasing relationship for this PHASE WINDOW is $Cl(x^{\circ})=C2(0^{\circ})$. Further, as outlined above, some crosslinguistic, *i.e.* universal motor and cognitive, constraints exist which limit the value of x. Let's postulate for our example that these limits are minimal and are something like a lower bound of 0° and an upper bound of 360°. That is, C2 may not start before C1 is activated or after its activation ceases entirely. It is quite conceivable that the correct upper bound could be higher. The PHASE WINDOW is, however, also additionally constrained in a way specific to English. Although very little cross-linguistic work has been done on this timing relationship, it does not seem unreasonable to assume that English consonant sequences are typically quite overlapped in comparison to other languages, as we find systematic perceptual assimilations (Byrd, 1992) and generally no acoustic releases of consonants in sequence (Jones, 1956; Catford, 1977; Hardcastle and Roach, 1979; Marchal, 1988). A language with systematically released consonants, like Tsou (Tung, 1964) or Salish (Flemming, Ladefoged and Thomason, 1993), would have a higher lower bound, and perhaps upper bound, on the window thereby yielding less overlapped sequences than in English. Furthermore, let's suppose for the sake of argument that English allows a wide range of possible timing relationships; as in fact appears to be the case. That is, in the case of our supposition, English has a large amount of variation in CC timing as compared to many other languages. These language specific constraints are learned by the child acquiring English. She learns that no more than a certain degree of overlap is allowed in speech for it to be intelligible to other speakers and that overlap is required to be at least a certain amount such that acoustic releases aren't present between the consonants. These constraints yield the CC PHASE

¹ Compare Kingston's (1990) suggestion that the timing cohesiveness (i.e. stability) between oral and laryngeal gestures within segments will be affected by the continuancy of the oral constriction.

WINDOW for English, for all (associated)² consonant clusters in general. The universal and language-specific boundaries are marked in the diagram below by thick and thin lines respectively.

FIGURE 7.1: THE PHASE WINDOW MODEL SHOWING THE COMBINED EFFECTS OF INFLUENCERS ON THE PHASE WINDOW.

We see that there is some probability, however small, of any value of x in the PHASE WINDOW occurring (otherwise it wouldn't be in the window). The combined influence of the linguistic and extra-linguistic conditions existing for the particular consonant cluster in the particular utterance determine the final probability density for the window. The more alike the contextual

²There may of course be some surface consonant clusters which are not phased to one another but only to the following vowel, e.g. potato or support (P. Keating, p.c.).

effects are from token-to-token, the more alike the combined *influencer* distributions will be. This will yield a high probability of similar organizations being realized in similar contexts-i.e. low token-to-token variability. Of course, the most interesting empirical and theoretical question is how this determination of the combined weighting of the PHASE WINDOW is arrived at.

We have seen in this study that a wide range of relative latencies between C1 and C2 can occur. Just as an overview of this range, consider the figure below.

bins (% of C1 at which C2 contact initiates)

FIGURE 7.2:

HISTOGRAM SHOWING THE FREQUENCY IN EACH BIN OF 5% FOR THE VARIABLE "C2 ONSET RELATIVE TO C1" FOR ALL CC SEQUENCES IN EXPERIMENTS ONE AND TWO.

This figure shows the relative latency for all the CC sequences in Experiments One and Two. These two studies manipulated overtly very few variables. While not intended to be a representation of phasing relationships, this figure does at least suggest that the timing of CC sequences does not demonstrate relative invariance. The modality of this distribution is not even clear, only a tendency for C2 contact to start sometime during C1 contact.

The system-specific and language-specific constraints suggested to apply to the PHASE WINDOW reflect issues previously considered in the literature. MacNeilage (1970) describes Hebb's (1949) contention that motor equivalence requires the use of learned perceptual information in addition to moment-to-moment information about ongoing motor activity. MacNeilage notes the similarity of these ideas to those of Piaget (Flavell, 1963, pp. 85-121 cited in MacNeilage, 1970). The nature of some *influencers* within the PHASE WINDOW Model will be identified by learning.

While not precisely analogous to our proposal, Turvey's discussion of ecological psychology is also relevant in considering a theory of timing which responds to task-specific requirements within a framework delimited by physical, systemic, and learned constraints. He says:

...Laws identify real possibilities. When circumstances--boundary conditions, constraints--are appended, actual events result...Nature, however, is not very economical with respect to patterns of coordination. There is a great diversity, with each pattern giving expression to the general laws and principles in very specific ways...Furthermore, in the province of coordinated movements, the circumstances appended to laws include intentions, plans, goals, and so on. Intentions function as exceptional boundary conditions on natural law...(Turvey, 1990, p. 941)

Our concept of a collection of influences on motor behavior is also in the spirit of previous comments in the literature. Keller's (1990) idea of mutual competition between codeterminers of speech timing and inclusion of perceptual and prosodic factors as influences on timing is in agreement with our theoretical proposal of a mechanism to incorporate motoric and non-motoric influences on speech timing in such a way as to allow variability in speech timing. We agree with Keller that:

...surface variability does not "hide' some ill-understood invariants under measurement error and articulator imprecision. On the contrary, surface variability is considered to be the inevitable and theoretically predicted concomitant of a communicative behavior capable of achieving its overall objectives by satisfying a large variety of competing demands...(Keller, 1990, p. $357)$

This approach suggests the pursuit of a research program to identify the factors influencing timing relationships and the nature of those influences. Gracco has recognized similar goals in the study of speech production noting that "it is becoming increasingly clear that any behavior is a reflection of multiple overlapping and interacting influences, each of which needs to be identified. The purpose of identifying the subcomponents is not strictly to assign function to structure but to evaluate their potential contribution to the overall process and hence allow development of realistic and biologically plausible working models of the system" (Gracco, 1992b, p. 27).

7.3 **EFFECTS OF INFLUENCERS ON THE PHASE WINDOW**

Let's consider how a variable may influence the probability density of the PHASE WINDOW. There are really three "dimensions" to consider here. First, a variable may cause a preference for a particular region of the PHASE WINDOW. This is related to how much overlap a contextual variable is associated with. For example, a fast speech rate will favor the "moreoverlapped" end of the window, and a slower speech rate the "less-overlapped" end of the window. Secondly, variables may differ in the extent of the window over which they have an influence. This corresponds to how much variability an *influencer* will allow. Lastly, the level of weighting or activation contributed by particular variables may differ. For example, the style of speech, casual versus formal, may have a greater influence on the final probability density than the rate of speech.

Some of the differences in timing that we have observed in the preceding studies can, tentatively (because of the unknown complexity of interactions between variables), be considered to evidence effects on the PHASE WINDOW. These include intrinsic influences such as constriction location and degree, influences of adjacent contexts, structural influences such as syllable constituency and boundary location, and the influence of speaking rate. We have seen that front consonants followed by back consonants tend to be more overlapped than back consonants preceding front ones. A front-back order in a consonant sequence might therefore weight a region in the "more-overlapped" end of the PHASE WINDOW. Similarly, stops were found to be more overlapped by a following stop than (sibilant) fricatives. The manner of C1 would then influence the final probability density of the PHASE WINDOW accordingly. Also we saw some evidence that onset clusters were less overlapped and less variable than coda clusters and heterosyllabic sequences. This would reflect a more narrow region weighted in the onset cluster context and a region more in the "less-overlapped" end of the window than for the other sequence types. In addition, although not completely clear in the preceding experiments, work by others (Hardcastle and Roach, 1979; Browman and Goldstein, 1988) has suggested that timing of the same gestures between words is more variable than within words. The presence or absence of a word boundary, or other prosodic boundaries, may influence the window over more or less narrow regions.

Let's consider just the influence of speech rate on the PHASE WINDOW. We observed in Chapter Six that rate has a roughly linear relationship to consonant cluster overlap. The schema below indicates the strength and region of influence in the phase window for particular speaking rates. The x-axis is the PHASE WINDOW, the y-axis shows sample planes representing probability distributions in a continuum of speech rate, and the z-axis indicates probability. $(N.B.:$ Only a single plane in the diagram is relevant for any particular utterance. This schema is intended to represent a three dimensional space, although only slices through the z-axis (speaking rate axis) of that space are shown to simplify presentation.)

FIGURE 7.3: SCHEMATIC FIGURE SHOWING THE EFFECT OF SPEECH RATE ON THE CC PHASE WINDOW

Recall that we found in Chapter Six that overlap increased at faster rates. This can be seen by the movement of the peak across the window. The linear nature of this change is emphasized by the line overlaid on the peaks. Suppose also, for the sake of illustration here, that variability in timing was found to decrease (i.e. coordination was more stable) at faster rates. Such a difference in timing variability would correspond to differences in the broadness of the distributions as rate changes. Such a difference is also illustrated in Figure 7.3.

Let's consider how the timing of a consonant cluster in a particular utterance would be described. Suppose the following: 1) the consonant cluster is sk , 2) it is in syllable onset position, 3) it is being spoken somewhat faster than "normal." Thus, we might find the following influences, each of which is represented by a plane schematically in the figure below: 1) consonant sequences of fricative-stop prefer less overlap, 2) onset clusters prefer less overlap, 3)

 $\hat{\mathcal{A}}$

medium fast clusters prefer somewhat more overlap. (Alternatively, the second influence could be seen as a mapping from a continuous variable of prosodic cohesiveness to the probability distribution in the PHASE WINDOW, much the same way as rate is. See Section 7.3.1 below for more on continuous versus categorical variables.) Note that, unlike Figure 7.3, the planes in this figure represent the *influencers* affecting a single utterance.

FIGURE 7.4 SCHEMATIC SHOWING MULTIPLE INFLUENCES ON THE CC PHASE WINDOW FOR AN #sk TOKEN AT A MEDIUM FAST RATE

The most difficult question is how the individual probability functions for each variable combine to determine the final probability density for the PHASE WINDOW for the utterance. It is likely that these interactions are in fact very complex. This complicated question is beyond the scope of this work, and in fact, would be very difficult to evaluate without a sophisticated computational model. The effects of each *influencer* could be added, convolved, overlaid³ and subjected to a peak-picking algorithm, or combined by any number of other possibilities. Here we outline two reasonable combinatorial possibilities. First, the probablity distributions of each *influencer* may simply be additive. The phase angle with the highest weighting or activation after the effects of all variables are added up specifies the phasing relationship with the highest probability of being realized. A second approach to combining effects of different *influencers* is to add the individual distributions, as before, and then take a weighted average in the form of the center of mass or centroid value. This value would be the output phasing. (This is the approach taken by Kosko (1993) for combining fuzzy sets.) These two approaches yield different empirical predictions. The first admits unpredictability and the possibility of multi-modal resultant distributions. The latter does not. However, testing for outliers and bimodality under identical *influencers* is, in practice, difficult. It would be impossible for an experimenter would have to ensure identical conditions on the part of the speaker for many tokens of an utterance.

The realization of the influences shown in the figure above will depend on which combinatorial procedure we use, the additive or the additive-centroid. (Note that because this study examined the *main* effects of certain variables on timing, we can make no claim about the nature of their interactions, and so can infer nothing about any weighting of the influences which might occur when they are combined. However, it would not be at all surprising to find that certain *influencers* were stronger than others.) In Figure 7.4 the additive process yields a timing with a high probability of being realized in the middle of the PHASE WINDOW but some chance as well of being realized in the more overlapped portion of the window. The additive-centroid combination does not admit any uncertainty in the outcome; it yields a single phase angle at which the CC timing occurs. Computationally and physiologically this approach is more appealing.

Cases in which substantially different behavior is predicted by the two combinatorial processes include tokens in which two disparate *influencers* are competing at opposite ends of the PHASE WINDOW; for example, as might happen in a very rapidly spoken onset cluster. The additive process predicts a bimodal distribution of the output timing across a large number of repeated tokens. The additive-centroid process predicts an output phasing consistently between the areas of the PHASE WINDOW preferred by each of the *influencers*. However, it's likely that there will nearly always be many synchronously operating *influencers* which prefer the middle range of the PHASE WINDOW, thus making instances of bimodality highly unlikely due to of the additive procedure used to incorporate all the *influencers*.

 3 Adding the distributions is in fact unrealistic, as it would eventually yield a flat line output.

7.3.1 DIFFERENT KINDS OF INFLUENCERS

Gracco (1992b) suggests two rationales for research focusing on developing models of motor control--"first, there is an inherent richness and intricacity [sic] to even the simplest problem of sensorimotor control, and second, [there is] an implicit assumption that higher functions such as cognition are not discontinuous with the lower level sensorimotor functions that implement them" (Gracco, 1992b), p. 27).

In the exploration of motor control, much attention has been given to physiologic factors determining timing. The efficiency of timing patterns has been determined in large part by physical factors such as energy and work requirements. There is an important way, however, in which speech movement differs from other types of body movement. For speech, the determination of efficient movement patterns must take into account a perceiver. That is, unlike control of the limbs in the study of gait, theories of articulatory coordination in speech timing must be able to account for the *communicative* goals, and hence communicative efficiency, of the movements. Is the listener in the room, in another room, hard of hearing, a non-native speaker, a child? All these factors could conceivably influence how the articulators are coordinated. The approach outlined above incorporates extra-linguistic influences in the same way as linguistic variables.

Recall that in Chapter 3, we suggested that acoustic goals might be taken into account in coordinating two gestures. This suggestion is in the spirit of Ladefoged, DeClerk, Lindau and Papcun (1972) and Johnson, Ladefoged and Lindau (1993) who have outlined an auditory theory of speech production in which speech movements are directed by auditory goals. Although acoustic influences on *timing* are not their focus, Johnson et al. do suggest that "the acoustic product of speaking is the crucial determinant of the *organization* of speech articulation" (p. 712, emphasis added). Ohala (1990) also states that the ultimate goals in speech are acoustic-auditory events and that acoustic goals can influence timing. Theories providing a specific mechanism for speech planning to utilize a speaker's knowledge about the crucial perceptual cues of his language offer greater insight into speech timing than those which rely solely on the physical properties of the movements themselves. According to the approach described here, timing is still implemented intrinsically, *i.e.* with reference to the temporal characteristics which are part of the linguistic unit, but is influenced by other knowledge as well.

We know that both linguistic and extra-linguistic variables may affect speech timing. We use "linguistic" here in a very narrow sense. We mean, for the sake of this discussion, categorical or phonological. We use extra-linguistic to refer to those variables which change continuously. We suggest that both types of *influencers* have commensurate types of effects on the PHASE WINDOW, although the shape of their probability distributions may differ. Consider that it is also not always clear whether a contextual influence is categorical or continuous. For example, resyllabification of the second consonant in a coda cluster to a following syllable might be a matter of *degree* or might be *all or nothing*. Conversely, changes in speech style and rate

might behave in categorical fashions or alter gradiently. Allowing phonological, phonetic, and non-linguistic factors to influence speech timing through the functioning of a single mechanism is preferred both by Occam's Razor and because it seems in agreement with the difficulties phoneticians have had in delimiting separate, non-interacting influences of such factors.

7.4 **WHAT ABOUT AN OVERALL OBSERVED PREFERENCE?**

Browman and Goldstein (1990b), on the basis of X-ray microbeam data, have observed a preference for C2 to start at 290° (*i.e.* the release) of C1 in English. This is their reason for proposing this phasing relationship for consonant sequences. Their observation can be captured here in one of two possible ways. The first possibility is that a mechanical or perceptual advantage proffered by this phasing relationship creates a strong, narrow weighting in this part of the PHASE WINDOW such that the combined preferences of other variables are generally much less strong. This would result in a high probability of the consonant sequence being realized with this phasing relationship. The second, and to my mind, more appealing, possibility is that the *typical* probability distributions for contextual variables will combine to yield a high probability peak around 290° in the PHASE WINDOW. This would result in an empirical bias for this phasing under many, but not all, conditions.

7.5 **DEVELOPMENT IN CHILDREN'S SPEECH**

Amerman and Parnell (1992) point out that the gradual acquisition of mature speech timing is indicative of "higher-level skilled motor or cognitive control" (p. 68). Recall above that we postulated that a child through learning will establish a PHASE WINDOW for a particular phasing relationship which lies within, *i.e.* permits less variability than, the physiologically possible limits on overlap. While studies on the acquisition of speech movement coordination are not numerous due to the problems of using movement tracking instrumentation with child subjects, there is some evidence that bears on this approach.

First, research suggests that a child's "loosely coordinated gestures [evolve into] the tightly coordinated patterns of articulatory movement characteristic of adult speech." and that "young speakers may exhibit great variation in (perceived) phonetic structure across attempts at the same utterance" (Nittrouer, 1993, p. 960). This gives some support to the idea that a speaker while learning to coordinate his speech movements is narrowing in on an acceptable PHASE WINDOW for particular relationships.

Second, Nittrouer (1993) notes that young speakers may exhibit greater overlap among articulatory gestures compared to more experienced speakers (Goodell and Studdert-Kennedy, 1993; Nittrouer, Studdert-Kennedy, and McGowan, 1989). Nittrouer (1993), citing Kent (1983), suggests that the initiation of gestures composing a syllable might take place sychronously for a child who has not yet established phasing relationships between the gestures. Kent goes on to suggest that "[s]ynchronous patterning may be a default principle that is overridden by phonetic and motor learning to yield the highly overlapping patterns that characterize rapid, fluent motor execution" $(p. 71)$.

The assumptions outlined above with respect to development and the PHASE WINDOW Model predict a particular sequence of events in the acquisition of speech timing. First, before PHASE WINDOWS have been created, *i.e.* before phasing relationships are implemented, children will be likely to initiate gestures (composing some prosodic unit like a syllable) synchronously. Next, after phasing relationships have been established but before they are subject to all the influences and limitations exhibited in adult timing, the coordination between gestures may be highly variable. Later, as coordination becomes more mature, systematic influences on a phasing relationship will yield more consistency in the timing of the phased gestures, although no single phasing relationship will be required.

7.6 **CONTRASTIVE VERSUS NON-CONTRASTIVE DIFFERENCES IN TIMING**

Browman and Goldstein (1991) hope to extend recent research on bimanual rhythmic movements which has demonstrated stable coordinative modes (citing Kay, Kelso, Saltzman, and Schöner, 1987; Turvey, Rosenblum, Kugler, and Schmidt, 1986). They hope that lexically contrastive patterns of gestural overlap can be understood similarly. Additionally, they suggest that critical differences in amount of overlap may yield qualitatively different acoustic, and hence perceptual, consequences. These are the mechanisms by which they hope to partition the "potential" continuum of overlap. However, they also assume an invariant timing relationship, as shown by the use of phasing rules, for the *non-contrastive* relationships as well, like that between consonants in sequence.

Above, we suggested that the PHASE WINDOW Model might be useful in capturing the timing variability observed between gestures composing different segments. The timing between gestures composing a single segment was considered to be stable and not necessarily subject to this model. These cases were excluded from discussion above. However, this approach has two drawbacks. First, it would obviously be more attractive to have a single timing mechanism at work both between and within segments. Second, the statements above confound traditional phonological/phonetic conceptions of the segment with the Articulatory Phonology approach assumed here which does not use segmental primitives but rather employs the gesture. This makes reference to segments within such a framework meaningless. The discussion below suggests that within segment and between segment timing have different characteristics but are not different in type.

It's unlikely to be coincidental that stable timing relations have been found generally for those pairs of gestures which are traditionally considered to belong to a single segment. It has been suggested within Articulatory Phonology that "syllable-structure is a characteristic pattern of coordination among gestures" (Browman and Goldstein, in press, p.13; see also Kelso, Saltzman, and Tuller, 1986). We would extend this and say that the percept and functionality of the segmental unit, to whatever extent it exists, results from *its* characteristic pattern of coordination. I suggest that this characteristic pattern of coordination is stability, *i.e.* a narrow PHASE WINDOW which is lexically specified. An independent speculation by Nittrouer et al. (1988) hypothesizes that it may be the case that intergestural overlap is more stable within segments than between segments; I infer that they mean that the presence of a segment will cause timing of its gestures to be stable. Similarly, Löfqvist (1991) offers the possibility that "gestures" forming a segment may show a greater degree of internal stability in the form of coherence of patterns of muscular activity and/or movement than those associated with different segments" (p. 346). Our suggestion is crucially different. We suggest that it is not the case that the quality-ofbeing-a-segment causes stable timing, but rather that stable timing causes the quality-of-being-asegment.

We've seen that in Articulatory Phonology, association, *i.e.* the definition of precedence relations between gestures, can be specified both within and outside the lexicon. Also implicit in Browman and Goldstein's work is that phasing of associated gestures can be specified within the lexicon and outside it. We suggest that lexically-contrastive phasing relationships are stable phasing relationships. That is, when a timing relationship is lexically distinctive, the PHASE WINDOW describing that relationship is part of the lexical representation of that word. These PHASE WINDOWS are narrow, *i.e.* yield stable or invariant relationships. By extension, these tightly cohesive gestures constitute what is traditionally considered a single segment. Other inter-gestural phasing relations are instituted outside the lexicon, as outlined in the above discussion of the PHASE WINDOW Model. Thus, we're suggesting that segmenthood might be epiphenomenal in production. In the case of contrastive phasing, we agree with Browman and Goldstein (1991) and Goldstein (1989) that quantal perceptual effects and perhaps natural oscillatory modes will determine the types of contrastive PHASE WINDOWS which may exist. (This is a separate issue from the supposition that PHASE WINDOWS exist and that they are narrow.) Specifically, we agree with Goldstein (1989, 1990) that quantal perceptual effects are important in determing contrastive timing relationships but not in actively constraining other types of inter-gestural timing.

Note that in this approach, the coordination of (associated) gestures in different "segments" (those gestures not having lexically specified, narrow PHASE WINDOWS) is exactly the same in type whether they are in the same or different words. Only the effect of prosodic *influencers* will create differences in stability and overlap.

From a language acquisition point of view, it would not be suprising if language learners lexicalize the stable timing relationships to which they are exposed; but learn as general principles of speech coordination the (inter-segmental) relationships which are influenced by a variety of factors.

This approach to the nature of "segmenthood" predicts that a doubly articulated stop like [kp] should have a very stable timing while a [kp] sequence should be more variable in timing.

(Note that Maddieson (1993) finds that the underlying [k] and [p] gestures are probably the same in both [kp] and [kp], with any movement trajectory differences being due mainly to aerodynamic conditions.) Although little articulatory data has been gathered on doublearticulated stops (but see Maddieson, 1993), particularly data addressing timing, the descriptions of these stops in the phonetic literature do seem to support the prediction of stable timing. Westermann and Ward (1933) state that "the two articulations *must* be simultaneous" (p. 58. cited in Maddieson, 1993; emphasis added). While the simultaneity (or lack thereof, cf. Maddieson and Ladefoged, 1989) isn't relevant here, the use of "must" suggests a stable timing pattern. Similar cases of relatively stable coordination within "segments" might be the coordination of labial and velar gesture in English [w] as compared to an English sequence of [kp] (although the latter differs in constriction degree from the glide); the timing of the closure and constriction gestures constituting an affricate, the timing relationships between the labial, pharyngeal, and tongue tip gestures used in certain productions of American [1] (Uldall, 1958; Delattre and Freeman, 1968; Lindau, 1985), timing of larynx raising/lowering gestures with oral gestures in ejectives/implosives, or timing of tongue body and oral gesture in clicks, etc.

This approach also suggests that if a timing relationship were to become diachronically more and more stable, it would be likely to be lexicalized by the language learner. (We should also note that in at least one case, a regular bimodal pattern of coordination (see Browman and Goldstein, in press, citing Sproat and Fujimura, 1993) has been interpreted as resulting in two allophones of a single "segment." The case in point being [1] versus [1].) This approach also predicts that inter-word timing (by definition post-lexical) will not exhibit the stability characteristic of lexically contrastive timing relationships. Lastly, it predicts that while certain between segment timing relationships may appear more stable than others (recall #sk), their timing will be affected by other *influencers*, like rate, in a way that a *within* segment timing relationship will not be, due to the difference in the width of the PHASE WINDOW which constrains the range over which *influencers* may have an effect. That is, there is a difference between a narrow PHASE WINDOW and a narrow *influencer*. Put more succinctly, this approach predicts that the relative timing of gestures constituting a segment will be less affected by rate (and other variables) than that of other gestures not constituting a segment.

Questions remain as to why there are not often lexically contrastive phasing relationships which, in addition to being very stable, have relatively little overlap (D. Steriade, p.c.). This suggests that stability might be a necessary but not sufficient condition to cause segmenthood. Again, Goldstein (1989, 1990) provides a rationale for the distinctive types of timing relationships that occur cross-linguistically. And, lastly, as in all studies of relative timing, it is difficult to experimentally determine what counts as stable or invariant. Regarding this point, we have mentioned above the predicted lack of effect of contextual influencers on stable phasings as compared to "inter-segmental" phasing relationships. This is one criterion for stability.

though there be no such thing as chance in the world, our ignorance of the real cause of any event has the same influence on the understanding and begets a like species of belief

David Hume⁴

One of the reasons for adopting a probabilistic approach to articulatory timing was to allow variability in inter-gestural timing.⁵ We began with the conjecture that all consonant clusters were not phased identically. Another means of avoiding invariant timing was the application of rules in a sequential fashion to alter timing. This was rejected for the reasons outlined in Section 7.1. However, we must realize that using probabilities to describe principles of motor timing is really describing only the observable characteristics of this process's output, not the process of timing itself. If we understood perfectly the nature of every *influencer* and could describe completely any set of *influencers* on an utterance, would the outcome still have uncertainty associated with it? Borden and Harris (1984) suggest a model of speech timing which uses internal feedback to plan speech, *i.e.* "a feed-forward predictive flow of information within the brain." Is it likely that such a speech planning mechanism of the brain works probabilistically? Probably, not. It is not even clear what this would mean. Yet, it is exactly this internal feedback or speech planning system that we ideally hope to model.

7.7.1 A FUZZY LOGIC APPROACH

probability dissipates with increasing information

B. Kosko (1990)

We saw above that the additive-centroid approach for combining various influences was appealing in that it did not yield an uncertain or probabilistic outcome; rather an actual phase angle resulted. The framework of fuzzy logic (Black, 1937; Zadeh, 1965) also is a systems control approach which might be able to capture the cognitive timing process more realistically.

 4 quoted in B. Kosko (1990)

⁵Some thinkers see no logical incompatibility between chance and determinism (e.g. Ruelle, 1991). Ruelle states: "...the state of a system at the initial time, instead of being precisely fixed, may be random...the initial state of our system may have a certain *probability distribution*. If such is the case, the system will also be random at any other time, and its randomness will be described by a new probability distribution, and the latter can be deduced deterministically by using the laws of mechanics. " (Ruelle, 1991, p. 30)

While an extended discussion of fuzzy logic won't be undertaken here, the reader is referred to Kosko (1990, 1992, 1993) on which much of the brief discussion below is based. In this section, we hope only to acquaint the reader with a framework within which the goals of the PHASE WINDOW Model outlined in this chapter may also be accomplished in a very similar fashion. Like the model proposed above, this approach has its advantages and disadvantages.

Fuzzy systems are collections of fuzzy rules, of the form if X is A, then Y is B, which all operate synchronously. The closer the input X is to the description of fuzzy set A , the greater the degree to which the output operation, the fuzzy set B, is applied to the output Y. The fuzzy system adds the outputs of all the rules and takes their centroid value (Kosko, 1993). All rules fire although some may fire to zero degree. The fuzzy system response resembles the response of a continuous system to all possible inputs (Kosko, 1993.) Neural networks can learn fuzzy rules (see Kosko, 1992).

What would fuzzy rules for speech timing be like? Here are some conceivable rules:

--If my language is not my listener's primary language, then my inter-gestural phasing should not be very overlapped.

Now, in terms of all or nothing this rule doesn't make much sense, but in fuzzy theory all things admit degree deterministically (Kosko, 1990). So, if my listener speaks my language somewhat well, I may increase my overlap to a degree corresponding to her fluency.

- --If my speech rate is very fast, then my inter-gestural phasing should be very overlapped.
- --If my speech rate is medium, then my inter-gestural phasing should be medium overlapped.

These rules are of a type typical in fuzzy control systems. Speech rate does not fall neatly into categories. A particular utterance may fall to some degree in the "very fast" fuzzy set and to some degree in the "medium fast" set. These degrees of membership are the degrees to which the two gestures are "very overlapped" and "medium overlapped" respectively.

FIGURE 7.5 SCHEMATIC SHOWING INFLUENCES ON THE CC PHASE WINDOW IN A FUZZY SYSTEM. (after Kosko, 1993)

These output sets are added and the centroid value or weighted average found to determine the output phasing. Other rules might be formulated to capture the effect of inter-gestural cohesiveness on inter-gestural phasing; or environmental noise, or distance from the listener, etc.

It is less intuitive how other, more categorical, influences should be captured in a fuzzy approach; for example, the effect of articulatory manner on timing. We saw some evidence that fricatives were less overlapped than stops by a following stop. Does the speech-planning control system operate taking into account some continuum of constriction degree? This certainly is not in accordance with linguistic understanding of the validity of categories in the mental representation of linguistic units, whether in featural or gestural theories. The use of probabilities (rather than fuzzy sets) in the PHASE WINDOW Model allows categorical and continuous influences to operate on timing commensurably, as outlined in Section 7.3.1.

There are two main points to gain from this quick look at fuzzy logic. First, a probabilistic approach to speech timing is not the only way to allow contextual variables to exert simultaneous influences on inter-gestural coordination. Second, fuzzy systems offer one alternative, deterministic approach to motor control which may be more harmonious with the physiological nature of speech planning and the multivalent nature of many variables shown to influence timing.

$7.7.2$ PROBABILISTIC EXPLANATION

We wish to close this section with a last comment on probabilistic explanation. Despite the difficulties of probabilistic explanation, it may be that because of its complexity, human behavior is still best illuminated by such an approach. Gracco suggests that "[p]erhaps speech perception and production should be appropriately represented as stochastic processes based on probability statements implemented through an adequate but imprecise control system. Strict determinism, invariance, and precision are most likely relegated to man-made machines working under rigid tolerance limits or simplified specifications, not to complex biological systems" (Gracco, 1992a, p. 20).

We concur with Diehl (1991) that phonological and phonetic knowledge might best be understood by research strategies employing a probabalistic approach. He comments that "in most cases scientists must at least provisionally settle for probabilistic forms [of explanation], because the full intricate skein of laws and relevant conditions is not completely known" (p. 129), and that "[t] the influence of any one factor may at times be obscured by the influence of other factors, which is why proposed explanations of recurrent phonological patterns must typically be probabilistic rather than deductive-nomological...[E]ven within the domain of phonetics, unique outcomes are not implied by known constraints" (p. 132). We would add also that in many cases such initial conditions cannot be known with precision in any practical way.

CHAPTER EIGHT: CONCLUSION

One of the most significant challenges in the study of speech production is to gain a theoretical understanding of how speakers coordinate articulatory movements. Many factors are known to affect articulatory timing but rarely have very many of these effects been studied concurrently using a single set of speakers and experimental techniques. This dissertation reports such an investigation, considering the timing of English consonant sequences. Electropalatography (EPG) was used to measure the tongue's contact with the palate over time. The analyses focused on dynamic patterns rather than static ones. A variety of factors related to the production of sequential consonants were also considered, such as duration and spatial magnitude. The following hypotheses were addressed: 1) The coproduction of consonantal gestures varies as a function of their articulatory places and manners. 2) The placement of syllable boundaries affects timing and displacement in consonant sequences. 3) As the number of consonants in a sequence is varied, the articulatory organization of the sequence with respect to its neighboring vowels changes. 4) Speakers adjust consonant sequence timing as a function of their speaking rate.

The results show that coda stops generally have less lingua-palatal contact than onsets. Coda fricatives are generally shorter than onsets. A tongue tip consonant is more overlapped by a following tongue body consonant than a tongue body consonant is by a following tongue tip consonant. Stop-stop sequences exhibit greater overlap than stop-fricative sequences. In examining prosodically different sequences, we find that an onset cluster is less overlapped and less variable in its timing than coda clusters and heterosyllabic sequences. In considering the coordination of long consonant sequences and neighboring vowels, our findings suggest that the near edge of the consonant sequence is relevant in timing relationships across a word boundary. Within-word relationships appear defined by the C-Center of the tautosyllabic consonants. However, each speaker and each sequence type is not necessarily consistent in what timing arrangement is most stable. Lastly, in the study of speech rate, we find that as rate increases, consonants spanning a word boundary shorten in duration and undergo a relatively linear increase in articulatory overlap. The speakers and sequences, however, do not behave identically. Rate has only a minimal effect on [d#g], which is almost completely overlapped at all rates. The consonants are also not equally susceptible to spatial reduction, although all shorten in duration. Reduction in lingua-palatal contact is the least consistent concomitant of fast speech, being found only in coda coronals.

Browman and Goldstein's Articulatory Phonology offers an innovative framework in which to consider articulatory timing. Within this framework, phasing rules synchronizing moments in two articulatory gestures have formed the basis for implementing timing. Within Articulatory Phonology, a specific phasing rule for consonant clusters has been suggested. This rule phases the onset of the second consonant in a CC cluster to co-occur with 290° in the first.

In contrast to an invariant timing approach using phasing rules, I suggest that the overlap between articulatory units is variable and may be influenced by a number of linguistic and nonlinguistic factors. If this is so, models of articulatory timing must incorporate such factors in a way which allows for timing variability. As an alternative to phasing rules, I outline a probabilistic approach to intergestural phasing--the PHASE WINDOW Model. I argue that timing relationships between articulatory gestures are constrained both physically (by human genetics) and language-specifically (by learning) to occur within permissible PHASE WINDOWS. Influencers which differ from utterance to utterance weight the PHASE WINDOW in a probabilistic manner, thereby determining where in the range of permissible overlap relationships a token is likely to be realized.

I have additionally suggested that the traditional dichotomy between segment and nonsegment be captured in terms of articulatory timing of gestural units. I argue that the percept and functionality of what has traditionally been called a segment results from a characteristic stable timing, *i.e.* a narrow PHASE WINDOW which is lexically specified. I suggest that it is not the case that segmenthood causes stable timing, but rather that stable timing causes the quality of being a segment.

The main goals of this dissertation were to determine the nature of various influences on articulatory timing and demonstrate the extent to which timing is variable, to offer a framework in which variability is allowed but constrained and, to some degree, predictable, and, lastly, to motivate discussion as to the nature of segmenthood and its relationship to gestural timing. It is hoped that the work presented in the preceding chapters makes modest progress toward these goals.

APPENDIX A: ARTIFICIAL PALATES AND DEFINED REGIONS

The following pictures show the artificial acrylic palates worn by each speaker. The pictures are to scale and are shown at 200% of their originial size. The front of the mouth is oriented to the top of the pictures. Some impression of the depth of the speakers' palatal vaults can be gained by noting the darkness in this area, with increased darkness being associated with increased depth. The electrodes are shown by small white circles. The electrodes which were excluded from both regions are shown by white X's. (See Chapter 2 for the criteria for region definition). The heavy dark horizontal line marks the division of the remaining electrodes into front and back region groups.

SPEAKER A

SPEAKER B

SPEAKER K

SPEAKER M

SPEAKER S

APPENDIX B: DESCRIPTIVE STATISTICS FOR EXPERIMENTS ONE AND TWO

SPLIT BY: SPEAKER, SEQUENCE

 $\mathcal{L}^{\text{max}}_{\text{max}}$ and $\mathcal{L}^{\text{max}}_{\text{max}}$

 ~ 10

 \sim

 $\hat{\mathcal{A}}$

 $\ddot{}$

SPLIT BY: SEQUENCE AND SPEAKER

 \bullet

 $1 -$

 \bullet

APPENDIX D: DATA FOR EXPERIMENT FOUR

SPLIT BY: SEQUENCE AND SPEAKER

 $\overline{1}$

 $\overline{}$

 $\hat{\mathcal{A}}$

33.333

 $\frac{50.000}{66.667}$ 70.000 70.000 0.000 66.667 72.727 20.000 50.000 70.000 50.000

 $\overline{53.333}$

 $\overline{57.143}$ 50.000 41.667 53.333 41.667

50.000 50.000 58.333 33.333

 62.500 0.000

45.455 33.333 33.333 54.545 40.000

 $\hat{\boldsymbol{\beta}}$

Abacus Concepts, (1989). SuperAnova. Abacus Concepts, Inc., Berkeley, CA.

- Amerman, J. D. & Parnell, M. M. (1992). Speech timing in the elderly, J. of Phonetics, 20, 65-76.
- Barry, M. (1985). A palatographic study of connected speech process. *Cambridge Papers in* Phonetics and Experimental Linguistics, 4, 1-16.
- Barry, M. (1991). Temporal modelling of gestures in articulatory assimilation. *Proceedings of* the XIIth International Congress of Phonetic Sciences, vol. 4, 14-17.
- Barry, M. (1992). Palatalisation, assimilation and gestural weakening in connected speech. Speech Communication, 11, 393-400.
- Barry, W. J. (1983). Some problems of interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, $\dot{9}$, 826-828.
- Bell-Berti, F. & Harris, K. S. (1981). A temporal model of speech production. *Phonetica*, 38, 9-20.
- Benoit, C. (1986). Note on the use of correlations in speech timing. *Journal of the Acoustical* Society of America, 80, 1846-1849.
- Black, M. (1937). Vagueness: An excercise in logical analysis. *Philosophy of Science*, 4, 427-455.
- Borden, G. J. & Harris, K. S. (1984). Speech Science Primer. 2nd ed. Baltimore: Williams & Wilkins.
- Browman, C. & Goldstein, L. (1986). Toward an articulatory phonology. Phonology Yearbook, 3, 219-252.
- Browman, C. & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. Phonetica 45:140-155.
- Browman, C. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, $201 - 252$.
- Browman, C. & Goldstein, L. (1990a). Gestural specification using dynamically-defined articulatory structures. J. Phonetics, 18, 299-320.
- Browman, C. & Goldstein, L. (1990b). Tiers in articulatory phonology, with some implications for casual speech. in J. Kingston and M. E. Beckman (eds.) Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech. Cambridge:Cambridge University Press, 341-376.
- Browman, C. & Goldstein, L. (1991). Gestural structures: distinctiveness, phonological processes, and historical change. in I. Mattingly and M. Studdert-Kennedy (eds.) Modularity and the Motor Theory of Speech Perception, New Jersey: Lawrence Erlbaum Associates, 313-338.
- Browman, C. & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49, 155-180.
- Browman, C. & Goldstein, L. (in press). Dynamics and articulatory phonology. Mind as Motion, (MIT press).
- Butcher, A. (1989). Measuring coarticulation and variability in tongue contact patterns. Clinical Linguistics and Phonetics, 3, 1, 39-47.
- Byrd, D. (1992). Perception of assimilation in consonant clusters: a gestural model. *Phonetica*, $49, 1-24.$
- Catford, J. C. (1977). Fundamental problems in phonetics. Bloomington:Indiana University Press.
- Choi, J. D. (1992). Phonetic Underspecification and Target Interpolation: An Acoustic Study of Marshallese Vowel Allophony, Ph.D. dissertation, UCLA, UCLA Working Papers in Phonetics, 82.
- Cohn, A. (1990). Phonetic and Phonological Rules of Nasalization. Ph.D. dissertation, UCLA, UCLA Working Papers in Phonetics, 76.
- Coleman, J. (1992). The phonetic interpretation of headed phonological structures containing overlapping constituents. *Phonology*, 9, 1-44.
- Collier, R., Lisker, L., Hirose, H., and Ushijima, T. (1979). Voicing in intervocalic stops and fricatives in Dutch. Journal of Phonetics. 7, 357-373.
- Daniloff, R. G., & Hammarberg, R. E. (1973). On defining co-articulation. J. Phonetics, 1, 239-248.
- Davis, S. (1990). An argument for the underspecification of [Coronal] in English. *Ling. Ing.* 21.2, 301-305.
- DeJong, K. (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica*, 48, 1-17.
- Delattre, P. and D. Freeman (1968). A dialect study of American R's by x-ray motion picture. Linguistics 44: 29-68.
- Diehl, R. L. (1991). The role of phonetics within the study of language. *Phonetica*, 48, 120-134.
- Dixon, W. J. (1988). BMPD Statistical Software Manual. The Regents of the University of California.
- Flavell, J. H. (1963). The Developmental Psychology of Jean Piaget. Princeton, NJ:Van Nostrand.
- Flege, J. E. (1976). Plasticity in adult and child speech production. Journal of the Acoustical Society of America, 79, S54.
- Flemming, E., Ladefoged, P., & Thomason, S. (1993). The phonetic structures of Montana Salish. to appear in UCLA Working Papers in Phonetics.
- Fletcher, S. G. (1992). Articulation: A Physiological Approach. San Diego: Singular Publishing Group, Inc.
- Fletcher, S. G., McCutcheon, M. J., Wolf, M. B., Sooudi, I., & Smith, T. L. (1975). Palatometricgnathometric study of speech articulation. Final Report. (NIH Grant No. NS10540).
- Foldvik, A.K., Husby, O., Kværness, J., Nordli, I.C. & Rinck, P.A. MRI (Magnetic Resonance Imaging) for filming articulatory movements. Proceedings of the XIIth International Congress of Phonetic Sciences, vol. 2, 34-36, 1991.
- Fowler, C. A. (1977). Timing Control in Speech Production. Bloomington, Indiana:Indiana University Department of Linguistics.
- Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. J. Phonetics, 8, 113-133.
- Fromkin, V. (1965). Some phonetic specifications of linguistic units: An electromyographic investigation. Ph.D. diss, UCLA, UCLA Working Papers in Phonetics, 3.
- Fujimura, O. (1986). Relative invariance of articulatory movements: An iceberg model.. in Perkell, J. & Klatt, D. (eds.) Invariance and Variability in Speech. Hillsdale, HIllsdale, NJ:Erlbaum, 226-234.
- Fujimura, O., & Lovins, J. B. (1978). Syllables as concatenative phonetic units. In A. Bell & J. B. Hooper (eds.) Syllables and Segments (pp. 107-120). New York: Elsevier North-Holland.
- Fukui, N. and Hirose, H. (1983). Laryngeal adjustments in Danish voiceless obstruent production. Annual Bulletin Research Institute of Logopedics and Phoniatrics, 17, 61-71.
- Gay, T. (1968). Effect of speaking rate on diphthong formant movements. J. Acoust. Soc. Amer., 44(6), 1570-1575.
- Gay, T. (1974). A cinefluorographic study of vowel production. J. Phonet., 2, 255-266.
- Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, 38, 148-158.
- Goldstein, L. (1989). On the domain of the quantal theory. J. Phonet., 17, 91-97.
- Goldstein, L. (1990). On articulatory binding: comments on Kingston's paper. Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. J. Kingston and M. Beckman (eds.) 445-450.
- Goodell, E. W., & Studdert-Kennedy, M. (1993). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study. *Journal of* Speech and Heaaring Research, 36.
- Gracco, V. L. (1988). Timing factors in the coordination of speech movements. Journal of Neuroscience, 8, 4628-4634.
- Gracco, V. L. (1992a). Characteristics of speech as a motor control system. Haskins Laboratories Status Report on Speech Research. SR-109/110, 13-26. to appear in G. Hammond (ed.) Cerebral Control of Speech and Limb Movements. Amsterdam: Elsevier.
- Gracco, V. L. (1992b). Sensorimotor mechanisms in speech motor control. Haskins Laboratories Status Report on Speech Research. SR-109/110, 27-44. to appear in H. Peters, W. Hulstijn, & C. W. Starkweather (eds.) Speech Motor Control and Stuttering.
- Gracco, V. L., and Abbs, J. H. (1989). Sensorimotor characteristics of speech motor sequences. Experimental Brain Research, 75, 586-598.
- Hamlet, S. L., and Stone, M. (1978). Compensatory alveolar consonant production induced by wearing a dental prosthesis. Journal of Phonetics, 6, 227-2481
- Hardcastle, W. J. (1972). The use of electropalatography in phonetic research. *Phonetica* 25: $197 - 215.$
- Hardcastle W. J. (1984). New methods of profiling lingual-palatal contact patterns with electropalatography. Phon. Lab. Univ. Reading Work in Prog., 4, 1-40.
- Hardcastle, W. J., Gibbon, F., & Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. Journal of Phonetics 19,3, 251-266
- Hardcastle, W. J. & P. Roach. (1979). An instrumental investigation of coarticulation in stop consonant sequences. in Current Issues in the Phonetic Sciences. H. Hollien & P. Hollien (eds.) Amsterdam: John Benjamins, pp. 531-540.
- Hayes, B. (1989). Compensatory legnthening in moraic phonology. *Linguistic Inquiry 20, 253*-306.
- Hayes, B. (1992). Comments on the paper by Nolan. Papers in Laboratory Phonology II: Gesture, Segment, Prosody. G. J. Dougherty & D. R. Ladd (eds.) Cambridge: Cambridge University Press, 280-285
- Hebb, D. O. (1949). The Organization of Behavior. New York: Wiley.
- Hirose, H. and Gay, T. (1972). The activity of the intrinsic laryngeal muscles in voicing control. *Phonetica* 25, 140-164.
- Hirose, H., Lisker, L. and Abramson, A. (1972). Physiological aspects of certain laryngeal features in speech production. Haskins Laboratories Status Report on Speech Research. $SR - 31/32$, $183 - 191$.
- Huffman, M. K. (1990). Implementation of nasal: Timing and articulatory landmarks. Ph. D. dissertation, UCLA, UCLA Working Papers in Phonetics, 75.
- Hume, D. (1748). An Inquiry Concerning Human Understanding.
- Johnson, K., Ladefoged, P. & Lindau, M. (1993). Individual differences in vowel production. J. Acoust. Soc. Am. 94,2, pt. 1, 701-714.
- Jones, D. (1956). An Outline of English Phonetics, Eighth Edition (Eighth Edition). Cambridge: W. Heffer & Sons, Ltd.
- Kay, B. A., Kelso, J. A. S., Saltzman, E. L., & Schöner, G. (1987). Space-time behavior of single and bimanual rhythmical movements: Data and limit cycle model. Journal of Experimental Psychology: Human Perception and Performance, 13, 178-192.
- Keating, P. A. (1990a). Phonetic representations in a generative grammar. J. of Phonetics, 18, $321 - 334.$
- Keating, P. A. (1990b). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. E. Beckman (eds.) Papers in Laboratory Phonology I. Cambridge:Cambridge University Press, 451-470.
- Keller, E. (1987). The variation of absolute and relative measures of speech activity. *Journal of* Speech and Hearing Research, 30, 223-229.
- Keller, E. (1990). Speech motor timing. in W. J. Hardcastle & A. Marchal (eds.) Speech Production and Speech Modelling., 342-364.
- Kelso, J. A. S., Saltzman, E., & Tuller, B. (1986). The dynamical perspective on speech production: data and theory. Journal of Phonetics, 14, 29-59.
- Kelso, J. A. S. & Tuller, B., (1987). Intrinsic time in speech production: Theory, methodology, and preliminary observations. in E. Keller & M. Gopnik (eds.) Sensory and Motor Processes in Language. Hillsdale, NJ:Erlbaum, 203-222.
- Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.
- Kent, R. D. (1970). A Cinefluorographic-Spectrographic Investigation of the Component Gestures in Lingual Articulation. Unpublished Ph. D. Dissertation, University of Iowa.
- Kent, R. D. (1983). The segmental organization of speech. in P. F. MacNeilage (ed.) The Production of Speech. New York: Springer-Verlag. 57-89.
- Kent, R. D. (1986). The iceberg hypothesis: The temporal assembly of speech movements. in Perkell, J. & Klatt, D. (eds.) *Invariance and Variability in Speech*. Hillsdale, NJ:Erlbaum, 234-242.
- Kent, R. D. & Bauer (1986). Toward an acoustic typology of vocalizations in the first year of life. Journal of the Acoustical Society of America, 79 , $S(1)$, $S53$ A.
- Kent, R. D. & Burkard, R. (1981). Changes in acoustic correlates of speech production. In D. S. Beasley & G. A. Davis (eds.) Aging: Communication Processes and Disorders, New York:Grune and Stratton, 47-62.
- Kent, R. D., and Moll, K. L. (1975). Articulatory timing in selected consonant sequences. Brain & Language, 2, 304-323.
- Kent, R. D., and Rosenbek, J. C. (1983). Acoustic patterns of apraxia of speech. J. Speech and Heariing Research, 26, 231-249.
- Kingston, J. (1990). Articulatory binding. In J. Kingston & M. E. Beckman (eds.), Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech. Cambridge: Cambridge University Press.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 59, 1208-1221.
- Kosko, B. (1990). Fuzziness vs. probability. Int. J. General Systems, 17, 211-240.
- Kosko, B. (1992). Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to *Machine Intelligence.* Englewood Cliffs, NJ: Prentice Hall.
- Kosko, B. (1993). *Fuzzy Thinking*. New York: Hyperion.
- Kozhevnikov, V., and Chistovich, L. (1965). Speech: Articulation and Perception (Translated and distributed by Joint Publications Research Service, Washington D. C., N. JPRS30, 543, Moscow-Leningrad. Originally published in Russian, 1965., Trans.). Washington DC: Joint Publications Research Service No. 30.
- Krakow, R. A. (1989). The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velar Gestures. Ph.D. dissertation, Yale University.
- Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *J. Phonet.*, 4, 303-320.
- Ladefoged, P. (1957). Use of palatography. J. Speech Hear. Dis., 22, 764-774.
- Ladefoged, P., DeClerk, J., Lindau, M., & Papçun, G. (1972). An auditory-motor theory of speech production. UCLA WPP, 22, 48-75.
- Lamontagne, G. A. (1993). Syllabification and Consonant Cooccurrence Conditions. Ph. D. dissertation, University of Massachusetts.
- Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, 5, (Suppl.):1-54.
- Levene, H. (1960). Robust tests for equality of variance. In I. Olkin, (ed.) Contributions to Probability and Statistics. Palo Alto: Stanford University Press.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- Lindau, M. (1985). The story of /r/. Phonetic Linguistics: Essays in Honor of Peter Ladefoged. Orlando, FL, Academic Press, 157-168.
- Lindau, M., Harshman, R., and Ladefoged, P. (1971). Factor analyses of formant frequencies of vowels. J. Acoust. Soc. Amer., 50, 117 (Abstract).
- Lindblom, B. (1963). Spectrographic study of vowel reduction. J. Acoust. Soc. Amer., 35(11), 1773-1781.
- Lindblom, B. (1964). Articulatory activity in vowels. STL-OPSR, 2, 1-5.
- Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (ed.) The Production of Speech (pp. 217-245). New York: Springer-Verlag.
- Lindblom, B., Lubker, J., Gay, T., Lyberg, P., Branderud, P. & Holmgren, K. (1987). The concept of target and speech timing. In R. Channon & L. Shockey (eds.) In Honor of Ilse Lehiste. Dordrecht, The Netherlands: Foris Publications, 161-181.
- Löfqvist, A. (1991). Proportional timing in speech motor control. *Journal of Phonetics*, 19, 343-350.
- Löfqvist, A. and Yoshioka, H. (1980a). Laryngeal activity in Swedish obstruent clusters. Journal of the Acoustical Society of America, 68, 792-801
- Löfqvist, A. and Yoshioka, H. (1980b). Laryngeal-oral coordination in American English obstruent production. Journal of the Acoustical Society of America, 68, S102(A).
- Löfqvist, A. and H. Yoshioka (1981a). Interarticulator programming in obstruent production. Phonetica 38: 21-34.
- Löfqvist, A. and Yoshioka, H. (1981b). Laryngeal activity in Icelandic obstruent production. Nordic Journal of Linguistics, 4, 1-18. (Also published in Haskins Laboratories: Status Report on Speech Research SR-63/64, 1980, 275-292).
- Löfqvist, A. and Yoshioka, H. (1981c). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. Speech Commun. 3, 279-289.
- Löfqvist, A. and Yoshioka, H. (1984). Interarticulator programming in obstruent production. Phonetica, 38, 21-34.
- Löfstedt, I. (1992). Swedish segment length and structure preservation. Studia Linguistica 46(2), 93-127.
- Lubker, J. (1986). Articulatory timing and the concept of phaase. Journal of Phonetics, 14, 133- $137₁$
- MacNeilage, P. (1970). Motor control of serial ordering of speech. *Psych. Rev.*, 77, 182-196.
- Maddieson, I. (1993). Investigating Ewe articulations with electromagnetic articulographyt. to appear in Forschungsberichte, Institut für Phonetik und Sprachliche Kommunikation, Universität München. (ACCOR Workshop on Electromagnetic Articulography.)
- Maddieson, I. & Ladefoged, P. (1989). Multiply-articulated segments and the feature hierarchy. Paper presented at the Annual Meeting of the LSA, Washington, D.C. Printed in UCLA Working Papers in Phonetics, 72, 116-138.
- Manuel, S. (1987). Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages. Ph. D. dissertation, Yale University.
- Marchal, A. (1988). Coproduction: Evidence from EPG data. Speech Communication 7: 287-295, 1988.
- Mattingly, I. G. (1980). Phonetic representation and speech synthesis by rule. *Haskins* Laboratories Status Report on Speech Research SR-61, 15-21.
- McCarthy, J. & Prince, A. (1986). *Prosodic Morphology*, ms., Deptartment of Linguistics, University of Massachusetts, Amherst, and Program in Linguistics, Brandeis University, Waltham, Massachusetts.
- Mermelstein, P. (1973). Articulatory model for the study of speech production. *JASA*, 53(4), 1070-1082.
- Munhall, K. G. (1985). An examination of intra-articulatory relastive timing. *Journal of the* Acoustical Society of America, 78, 1548-1553.
- Munhall, K. G., Fowler, C., Hawkins, S., & Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. J. of Phonetics, 20, 2, 225-240.
- Munhall, K. G., & Löfqvist, A. (1992). Gestural aggregation in speech: Laryngeal gestures. J. of *Phonetics*, 20, 93-110.
- Munhall, K. G., Löfqvist, A., & Kelso, J. A. S. (1986) Laryngeal compensation following sudden oral perturbation. J. Acoust. Soc. Am., 80(Suppl. 1), S109.
- Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. Journal of Experimental Psychology, 4, 457-474.
- Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: evidence from an acoustic study. Journal of Speech and Hearing Research, 36, 959-972.
- Nittrouer, S., Munhall, K. G., Kelso, J. A. S., Tuller, B., & Harris, K. S. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. J. Acoust. Soc. Am., $84(5)$, 1653-1661.
- Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. Journal of Speech and Hearing Research, 32, 120-132.
- 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.* Cambridge:Cambridge University Press, 261-289.
- Ohala, J. (1990). The generality of articulatory binding. Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. J. Kingston and M. Beckman (eds.) 435-444.
- Ohala, J., Hiki, S., Hubler, S., and Harshman, R. (1968). Photoelectric methods of transducing lip and jaw movements in speech. UCLA Working Papers in Phonetics, 10.
- Öhman, S. E. (1967). Numerical model of coarticulation. J. Acoust. Soc. Am., 41, 310-320.
- Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movements in speech. J. Exp. Psychol. Human Percept. Perform., 9, 622-636.
- Ostry, D. J., and Munhall, K. G. (1985). Control of rate and duration of speech movements. J. Acoust. Soc. Am., 77(2), 640-648.
- Petursson, M. (1977). Timing of glottal events in the production of aspiration after [s]. *Journal* of Phonetics, $5, 205-212$.
- Pierrehumbert, J. (in press). Syllable structure and word structure: A study of triconsonantal clusters in English. in P. Keating (ed.) Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III. Cambridge: Cambridge University Press, 168-187.
- Prator, C. H., & Robinett, B. W. (1985). Manual of American English Pronunciation, New York: Holt, Rinehart and Winston.
- Prince, A. (1980). A metrical theory for Estonian quantity. *LI*, 3, 511-562.
- Ramig, L. (1983). Effects of physiological aging on speaking and reading rates. Journal of Communication Disorders, 16, 217-226.
- Recasens, D., Fontdevila, J., & Pallarès, M. D. (1993). An electropalatographic study of stop consonant clusters. Speech Communications, 12, 335-355.
- Ringel, R. L., Chodzko-Zajko, & Offenbach, S. I. (1986). Age, health and the speech process, ASHA Mini-seminar, Detroit, Michigan.
- Ruelle, D. (1991). Chance and Chaos. New Jersey: Princeton University Press.
- Saltzman, E. & Munhall., K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1:333-382.
- Selkirk, E. O. (1982). The syllable. In H. Van der Hulst & N. Smith (eds.) Structure of *Phonological Representations, Part 2 Foris Publications.*
- Sock, R. & Jah, O. (1986). Old wine in new bottles, new wine in old bottles: Action theory and (para)metric phonology views in the domain of Wolof. Bull. Inst. Phon. Grenoble, 15, $117 - 133.$
- Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. J. of Phonetics, 21, 291-312.
- Steriade, D. (1982). Greek Prosodies and the Nature of Syllabification. Ph. D. Dissertation, MIT.
- Steriade, D. (1993). Closure, release and nasal contours, in M. Huffman and R. Krakow (eds.) Phonetics and Phonology, Volume 5, Nasals, Masalization, and the Velum. Academic Press, 401-470.
- Stevens, K. N. and House, A. S. (1963). Perturbation of vowel articulation by consonantal context: an acoustical study. Journal of Speech and Hearing Research, 6, 111-128.
- Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. Language, 65, 81-106.
- Stone, M., Faber, A., & Cordaro. M. (1991). Cross-sectional tongue movement and tongue-palate movement in [s] and [f] syllables. in *Proceedings of the XIIth International Congress of* Phonetic Sciences. Université de Provence, 354-357.
- Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. J. Speech Hear. $Res.$, $16.397-420.$
- Tuller, B., & Kelso, J. A. S. (1984). The timing of articulatory gestures: Evidence for relational invariants. J Acoust Soc Amer, 76(4), 1030-1036.
- Tuller, B., Kelso, J. A. S., and Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. J Experimental Psychol: Human Perception & Performance, 8, 460-472.
- Tung, T. (1964). A Descriptive Study of the Tsou language, Formosa. Taipei: Institute of History and Philology, Academia Sinica Special Publications, No. 48.
- Turvev. M. T. (1990). Coordination. American Psychologist, August, 938-953.
- Turvey, M. T., Rosenblum, L. D., Kugler, P. N., & Schmidt, R. C. (1986). Fluctuations and phase symmetry in coordinated rhythmic movements. Journal of Experimental Psychology: Human Perception and Performance, 12, 564-583.
- Uldall, E. T. (1958). American 'molar' r and 'flapped' r. Revista do Laboratorio de Fonetica Experimental (Coimbra) 4, 103-106.
- Vatikiotis-Bateson, E. & Kelso, J. A. S. (1993) Rhythm type and articulatory dynamics in English, French and Japanese. Journal of Phonetics, 21, 3.
- Welford, A. T. (1977). Causes of slowing performance with age. Interdisciplinary Topics in Gerontology, 11, 43-51.
- Westermann, D. & Ward, I. C. (1933). Practical Phonetics for Students of African Languages. Oxford University Press for International African, Institute.
- Winer, B. J. (1971). Statistical Principles in Experimental Design. New York:McGraw-Hill Book Company.
- Winitz, H., Scheib, M. E., & Reeds, J. A. (1972). Identification of stops and vowels. *J Acoust* Soc Amer, 51(4), 1309-1317.
- Yoshioka, H., Löfqvist, A., & Hirose, H. (1981). Laryngeal adjustments in the production of consonant clusters and geminates in American English. J Acoust Soc Amer, 70, 1615-1623.
- Yoshioka, H., Löfqvist, A., & Hirose, H. (1982). Laryngeal adjustments in Japanese voiceless sound production. *J. of Phonetics*, 10, 1, 1-10.
- Zadeh, L. A. (1965). Fuzzy Sets. Information and Control. 8, 338-353.
- Zec, D. (1989). Sonority Constraints on Prosodic Structure. Ph.D. Dissertation, Stanford University.