An Articulatory Basis for the Syllable

by

Melissa Annette Redford, B.A., M.A.

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 1999

An Articulatory Basis for the Syllable

Approved by
Dissertation Committee:

Randy L. Diehl

Acknowledgments

I am grateful for the rich, rigorous, and supportive intellectual environment that my committee members have provided me over the years leading up to the completion of this dissertation. Leslie Cohen, Randy Diehl, Catharine Echols, Bjorn Lindblom, Risto Miikkulainen, and, my advisor, Peter MacNeilage have all contributed to my development as a thinking person and academic.

While I was still a student in the Department of Linguistics, I had the opportunity to begin working with Leslie Cohen and Randy Diehl on different projects. I was welcomed into both their labs and, under their guidance, I was able to learn first-hand how to design, implement, and report experiments. Les also taught me the basics of the field of infant cognition and together we briefly forayed into animal cognition. Les also provided me with my first experience of academic life as he constantly encouraged me to join in all the activities (both academic and not) undertaken by the extended "family" in his lab and area. Eventually, however, I concentrated more and more on phonetic questions and so moved from part-time residence at the Children's Research Lab to full-time residence in Randy's Speech Perception lab. Randy never asked me to leave even after he and I finished our project together. For years, he has generously provided me with a home and resources with which to conduct experiments unrelated to his own work. All of the work presented in this dissertation was realized because of his generosity.

I met Catharine Echols when she gave a guest lecture in Risto's Cognitive Science class when I was still a student in Linguistics. I never had the opportunity to work directly with Cathy on a particular project, but have benefited from discussing with her all of the projects that I have ever undertaken in graduate school. Cathy was always willing (and incredibly able) to think deeply about any one of the diverse topics I had launched into and she provided me with many far-reaching comments about the content or structure of my arguments. In addition, Cathy has always found time to offer me personal support whenever I have needed it.

Although I met Risto Miikkulainen the same semester that I met Cathy, it was not until I went to him with an idea for a project that I really got to know him. He welcomed me with his characteristic energy and enthusiasm for new ideas. Risto taught me the meaning of brainstorming and I have had a lot of fun working with him. In addition, I have benefited tremendously from his practical and emphatic approach to work issues ranging from conference going to dissertation writing.

My development as a phonetician and thinker has been most influenced by Bjorn Lindblom and Peter MacNeilage. I met Peter at the reception for new graduate students in Linguistics. He directed me to enroll immediately in Bjorn's Introductory Phonetics course since Bjorn was scheduled to return to Sweden for an indefinite period of time the following semester. The material covered in that course captivated me and the ideas that Bjorn presented, his ideas, shaped my thinking about phonetics. Bjorn returned to Texas a couple of years ago on a part-time basis and I have continued to learn from him every time that I meet with him. His influence on my thinking can be seen throughout the pages of this dissertation as many of the specific ideas that I develop emerge from his previous work.

This brings me finally Peter MacNeilage, my advisor and mentor, to whom I owe an enormous debt of gratitude for shaping my intellectual development over the past years and for supporting my emotional development. Peter has tried to teach me to look at the details of many things and understand them as one big thing. He has provided me with a singular way to ask questions, evaluate evidence, and organize an argument. Peter has also provided me with an optimal learning environment, where intellectual support was available without any intellectual restraints. In addition to these big things, Peter has over and over contributed to my development as a thinking person and academic in smaller ways. For instance, we once spent 6 solid hours going over my master's thesis page by page. He illuminated, in his inimitable way, all the major and minor flaws of the paper (and there were many) without once ever making me feel inadequate. Peter has subtly supported, shaped, and encouraged all my endeavors over the past six years. These endeavors have culminated in this dissertation, which so clearly bears his mark in both content and tone.

In addition to being grateful to the members of my committee, I am grateful to my family and all my friends for always supporting in my work. My parents have always been extremely encouraging of my endeavors. They are always duly impressed and very proud, even when they appear somewhat mystified by what exactly I am doing. Of my friends, I would most especially like to thank my best friend and husband, Sergei Bogdanov, for being willing to intelligently discuss at interminable lengths any and all issues that have arisen in the process of my conceiving of and writing this dissertation.

MELISSA ANNETTE REDFORD

The University of Texas at Austin
December 1999

An Articulatory Basis for the Syllable

Publication No. ____________________

Melissa Annette Redford, Ph.D.
The University of Texas at Austin, 1999

Supervisor: Peter F. MacNeilage

The syllable has been difficult to define in phonetics and so it has often been assumed to be without uniform or direct phonetic correlates. The lack of a concrete definition for the syllable has encouraged phonologists and phoneticians to treat the syllable as an innate, higher-order mental unit within which segments are organized, rather than as an actual unit of speech derived from experience with production and perception. It is argued that previous attempts to define the syllable in phonetic terms might have failed either because the definition rested on a sequential analysis of speech production or because the definition was not linked to syllable perception. In addition, previous analyses have ignored the cross-language patterns in syllable structure identified by phonologists, which a phonetic account of the syllable should also explain.

In contrast to previous attempts, the present attempt to provide a phonetic basis for the syllable focuses on the relationship between the hierarchical structure of the supraglottal vocal tract and the acoustic patterns of speech that are associated with the perception of syllables. Specifically, it is hypothesized that the regular, fixed, open-close movement of the jaw provides a mechanical and temporal constraint on the action of the more versatile segmental articulators. This constraint is predicted to yield the phonological pattern of preferred segment sequences as well as the phonetic pattern of different relative segment durations. In addition, it is argued that inherent properties of the cycle, such as the asymmetries in duration, displacement, and velocity of the opening and closing phases, provide an articulatory basis for certain cross-language preferences in syllable structure. Measures of jaw movement were used either to test against alternative phonological/phonetic

hypotheses or to predict acoustic and perceptual data according to the hypotheses. The results of these tests supported the idea that the constraint of the jaw cycle provides an articulatory basis for the syllable.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Here are the verses that describe the waters of the South Pole in the "Rime of the Ancient Mariner" by S.L. Coleridge:

And now there came both mist and snow,
And it grew wondrous cold:
And ice, mast-high, came floating by,
As green as emerald.

And through the drifts the snowy clifts
Did send a dismal sheen:
Nor shapes of men nor beasts we ken –
The ice was all between.

The ice was here, the ice was there,
The ice was all around:
It cracked and growled, and roared and howled,
Like noises in a swound!

The rhythms of this epic poem jump out at the reader. These rhythms are not due to the individual sounds or the words, though these create the rhyme. Instead the rhythm is created with a regular alternation in the number of syllables per line. In these three verses, the regular alternation is eight syllables, then six syllables. Of course, the reader is not instructed as to the number of syllables per line, but by the time one arrives at the line *As green as emerald*, it is difficult not to draw out the three syllables of the final word and pronounce: *As green as em-mer-rald.*

The verses presented above are meant to demonstrate that the syllable is a psychologically real unit for language speakers. This is evident not only in the rhythms of poetry and song, but also in the word games, writing systems and morphological processes of various languages. Unlike the phoneme, the idea of the syllable does not require familiarity with a particular writing system, or any writing system at all (Ladefoged, 1993). For instance, Derwing (1992) showed that literate and illiterate speakers of a diverse set of languages with a diverse set of writing systems were remarkably consistent in their divisions (or syllabifications) of words with single intervocalic consonants or consonant sequences. Notwithstanding Derwing's results, linguists often find that, even though syllables are intuitive to all language speakers, intuitions regarding the exact location of syllable boundaries often vary from speaker to speaker. This variation in native speaker intuitions has created a major problem for a linguistic theory of the syllable. The problem is exacerbated by the perceived lack of direct phonetic correlates to the syllable.

Perhaps because the syllable is perceived to lack direct phonetic correlates, the syllable is assumed by phonologists and many phoneticians to exist as a mental representation for segment organization (see, for instance, Blevins, 1995; Ladefoged, 1993). This does not mean that syllables are not realized in speech. Phonological patterns of segment sequencing and phonetic patterns of relative segment duration are both associated with the syllable, but insofar as "phonological representations provide input to the phonetic interpretive component (Blevins, 1993:232-233)," syllables are thought to exist first and foremost in the language speaker's mind. This view of the syllable allows phonologists to characterize various regular sound processes within a single language, such as the pattern of stress assignment in English. The view fails, however, to explain how the syllable comes to exist in a speaker's mind and/or how it emerges as a unit of sound organization in language.

In order to explain how syllables emerge as units of sound organization in language, it is necessary to appeal to an approach that views language in diachronic terms. One such approach assumes that language has evolved as an optimal communication system. Optimal here means that the system has as low costs and high benefits as possible, i.e., the signaler expends the minimum amount of energy in producing a signal and the signal produced is perceptually distinctive to the receiver. In this framework, sound patterns in language are understood to emerge in response to production and perception constraints imposed by language users on the system. Since the emphasis of this approach is on the functional nature of sound patterns, it assumes that sound patterns exist and are accessible in the physical stimulus. This assumption entails that concepts, such as the syllable, come to exist in the mind of language speakers through experience with language.

The present dissertation adopts this explanatory approach in order to understand how syllables emerge as units of sound organization in language. Since the syllable is currently thought to exist only as a mental representation, the present attempt will focus on establishing a possible phonetic basis for the syllable. Specifically, the focus will be on establishing an articulatory basis for the syllable. Articulation is chosen over perception because it is assumed that the syllable is perceived from information available in the speech stream. I will therefore try to answer the question of how that information gets into speech.

In the following chapter, the problems associated with defining a syllable are enumerated. In addition to the previously mentioned problem of variable syllable boundary judgments, there is the problem of explaining segment sequencing patterns in language, phonetic patterns associated with syllables, and syllable structure, which varies significantly from language to language. It is argued that any account of the syllable must attempt to solve each of these problems. The proposed solution to these problems is also presented in Chapter 2. The solution posits that the syllable, a superordinate unit in the hierarchy of speech sounds, emerges from the hierarchical structure of the supraglottal vocal tract. Specifically, the regular, open-close motion of the jaw is hypothesized to constrain segmental articulation in such a way as to yield the phonological and phonetic patterns that form the basis of syllable perception. More specific hypotheses emerge from this general idea and these hypotheses are further developed, tested and discussed in Chapters 3, 4, and 5. The cumulative evidence from these latter chapters is discussed in Chapter 6 and presented as support for the general idea that the jaw cycle is the defining articulatory factor in syllable production. By way of conclusion, Chapter 6 also considers some of the implications of the hypotheses, limitations of the evidence presented, and future work to extend and solidify the basic ideas proposed in this dissertation.

Chapter 2

The Problem of Syllable Definition

2.1 The problem with syllables

In discussing sound sequences in speech Ferdinand Saussure noted that, "The ear perceives syllabic division in every spoken chain; it also perceives a sonant in every syllable. One can accept both facts and still wonder why they should hold true (1959:58)." The two facts are related because the then, and in some ways still current, theories of syllabic division were based on a perceived division between the sonants and those sounds that combined with the sonants – the consonants. In these theories, the syllable was an assembly of the two sounds types. Sonants were flanked by consonants in the speech stream and the least sonorant consonants divided the sonants and their flanking consonants from one another. While Saussure criticized these theories because they were largely descriptive rather than explanatory, his own proposal, based on combinations of "implosions" and "explosions" in the speech chain, did not resolve the question. As a result, Chomsky and Halle (1968), for simplicity's sake, ignored the question completely in their famous phonological analysis, The Sound Patterns of English. More recently, however, those who study sound patterns have found that syllables and, by extension, syllable boundaries are impossible to ignore.

In an introductory text on phonology, Kenstowitz (1994) expresses the prevailing sentiment that the concept of the syllable is motivated in phonology by three factors: (1) the existence of segment sequencing constraints in language; (2) the fact that phonological rules, such as those for linguistic stress assignment, are simplified by the concept of syllable; and (3) that certain phonological operations, such as the insertion of an epenthetic vowel, are best understood with reference to syllable structure. These statements complement the stated belief of the author that the syllable is an abstract and conceptual unit with no "uniform or direct phonetic correlates (1994:250)." By and large, this belief is supported in the phonetic literature where the syllable is not associated with any specific phonetic event, but rather with patterns of differences in acoustic duration, amplitude, and frequency (e.g., Lehiste, 1970; Price, 1980; Ainsworth, 1986). The lack of a one-to-one correspondence between the acoustics and the perception of syllables has led many phoneticians to assume that the syllable represents one unit of neural organization in speech programming (e.g., Fry, 1964; Kozhevnikov and Chistovich, 1965; Fromkin, 1968; Lehiste, 1977). Thus, phoneticians, like phonologists, have usually treated the syllable as a higher-order, mental unit within which segments are organized.

Phonetic analysis of syllabic attributes such as quantity, tone, and stress can continue without a concrete definition of the syllable, even though this may not be preferred. Phoneticians can measure differences in the duration and frequency of segments that form unambiguous syllables in the perception of a native speaker (i.e., the phonetician). In contrast, phonological analyses of these attributes require that any speech string be divided into syllables before syllable-referencing phonological rules may apply. Accordingly, phoneticians have mostly left the task of defining syllables to phonologists.

2.2 The phonological solution

Phonologists have addressed the problem of syllable definition by referring to the language data. General trends are induced from syllable systems across languages. These trends are formalized as rules that guide syllable boundary assignment or syllabification. Syllabification rules are generalized to new data and used to define syllable units therein. Thus, in phonology, an understanding of the syllable requires an understanding of the language data upon which syllabification rules are derived. These language data have been exhaustively summarized by Bell and Hooper (1978:8-11) in the following fifteen separate statements:

1. Within the section[1], VV sequences ("hiatus") are not permitted by about one-half of the world's languages, e.g., Berber.
2. CC sequences are not permitted by about 10 to 15 percent of the world's languages, e.g., Fijian.
3. No languages *require* that all sections with two or more vowels contain a hiatus, nor that those with two or more consonants contain a sequence of consonants.
4. About 10 to 15 percent of the world's languages that permit consonant sequences within the section, permit none initially or finally. Almost all of these are limited to no more than two consonants, with Kannada, which possesses medial -CCC- but no initial clusters nor final consonants, being the best known exception.
5. Sections must begin with a consonant in about 20 to 40 percent of the world's languages, e.g., Hottentot.
6. Sections must end in a vowel in about 10 to 25 percent of the world's languages, e.g., Luganda.

---

[1] Although Bell and Hooper (1978) used the term 'section', which referred roughly to a word unit, they presented these statements as statements about syllable structure.

7. There are virtually no languages whose sections obligatorily begin with a vowel or end with a consonant.

8. Languages are more likely to have initial consonant clusters than final clusters. The world's languages are split about evenly between those with initial clusters and those without. But less than half, perhaps as few as one-quarter, have final clusters.

9. In final position, the single consonants that may occur are a small subset of the total segment inventory in many languages; this does not appear to occur in initial position.

10. Glides (nonsyllabic vocoids) are the most preferred interior segments.

11. Liquids are preferred over nasals as interior segments.

12. Liquids are preferred over obstruents as interior segments.

13. obstruent - nasal - liquid - glide (This refers to Bell and Hooper's suggested consonant cluster hierarchy where obstruents are the least preferred interior segment).

14. Segments of a syllable must be arranged in such a way that their sonority increases from the onset to the nuclear peak, and decreases thereafter.

15. stop - fricative - resonant - vowel (This refers to Bell and Hooper's suggested nuclear peak hierarchy where stops are the least preferred).

The variety and number of these statements underscores the fact that the syllable systems of languages differ considerably. Nevertheless, if attention is directed to the similarities between the statements, just three generalizations emerge: languages prefer sequences with the fewest number of identical segment types (1 - 4); languages prefer initial consonants and disfavor final consonants (5 - 9); and, a manner-of-articulation hierarchy (e.g., stop-fricative-nasal-liquid-glide-vowel) describes the sequential organization of segments within a syllable (10 - 15). Where, the third generalization characterizes the most frequent sequential organization of identical segment types within syllables, the first two suggest a primitive syllable type, the consonant-vowel (CV) syllable. In this syllable type, only one segment from each class occurs and the consonant occurs in onset position. Evidence suggests that this syllable type is also the most frequent type in each language. Although trivially true in languages that only allow a few different syllable types, such as in Hawaiian and other Polynesian languages, this observation is also supported in languages that allow many more syllable types, such as English (Greenberg, 1997). Little wonder, then, that in some phonological accounts, the CV syllable type represents the underlying structure of any syllable type (e.g., Clements and Keyser, 1983).

In phonology, the three cross-language generalizations on syllable systems are formalized by two major principles. The Maximal Onset Principle mandates that consonants in a string should behave as syllable onsets even if this requires the formation of onset clusters (Venneman, 1972; Hooper, 1976; Selkirk, 1982). In addition to corresponding to the cross-language preference for consonantal onsets, the Maximal Onset Principle reflects the cross-language preference for vocalic offsets by ensuring that consonants are used in syllable-initial rather than syllable-final position as much as possible (Bell and Hooper, 4-9). The second principle, the Sonority Sequencing Principle, is perhaps more important, since the facts formalized by this principle motivate the concept of a syllable in phonology (e.g., Kenstowitz, 1994). The Sonority Sequencing Principle states that sonority should be greatest at the syllable nucleus and should drop off towards the edges of the syllable (Hooper, 1976; Clements, 1990). Sonority is defined in terms of a manner-of-articulation hierarchy where obstruents are the lowest and vowels are the highest in sonority. This principle, therefore, reflects the cross-language preference for alternations of segment type (Bell and Hooper, 1-3). The details of the sonority hierarchy closely parallel the cross-language segment sequencing preferences described in Bell and Hooper's statements 10 - 15.

The Sonority Sequencing Principle has been the cornerstone of most syllabification routines in phonology since Saussure's time. A version of the sonority principle was first described by Sievers and then by Jespersen at the end of the 19th century (Jakobson and Waugh, 1979/1987). Saussure referred to Sievers' version as an example of the circular reasoning prevalent in descriptive phonology, but did not really succeed in interpreting the hierarchy in more concrete terms. Given its central role in defining the syllable in phonology, the segment sequencing constraints described by the Sonority Sequencing Principle has to be accounted for in any alternative definition of the syllable.

The comparison between the language data and the formal principles of phonology illustrate that the specific content of the principles relating to a syllable unit is motivated by the data itself. The formalized generalizations of the data refer, in theory, to a speaker's internalized and innately specified grammar. In this sense, phonology defines a syllable as an abstract, mental concept that occurs in our universal and innate grammar. While the goal of phonological theory may not require concrete definitions of the fundamental units of sound organization, explaining these units in terms of an innate grammar creates definitions without explanatory or predictive value outside of the specifics of the theory. By defining principles in terms of how syllables are usually perceived in language, phonological theory cannot answer the question posed by Saussure of why "we perceive syllabic division." More generally, the theory cannot answer the questions of why or how syllables emerge as important units of sound organization. Similarly, phonological principles specify that less frequent syllable types are ill-formed or "marked," but they cannot explain why this should be so, except in the most trivial and circular sense.

It should, of course, be noted that phonetic theories that define the syllable as a cognitive or motor unit fail in the same explanatory tasks as phonological theories. The syllable is often invoked in phonetics to organize data of suprasegmental acoustic patterns. The definition of a syllable in phonetics is therefore also often descriptive and not explanatory or predictive. Whereas phonology may be able to incorporate its descriptive definition into a larger theoretical framework, a descriptive definition of the syllable in phonetics is merely descriptive. A preferred alternative available to phonetics would be to define the syllable as a functional and physical unit of sound organization. If this type of definition also succeeded in motivating the phonological and acoustic data associated with syllables, it could provide answers to questions of why and how syllable units emerge. Such a definition may also contradict the notion that the syllable is an innate unit of grammar or cognition. In addition, it would carry predictive power about the shape of syllable systems.

2.3 Alternative solutions

As noted in the introduction, this dissertation is cast in a theoretical framework that sees the sound patterns of language as having emerged (and continuing to evolve) in response to the selection pressures provided by listeners and speakers. This approach to the study of sound systems may be seen as a special case of the Saussurian approach to the study of language. Saussure famously situated *la langue* in the space between speaker and listener. He situated the executive and receptive functions of language, *la parole*, within each individual. This view of language is represented by a diagram of the "circuit of speech" reproduced in Figure 2.1. Harris (1991) notes that Saussure was the first to emphasize that language is born from the circuit that attaches speaker to listener. One consequence of this emphasis is that linguistic sounds (or "sound images") can be understood in articulatory

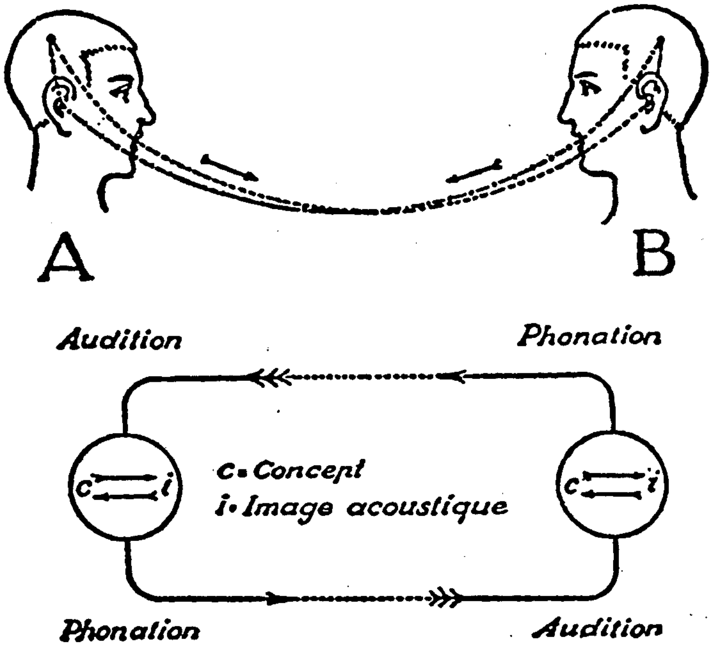


Figure 2.1: **Saussure's circuit of speech.** Language is shown to exist in the space between a speaker and listener. Though knowledge of the language resides in the heads of the interlocutors, language is also always tied to the production and perception systems.

and perceptual terms. Neither articulatory, nor perceptual criteria alone suffice to explain this aspect of language. Given his novel insight on the sounds of language, it might at first seem contradictory that Saussure is also famous for his elevation of the abstract and arbitrary nature of language as the true subject of linguistics, and his concomitant subordination of the natural or phonetic study of language as an auxiliary science of language (Harris, 1991). Nevertheless, this distinction may be interpreted to mean that even though the sounds of language represent the foundation upon which the conceptual structure of language is overlaid, sounds are also inextricably tied to more general, physical principles, which may be studied without reference to language. Saussure's treatment of the problem of syllables provides some support for this interpretation of the connection between the valued linguistics of *la langue* and the undervalued linguistics of *la parole*. Because Saussure was the first to cast the problem of syllables in the theoretical framework adopted in this dissertation, his theory of syllabic division is reviewed here.

2.3.1 Combinations of implosion and explosion

At the beginning of the 20th century, Saussure admonished phonologists for "considering abstractions real units without examining more carefully the definition of the unit (1959:53)." It would appear from the current phonological theories of syllabification that his admonition was not heeded. The correct approach to defining a syllable, according to Saussure, is to find the "irreducible units" in the speech stream and their laws of combination. For Saussure the relevant irreducible units for syllables were acoustic/perceptual in nature, but they were derived from underlying articulatory configurations. The laws were to be established in a similar manner. He insightfully argued that "freedom in linking phonological species (i.e., phonemes) is checked by the possibility of linking articulatory movements (1959:51)." Thus, Saussure advocated a functional approach to explaining phonological phenomena. In spite of this urging, Saussure was not a phonetician and so the specifics or phonetics of his proposal were borrowed from others: for instance, he adopted Jespersen's framework of sound distinctions (Harris, 1991). In evaluating Saussure's theory, it is therefore important to recognize that his contributions lie in the questions posed and the approach used to answer these questions.

To answer to the question of why we perceive syllabic divisions in every speech chain, Saussure turned first to examining the irreducible units of the speech stream. This examination apparently lead him to propose implosion (>) and explosion (<), or closure and release, as irreducible units. These units apply between segments, but also within the articulation of a single segment. Thus, he notes that *appa* may be perceived as *ap.pa*[2] because the first [p] is the closed variant and the second is the released variant (><) of the phoneme /p/. The conjunction of these two variants results in a perceptible sound, which marks a syllable boundary. Similarly, he notes that the implosive and explosive variants of a single segment also exist for most vowels. For example, the dual nature of [i] can be seen in the orthographic realization of the segment as "y" for the closed [i] variant and "ee" for the open

---

[2] Throughout this dissertation, a '.' will serve to denote a syllable boundary.

variant. Because almost every segment can be realized as both a closing and opening sound, Saussure recommended that the phonemic inventory be expanded to include both types. The possible combinations of closing and opening sounds could then be discussed.

An exhaustive list of four combination were proposed (<>, ><, <<, >>), but Saussure noted that the perception of a syllable boundary occurs only when implosion and explosion units are combined (><) and only if that combination produces a sound. This particular definition of a syllable boundary is important because it stresses not only the importance of articulation, but also the importance of linking articulation to an audible acoustic consequence. By extension, the definition also makes an important distinction between production and perception. Saussure recognized that even though perception usually follows from production, the two are distinct in that a perceptible sound may not arise from an articulatory event or different articulatory events may give rise to the same percept. The importance of this distinction is demonstrated in Saussure's account for why the implosive/explosive sequence found in the combination [sp] may be perceived as an onset cluster rather than split. He explained that the "furtive sound" produced by the combination of these two sounds "in no way interferes with the succession of the chain (1959:55)."

The theory also accounted for why the same combination of segments may be divided differently. For instance, the sequence *apa* is heard as *a.pa* when the explosive variant of [p] is used (e.g., ><<), but at *ap.a* when the implosive variant is used (e.g., >><). This account, however, is less satisfactory than the account of the [sp] cluster. To explain the different syllabifications of the same segment sequence Saussure relied on the fact that he had expanded the segment inventory to include the implosive and explosive variants of a single phoneme without indicating when one should be used instead of the other. There is no principled explanation for why *apa* is sometimes produced *a.pa* and other times *ap.a*. According to the theory, both linkages are equally possible in spite of the fact that *a.pa* would be by far the more frequent pronunciation of the string *apa* across languages. Thus, we find that the linkages between implosion and explosion units are not sufficiently "checked by the possibility of linking articulatory movements" and no explanation can be provided for why certain combinations are preferred over others. In losing sight of the articulatory underpinnings of combination, Saussure replicates the phonologists' descriptive approach to defining syllable structure.

A possible reason why Saussure did not manage to establish an independently-motivated combinatorial system may be that Saussure's analysis of the articulatory underpinnings of syllables was based on adjacent segment sequences rather than on the whole syllable. Saussure attempted to derive syllables from the sequentially organized movements of the lips and tongue. This type of analysis is problematic because, although a syllable is made up of sequentially organized segments, the percept of the syllable is often formed from the relationship between multiple, nonadjacent segments. For instance, the duration of all segments varies as a function of syllable position (Lehiste, 1970; Oller, 1973; Klatt, 1976), but the overall pattern may not be perceived if only adjacent segment durations are compared (but see, Tuller and Kelso, 1991). It is therefore most likely that syllable perception is rooted in an analysis of relative duration that extends across multiple segments (e.g., Boucher, 1988; Anderson and Port, 1992). Accordingly, a production-based theory of the syllable needs to focus on movement patterns whose acoustic effects would be spread over multiple segments in a predictable fashion. Such a theory would provide an articulatory basis for the segment combinations that form syllables.

2.3.2 Chest pulses as the suprasegmental movement that defines syllables

In contrast to Saussure, Stetson's (1951) analysis of the syllable was not limited by a sequential view of the organization of speech or a naive understanding of articulatory processes. Stetson recognized that syllables were not usefully defined as assemblies of individual sounds, just as the movements of individual articulators were not usefully conceived of without respect to the whole speech system. In his words:

> The various boundary markers ("Grenzsignale"), stress and intonation patterns which have been noted, are not independent traits, appearing isolated as members of a series of symbols; they are rather cues to these basic, coordinated movement units which make up connected articulate speech (1951:4).

Stetson's special emphasis was on the action of the breathing mechanism during the articulation of speech. He showed that, contrary to what one might assume, the chest does not provide steady pressure during exhalation in speech. Instead, the chest muscles (intercostals) contract in short intervals giving rise to air pulses upon which vowels sounds are formed. Stetson argued that air pulses could be released and arrested by the activity of the intercostals alone or by consonantal closure formed in the mouth by the tongue or lips. More than one consonant might occur at the beginning or ending of an air pulse, so long as only one of the consonants arrested air flow. According to Stetson, a syllable could therefore be defined by a single chest pulse and delimited by the release and arrest of the air pulse.

Stetson's motor definition of a syllable was disputed by Ladefoged and colleagues (1967; Ladefoged, Draper, Whitteridge, 1958). These researchers found that a one-to-one correlation between chest pulses and syllables did not always exist, it only 'usually' existed. On the strength of this counterevidence, Stetson's definition was dismissed and the mentalist definition of the syllable prevailed.

But the fact that Stetson's definition was so readily overturned may have less to do with the data presented by Ladefoged and colleagues and more to do with Stetson's own position on speech, which seems improbable. Stetson argued that: "In the individuality of the syllable the sound is secondary; syllables are possible without sound. Speech is rather a set of movements made audible than a set of sounds produced by movements (1951:33)." Given that the function of speech is to communicate information to a listener via an acoustic medium, why would sound be secondary? It may be that in Stetson's theory sound is secondary by necessity: no sound structure is inherent in an air pulse. Sound itself is derived from vocal fold vibration and sound structure from the different, sequential configurations of the supraglottal vocal tract. Because air pulses have no inherent sound structure, Stetson's attempt to account for sound patterns associated with syllables seems particularly *ad hoc*. Hence, the fact that syllables are not always accompanied by a chest pulse may be a less relevant critique of the theory, than the fact that Stetson's syllables were divorced from the relevant acoustic aspects of speech.

2.3.3 Combining approaches for a solution

In formulating a solution to the problem of the syllable, it is important to recognize the virtues of both Saussure and Stetson's theories and to use these virtues as a model from which to proceed. Saussure and Stetson both tried to explain syllables and the sound patterns associated with syllables in physiological terms rather than in terms of the language data itself. Specifically, both hypothesized that articulatory events gave rise to a well-defined syllable. In Saussure's case, though, the syllable was also a perceptual phenomenon. The articulatory underpinnings were important insofar as they had some acoustic effect on the speech stream. Saussure recognized that production and perception were usually, but not always, linked. Production may define the types of sounds that are produced and the manner in which they can be linked, but perception gets away from production when the acoustic effects of a particular articulation are not audible. Syllable production should therefore be considered in conjunction with syllable perception.

Even though Saussure recognized that the audible outcome of articulation was important, he did not have Stetson's integrated view of the speech system. According to Saussure, the important parts of the speech system were the lips, the tongue, the teeth, the hard and soft palates, and the uvula. In other words, for Saussure, the parts of the speech system that mattered were those clearly involved in segmental articulation. Consequently, Saussure located the perception of a unit larger than the segment in the sequential articulation of units even smaller than the segment. As was previously indicated, a sequential or segmental analysis of the syllable cannot explain the perceptual gestalt of the syllable if that gestalt is assumed to emerge somehow from articulatory factors. The real virtue of Stetson's theory was that the syllable movement was assumed to be slower than the movements that define segmental articulation. In this way, Stetson was able to derive a suprasegmental unit of sound organization from the hierarchical organization of the speech system. The problem, however, was that the movement chosen was so far removed from the local acoustic structure of speech that Stetson was unable to define syllable boundaries or internal structural attributes of syllables, which, as the phonological theories emphasize, are the main obstacles to defining a syllable.

Saussure and Stetson's approach to the problem of defining the syllable are worthy of emulation because they employ an approach in which non-language data are used to explain sound patterns in language. But in attempting to build a bottom-up account of the syllable both theorists are ultimately foiled by the language data. Both Saussure and Stetson redefine the Sonority Sequencing Principle in their accounts, but this effort is descriptive. With respect to syllable types, Saussure departs from a data-first approach and doubles the phoneme inventory. He then allows for this new phoneme set to be recombined by speakers in language-specific ways without acknowledging that certain combinations are more prevalent than others. Stetson, on the other hand, mostly ignores the problem of sequencing constraints and syllable structure altogether. A bottom-up account should, in principle, be able to explain these language data since they reflect information about the essential characteristics of a syllable. If syllables can be usefully defined in terms of speaker-related or articulatory factors, as proposed by both Saussure and Stetson, then these same factors should help explain the cross-language preferences in segment sequencing and syllable structure.

2.4 An articulatory basis for the syllable

In any attempt to define the syllable, its boundaries, and its variable but constrained structure, a reasonable point of departure is to explain the emergence of its most basic and ubiquitous form. As noted in section 2.1, the basic syllable is the consonant-vowel or CV syllable. This simple structure is interesting because it is built on the unity of two sounds, which are universally considered to be of distinct classes (Jakobson and Waugh, 1979/1987). In phonetics, the distinction between these two classes is made on the basis of articulation. Consonants, as a class, are those sounds that are produced with a completely or mostly constricted vocal tract, whereas vowels, as a class, are produced with a relatively open vocal tract (e.g., Straka, 1979). In a certain sense, this is also the distinction that Saussure noted and upon which his theory was based. But if these are distinct sounds, how do they come to be coordinated and why does this coordination usually take the form of a consonant, then vowel sequence? It was previously argued that a sequential analysis based on the movements of the segmental articulators, such as the one preferred by Saussure, does not provide adequate insight into this question. The alternative, exemplified by Stetson, is to use the hierarchical organization of the speech system to account for the similar organization of speech sounds.

2.4.1 Frame/Content organization

The speech system is multi-tiered, just as are speech sounds. A certain pattern of correspondences between the articulatory levels and sound levels provides us with reason to believe that one may be a partial reflection of the other. For instance, prosodic changes, such as downdrift, that occur across entire phrases are attributable to breath control (e.g., Hauser and Fowler, 1992); pitch changes that may occur over one or more words are controlled at the glottis; segmental changes are most often accomplished by the versatile articulators of the supraglottal vocal tract. The acoustic changes that relate to syllable perception usually take place over adjacent consonant and vowel segments. If articulators such as the tongue, lips and velum define segmental articulation, we might wonder what articulator would define changes over two or more segments. Given the relatively local nature of syllable-related changes, it is likely that the relevant articulator in syllable production is also supraglottal.

Like the rest of the production system, the supraglottal vocal tract is hierarchically organized. The movement paths of the segmental articulators are coordinated with the slower, cyclic movement of the jaw (Perkell, 1969; Munhall, Ostry, Flanagan, 1991; Gracco, 1994). MacNeilage (1998) has argued that this organization may have provided a basis for the emergence of syllable-like units in phylogeny and ontogeny. Specifically, he has argued that the simple movement of the jaw, from rest position to an open position and back, provides a "frame" within which the segmental articulators position themselves for close and open configurations, thus producing consonants and vowels – the "content" of the frame. This view is formalized as the Frame/Content Theory of speech production. Although many phoneticians have remarked on the coordinated movements of the lips, tongue, and jaw, the traditional and dominant view has been that the jaw moves *in service* of the segmental articulators (e.g., Stetson, 1951; Perkell, 1969; Keating, Lindblom, Lubker, Kreiman, 1994) and not that the movements of the segmental articulators are nested into a continuous jaw cycle. The view that the jaw moves in service of the fast articulators is reminiscent of Saussure's sequential analysis of speech sounds and does not provide insight into why sounds are organized as CVCV sequences as opposed to CCCC or VVVV sequences.

In MacNeilage's view the cyclic movement of the jaw not only provides an explanation for the CV alternation in speech, but also provides a scaffolding upon which the rapid sequence of segmental articulations in speech are first pegged. The scaffolding function of jaw movement is evident in babbling. MacNeilage and Davis (1990; Davis and MacNeilage, 1995) have provided evidence that the varied CV sequences of infant babbling, which are perceived by listeners as syllables, are produced almost entirely by jaw movement during phonation with little or no contribution from other articulators. The entire CV sequence therefore reflects either tongue-fronting and jaw movement, as in [dididi], tongue-retraction and jaw movement, as in [gugugu], or pure jaw movement as in [bəbəbə]. When variegated babbling occurs, the consonants and vowels in the sequences differ not in tongue movement (place-of-articulation), but rather in jaw height (manner-of-articulation) (Davis, MacNeilage, Matyear, 1999). For example, a typical variegated sequence of babbling might be comprised of a [glideVstopV] sequence where the intervening vowels were of different heights and all segments exhibit the same relative fronted, retracted, or null place-of-articulation. Example utterances of this type would be [jedi] or [wəba]. MacNeilage and Davis have explained that these variegated babbled sequences result from amplitude modulations of the jaw cycle. Sequential variation that involves multiple changes in tongue position appears later in development when the infant has gained greater control over the fine musculature of the tongue, but frame dominance – the dominance of the jaw cycle in sound production – continues even into the first words (MacNeilage, Davis, Matyear, 1997) and to some extent throughout life (*viz* CV co-occurrence constraints across languages, MacNeilage, Davis, Kinney, Matyear, 1999).

2.4.2 The frame in adult speech

According to the Frame/Content theory of speech production, the jaw cycle provides structural support for segmental articulation. This view works for babbling because infants have not developed independent control over their articulators. In contrast to babbling, the sound combinations of adult speech rely much more heavily on the contributions of articulators other than the jaw. Nevertheless, the rhythmic open-close cycle of the mandible, which defines infant babbling, also characterizes adult speech production (Stone, 1981; Erickson, Lenzo, Fujimura, 1994). Given mature control over segmental articulation, how might the jaw cycle function in adult speech? One possibility continues the metaphor of frame and content.

A frame may offer structural support for the presentation of content, but it also imposes limits on its realization. As previously indicated, the movements of the segmental articulators are coordinated with the cyclic movement of the jaw in adult speech. This coordination must arise in large part from the physical construction of the supraglottal vocal tract. The most versatile and important segmental articulator, the tongue, is attached to the mandible and so is the lower half of the other major articulator, the lips. This means that when the jaw moves, the body of the tongue and lower lip move with it. If we assume, like MacNeilage, that the jaw cycle is basic to speech, then it is reasonable to assume that the pattern established in development will continue and that the movements of the segmental articulators in adult speech will also conform more to the movements of the jaw than vice versa. In this way, the simple open-close cycle of the jaw may act as a mechanical constraint on segmental articulation. Since the jaw also appears to have a preferred rate of movement, with a normal cycle duration of around a quarter of a second (e.g., Ohala, 1975; Nelson, Perkell, Westbury, 1984), the cycle may also act as a temporal constraint on segmental articulation. In this dissertation, it will be argued that the mechanical and temporal constraint of the jaw cycle is manifested in the phonological and phonetic sound patterns that are perceived as syllables.[3] For this

---

[3] Note that the constraint of the jaw cycle is circumvented when jaw movement is circumvented, as in glottal or pharyngeal articulations. Articulate speech is, however, by and large supraglottal. The constraint of the jaw cycle may be circumvented in other special cases, but only at the expense of naturalness. For instance, bite-block experiments and pipe-speech show that perceptible speech is possible with a clenched jaw. Given the necessary activation of the anterior digastric - one of the primary opening muscles of the jaw – during production involving the depression and retraction of the tongue, it is possible that the jaw may remain clenched during these productions only because the large closing/clenching muscles of the jaw (e.g., the masseters, which are normally not used in

reason, it is proposed that the jaw cycle may provide an articulatory basis for the syllable in language.

2.4.3 Defining the syllable

The basic and most frequent segment sequence in adult languages is the consonant-vowel sequence, perceived as the CV syllable. As in babbling, this sequence is the product of the articulatory frame provided by the jaw cycle. Vocal tract configurations associated with consonant production take place during the least open portions of the cycle and those associated with vowel production during the most open portions. Unlike babbling, though, the resultant consonant-vowel sequences are not fortuitous byproducts of the cycle, they are simply the most efficient sequence type given the constraints of the cycle.

The simple constraint of where consonants and vowels may occur in the cycle is demonstrated in the speech error data. Speech errors, specifically spoonerisms, involve the exchange of segments within or between syllables. Vowels and consonants may switch places with the vowels and consonants of other, usually adjacent, syllables. Consonants may also be switched within a single syllable. But consonants and vowels never exchange places within or between syllables. MacNeilage (1998) argues that consonants and vowels are never interchanged because these must always occur in different portions of the jaw cycle.

Although the most efficient sequence type, given the constraint of the cycle, involves a single articulation for a consonant and one for a vowel, multiple segments may occur within a cycle when the segmental articulators increase their rate of movement. Each of these additional segments will, however, be articulated with varying degrees of jaw opening corresponding to where they occur in the cycle. Since different consonant and vowel segments are preferentially articulated with different degrees of jaw opening (Keating, Lindblom, Lubker, Kreiman 1994), a normal sequence of segments will emerge that is best defined by the jaw cycle. Thus, the jaw cycle may also provide a basis for the normal sequential organization of segments. If this normal organization parallels the organization described by the Sonority Sequencing Principle, we can see how the jaw cycle may provide an articulatory basis for the syllable.

The constraint of the jaw cycle on segmental articulation is also manifested in the phonetic patterns associated with syllable perception. One of the main cues associated with the perception of syllables is the relative acoustic duration of all

---

speech), are able to overcome the action of the smaller opening muscle.

segments within a syllable (Boucher, 1994; Anderson and Port, 1994). Although the absolute amplitude of the jaw cycle may be determined by the type of vowel to be produced (high, mid, or low), the relative duration of the segments is affected because the amplitude or size of the cycle correlates positively with its duration or length. For instance, Lindblom (1967) showed that differences in vowel duration could be explained in terms of this model. High vowels, articulated with a relatively closed jaw were of shorter duration than low vowels, articulated with a relatively open jaw. In the case of additional segments, the absolute amplitude of the cycle may or may not increase (Munhall, Fowler, Hawkins, Saltzman, 1992), but in either case the duration of the cycle does not increase in a linear fashion according to the number of segments added (Sigurd, 1973). As a result, multiple consonant (or vowel) segments will not be articulated with the same duration as when they occur by themselves during the relatively closed (or open) portions of the cycle.

So far in this dissertation, the jaw cycle has been assumed to constrain segments in a uniform manner. Given that the cycle is not symmetrical, this assumption is unlikely. The opening phase of the cycle is of longer duration and executed at a slower speed than the closing phase (Sussman, MacNeilage, Hanson, 1973; Kuehn and Moll, 1976; Kelso, Vatikiotis-Bateson, Saltzman, Kay, 1985; Gracco, 1994). This consistent difference between the phases of the cycle may motivate certain structural characteristics of the syllable captured by the Maximal Onset Principle of phonology. For instance, the cross-language preference for syllable-initial consonants over syllable-final consonants may result from the asymmetrical phases of the jaw cycle. Initial consonants may be produced more distinctively and with less variability than final consonants (e.g., Byrd and Tan, 1996; Sussman, Bessell, Dalston, Majors, 1997; Redford and Diehl, 1999) because there is more "room" within the initial part of the cycle compared with the final part of the cycle.

In sum, three major hypotheses are proposed regarding the relationship between jaw movement and syllables. (1) The mechanical constraint of the jaw cycle on segmental articulation may provide a basis for the normal sequencing of segments within the cycle. (2) The temporal constraint of the jaw cycle may provide a basis for acoustic patterns associated with syllable perception. In addition, (3) the opening and closing phases of the jaw cycle are not symmetrical and may therefore motivate the preferred cross-language structure of syllables. Each of these hypotheses is examined in the chapters that follow in an attempt to support the idea that the jaw cycle provides an articulatory basis for the syllable.

Chapter 3

The Constraint of the Jaw Cycle

The argument that the jaw cycle provides an articulatory basis for the syllable is dependent upon the assumption that the jaw cycle constrains the movements of the segmental articulators and that this constraint is realized in sound patterns perceived as syllables. The goal of this chapter is to support these assumptions with evidence. Specifically, this chapter focuses on the cross-language occurrence of segment sequencing constraints. It is hypothesized that the normal sequential organization of segments within a syllable emerges naturally from the constraint of the jaw on segmental articulators. The alternative hypothesis is that the jaw moves in service of the segmental articulators and according to the cognitively-based Sonority Sequencing Principle of phonology. In order to distinguish between these two possibilities the jaw movements and acoustic durations associated with the production of reversed-sonority and normally-sequenced consonants are examined.

3.1 Background

In articulatory terms, consonants and vowels are most easily contrasted along one dimension, namely, the degree to which the vocal tract is constricted (e.g., Straka, 1979). Consonants, relative to vowels, are produced with a greater degree of vocal tract constriction, but not all consonants or vowels are produced with the same degree of vocal tract constriction. Obstruent consonants, such as fricatives and stops, are produced with more vocal tract constriction, than sonorant consonants, such as liquids and glides. Similarly, high vowels are produced with more vocal tract constriction than low vowels (Perkell, 1969). The vocal tract configurations required to produce different segments are described in terms of tongue and lip movements, but the degree of constriction required for a particular segment is correlated with the degree to which the jaw is raised or lowered (e.g., Lindblom, 1983; Keating, Lindblom, Lubker, Kreiman, 1994). The different relative degrees of vocal tract constriction or jaw opening have perceptible acoustic consequences. These consequences are captured by the qualitative term "sonority," which Jesperen described as "auditory prominence" (1921, cited by Butt, 1992). Perhaps as a consequence of this description, sonority is usually thought to be most closely related to the acoustic parameters of relative intensity and duration (Price, 1980; Ladefoged, 1993). Sounds with low sonority are produced with more vocal tract constriction and a more closed jaw than sounds with high sonority.

In Chapter 2 it was noted that the normal sequential organization of segments within a syllable is characterized in terms of a manner-of-articulation hierarchy referred to as the sonority hierarchy (and formalized as the Sonority Sequencing

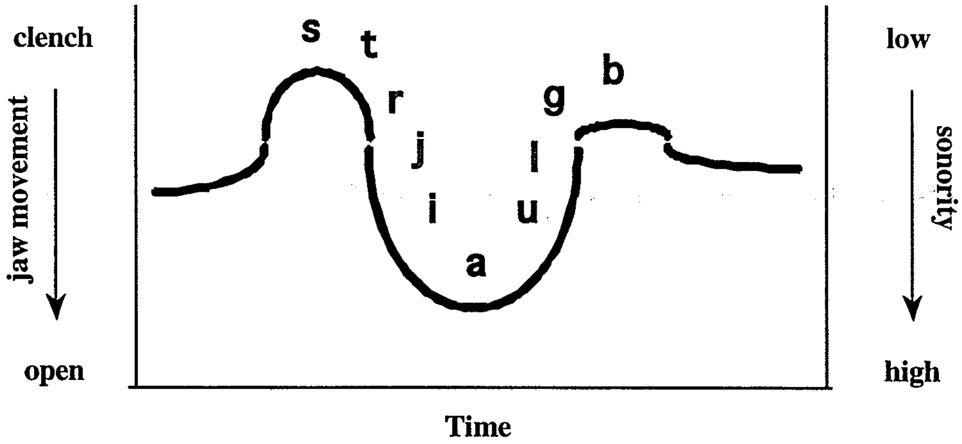


Figure 3.1: **Correspondence between jaw openness and sonority scale.** Schematic of Lindblom's (1983) results, which show a relationship between the sonority hierarchy and relative jaw opening.

Principle). The least sonorous segments define the edges of the syllable and the most sonorous element defines the peak or nucleus. Sonority increases from the edges of the syllable to the peak. Given the previously described relationship between sonority and jaw opening, it might be expected that the sonority hierarchy could be redescribed in terms of relative jaw opening. Lindblom (1983) expected as much and so measured average jaw height during the articulation of different Swedish consonants and vowels. When he plotted relative jaw height against relative sonority he found a remarkable correspondence between jaw openness and the sonority hierarchy. This finding is schematized here in Figure 3.1. The correspondence between sonority and jaw height might be interpreted in different ways. Lindblom interpreted his results to mean that the sonority hierarchy reflects speakers' "propensity" to coarticulate consonant segments with the vowel. Yet this propensity may ultimately reflect either a cognitive constraint or a mechanical constraint, depending on how one views jaw movement.

It has often been said that the jaw *supports* segmental articulation (e.g., Perkell, 1969; Gracco, 1994; Stone and Vatikiotis-Bateson, 1995). On this view, jaw height is dependent on the flow of segments, which is defined elsewhere, probably by the Sonority Sequencing Principle. One of the problems with this view is that it cannot answer the question of why languages universally organize segments in a particular manner. The alternative view proposed here is that jaw movement is basic to articulate speech and that the action of the segmental articulators tend to conform more to the continuous open-close jaw cycle than vice-versa. On this view, the sequential organization of segments emerges naturally in speech in a manner that tends to give rise to the sonority hierarchy. One way to distinguish between these possibilities is to determine whether sequences that are perceived as single syllables by native speakers, but that violate the sonority principle, still conform to the open-close jaw cycle. For instance, the sonorant-stop onsets from a small set of monosyllabic Russian words (e.g. [lba] "forehead," [lgat^j] "to lie"). The present study uses exactly this type of test to establish whether jaw movement conforms to the sonority hierarchy or to its own basic cyclicity.

Data were collected on the jaw movement of 3 native Russian speakers while they produced different types of legal Russian syllables. These included simple syllables with a single consonantal onsets, syllables with initial clusters that obeyed the sonority principle, and syllables with reversed sonority clusters. Measurements were taken on the relative jaw position during articulation of the segments. It was predicted that the mechanical constraint of the jaw cycle would be a better predictor of relative jaw height than the relative sonority of a segment. Specifically, it was predicted that stops in the first consonantal (C1) position of a cluster would be articulated with a relatively closed jaw configuration compared with when they occurred in the second consonantal (C2) position of the cluster.

In Chapter 2, it was hypothesized that, in addition to a mechanical constraint, the jaw cycle provides a temporal constraint on segmental articulation. This temporal constraint is thought to influence relative segment duration within a cycle. For instance, the relative amplitude of the cycle, though initially specified by segmental content, affects cycle duration, which in turn affects segment duration (e.g., Lindblom, 1967). Although the relative duration of the cycle may change depending on the amplitude of the cycle, the change is moderated by the fact that the jaw appears to have a preferred oscillating frequency of about 4 cycles per second (e.g., Ohala, 1975; Nelson, Perkell, Westbury, 1984).[1] This preferred oscillating frequency may provide a further temporal constraint on the articulations of segments within the cycle so that segment duration will be inversely correlated with the number of segments articulated within a single syllable. Thus, the constraint of the jaw cycle might provide an explanation for why vowel duration decreases when final consonants are added to the syllable (Lindblom and Rapp, 1973; Munhall, Fowler, Hawkins, Saltzman, 1992) and why the sum of the durations of consonants in a cluster is not equal to the sum of the durations of the same consonants when they occur singly (Sigurd, 1973). To determine whether differences in segment duration were attributable to the jaw cycle, measurements of acoustic duration were made on each of the segments. It was expected that, as in Lindblom (1967), a correlation would be found between jaw height and acoustic duration. In addition, it was expected that, as in Sigurd (1973), the relative duration of the consonants would differ as a function of the number of segments in the syllable and as a function of position in the syllable.

3.2 Study Methods

3.2.1 Stimuli

One female and two male native Russian speakers produced 42 single syllables in a frame sentence. The tokens were consonant-vowel (CV), sonorant-vowel (SV), consonant-sonorant-vowel (CSV), and sonorant-consonant-vowel (SCV) syllables. The consonants were the voiced stops [b] and [g], the sonorant was the liquid [l]. Russian has two variants of this liquid, a palatalized and pharyngealized variant. In the present stimuli all liquids were pharyngealized. The vowels were the point vowels [i], [u], [a]. Most of the SCV tokens were actual monosyllabic Russian words, for example, [lba], "forehead" (sing. gen.), [lgu], "I lie". In contrast, the CV, SV, and CSV tokens, though also legal syllables in Russian, were not actual Russian words, for example, [glu] as in the first syllable of [glu.xa], "deaf" (fem.). Each syllable type was said twice in the sentence [poi ___ s nova].

The speakers read the written form (Cyrillic) of the tokens in the frame sentence from a randomized list of the tokens. The sentences were recorded with a Nakamichi CM700 microphone directly into a pentium PC using a waveform editor developed in the Speech Perception Laboratory. The audio data was sampled at

---

[1] If we pursue the resonance metaphor, it is helpful to recall that resonances have bandwiths. If a system is relatively damped, as in the case at hand, we can expect that the bandwith will be relatively wide and the system will therefore have a range of preferred oscillating frequencies.

11025 Hz. Jaw movement was recorded simultaneously by means of two strain gauges attached to a depressor. The depressor was fixed under the speakers chin with a light-weight head-mount, which the speaker wore while producing the stimuli. Jaw movement was sampled at 100 Hz. Movement calibration was achieved by recording the speaker with a clenched jaw and with a 1 cm spacer inserted between the premolars. Two calibration recordings were made at the beginning and ending of each 10 minute recording session.

3.3 Measurements

The temporal onset and offset of each segment of a token was measured. The temporal onset and offset of a segment was determined by visual inspection of the waveform and by auditory analysis. The onset/offset of stop segments corresponded to abrupt changes in the amplitude envelope of the waveform, though some periodicity was generally present throughout the stop closure. The boundary between the liquid and a vowel also corresponded to changes in the shape of the waveform, but with sonorant characteristics. Demarcation of the liquid boundary was coupled with auditory judgments. Vowel offsets corresponded to the onset of frication of the following [s] from the frame sentence. The midpoint of each segment equaled the exact midpoint between the onset and offset of the segment.

The temporal points and corresponding jaw heights were recorded at minimum and maximum jaw opening for each of the tokens. The minimum and maximum points corresponded to the onset and the midpoint of the cycle respectively. Measurements of jaw height were also taken for each segment. The acoustic and movement waveforms were aligned. Jaw height measures were taken at the absolute midpoint of the acoustic segment. Figure 3.2 shows a schematic of acoustic and movement waveform, along with the measurement points.

3.3.1 Analyses

Syllables with different stop types were collapsed in the analyses. The collapsing of stop types meant that there were fewer observations for SV syllables than for any other syllable type. Parity between SV syllable observations and observations for other syllable types was restored by using the average value as the values for missing observations. Due to the limited repetitions of each token, the data were also collapsed over speakers. Consequently, individual differences are not explored in this study.

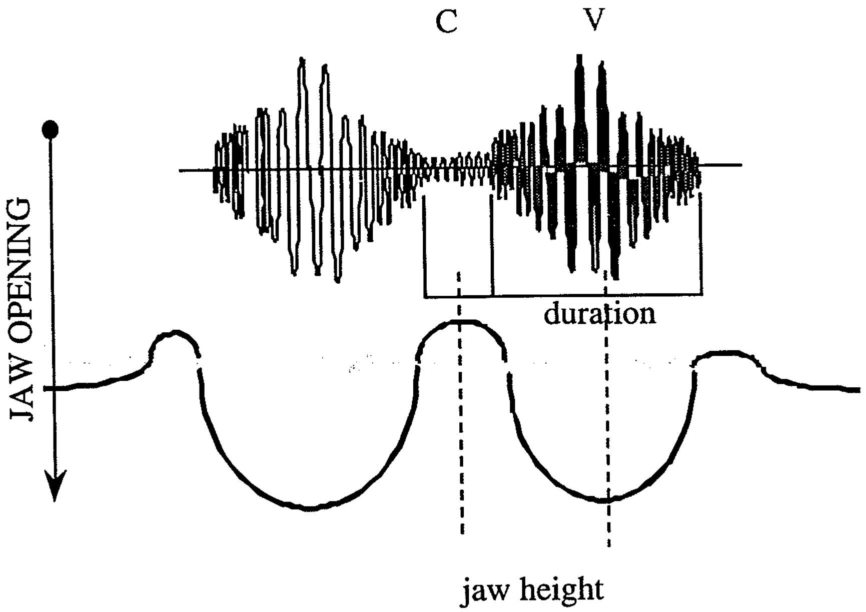


Figure 3.2: **Illustration of measurement locations.** Acoustic and movement waveforms were aligned. Jaw openness measurements were taken at the absolute midpoint of the acoustic segment.

3.4 Results

Each of the measurement sets is first considered individually. The patterns of jaw displacement and relative degree of jaw opening are reported primarily as a function of syllable and segment type. The results reported for different patterns of acoustic duration also focus on this measure as a function of syllable and segment type. In a final set of analyses, data on jaw movement and acoustic duration are directly compared.

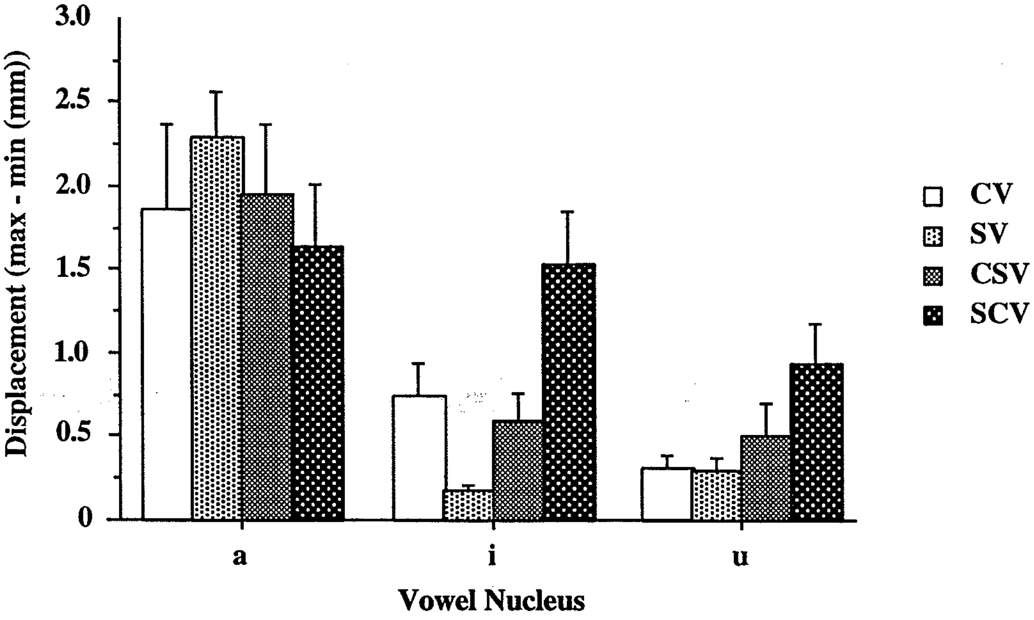


Figure 3.3: **Displacement as a function of syllable type and vowel nucleus..** Displacement was measured during the opening phase of the jaw cycle as the distance (in millimeters) between minimum and maximum jaw opening. The four syllable types were the consonant-vowel (CV) syllables, the sonorant-vowel (SV) syllables, the syllables with normally-sequenced onset cluster (CSV), and syllables with reversed-sonority clusters (SCV).

3.4.1 Jaw openness

Overall

Total jaw displacement during opening was considered to be the distance in millimeters between minimum and maximum jaw opening. A two-way Analysis of Variances (syllable type x vowel type) indicated a main effect for vowel type ($F(2,22) = 26.080, p < 0.01$). As expected, when the syllable nucleus was the low vowel [a], displacement was greater than when the syllable nucleus was either of the high vowels [i] or [u] ($F(1,22) = 50.800, p < 0.01$). This main effect can also be seen in Figure 3.3, which shows displacement as a function of vowel and syllable type.

While there was no main effect of syllable type, a non-significant trend was that syllables with reversed-sonority onset clusters were articulated with greater displacement than syllables with normal onset clusters or single onsets. A statistically significant interaction between syllable type and vowel type (shown in Figure 3.3) indicated that syllables with reversed-sonority onset clusters were only articulated with more displacement when the syllable nucleus was one of the high vowels (for [i]: CV vs. SCV ($F(1,66) = 6.601, p < 0.05$); SV vs. SCV ($F(1,66) = 19.594, p < 0.01$); CSV vs. SCV ($F(1,66) = 9.227, p < 0.01$); for [u]: CV vs. SCV ($F(1,66) = 4.200, p < 0.05$); SV vs. SCV ($F(1,66) = 4.301, p < 0.04$)).

Individual segments

Next, the relative degree of jaw opening was analyzed for the different consonants as a function of syllable type and vowel nucleus. A three-way (consonant type, syllable type, vowel type) ANOVA indicated that, overall, stops and liquids were not articulated with different degrees of jaw opening. Instead, jaw opening for the articulation of the consonants differed as a function of the following vowel ($F(2,22) = 24.635, p < 0.01$). Consonants were articulated with more jaw opening when the following vowel was the low vowel [a] than when it was either of the high vowels [i] or [u] ($F(1,22) = 36.61, p < 0.01$). Consonants preceding [i] were articulated with more jaw opening than those preceding [u] ($F(1,22) = 12.657, p < 0.01$).

The interaction between consonant type and syllable type, shown in Figure 3.4, was not significant, but mean comparisons indicated a statistically significant difference in jaw opening between the stop and liquid consonants of syllables with reversed-sonority onset clusters (stop vs. liq, in SCV, ($F(1,22) = 9.688, p < 0.01$)). The liquid consonant is articulated with significantly less jaw opening than the stops when it is in the first position of the cluster. Interestingly, the liquid consonant is articulated with the same degree of jaw opening in both first and second position of the onset cluster. On the other hand, the stop consonants are articulated with more jaw opening when they appear in the second position of the onset cluster than when they appear as the first consonant (stop in CSV vs. SCV ($F(1,22) = 5.068, p < 0.05$). While the lack of difference in jaw opening between the first and second consonants of the normally-sequenced onset cluster (CSV) does not constitute a violation of the jaw cycle, it is still surprising.

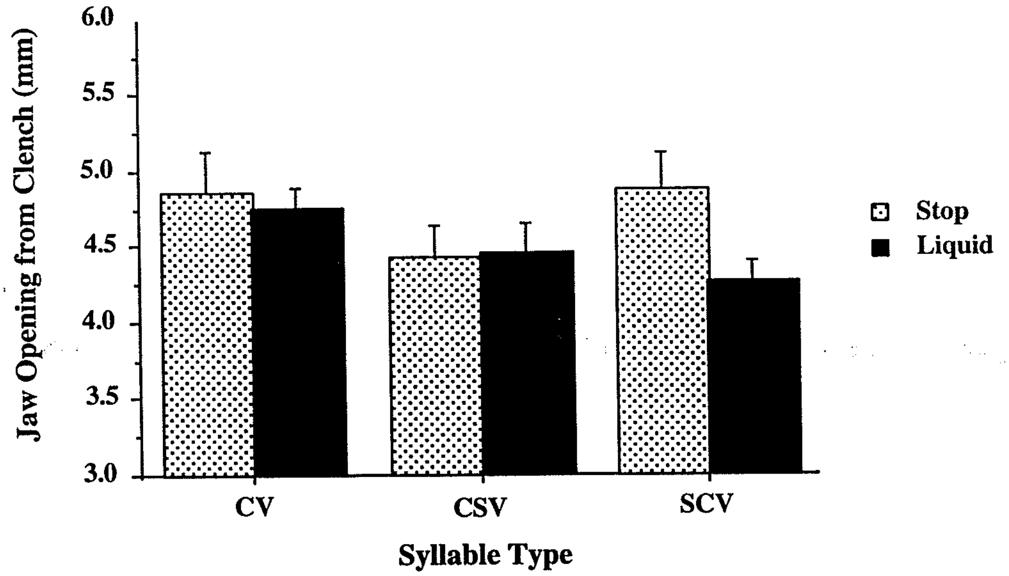


Figure 3.4: **Jaw opening for consonants as a function of syllable type..** Jaw opening was measured from clench (in millimeters) for the stop and liquid consonants of the four syllable types. The four syllable types were the consonant-vowel (CV) syllables, the sonorant-vowel (SV) syllables, the syllables with normally-sequenced onset cluster (CSV), and syllables with reversed-sonority clusters (SCV).

3.4.2 Acoustic duration

Overall

Differences in total acoustic duration of the syllables[2] was tested in a two-way (syllable type and vowel type) ANOVA. Both main effects were statistically significant: overall syllable duration differed as a function of syllable type ($F(3,33) = 73.162, p < 0.01$) and vowel type $F(2,22) = 11.704, p < 0.01$). Mean comparisons indicated that syllable with onset clusters were of greater duration than those without (CV+SV vs. CSV+SCV ($F(1,33) = 191.79, p < 0.01$)) and that differences in syllable duration as a function of vowel type paralleled the differences in opening jaw displacement. When the syllable nucleus was the low vowel [a], duration was greater than when the syllable nucleus was either of the high vowels [i] or [u] ($F(1,22) = 20.78, p < 0.01$).

---

[2] Syllable duration equaled the summed durations of the individual consonant and vowel segments.

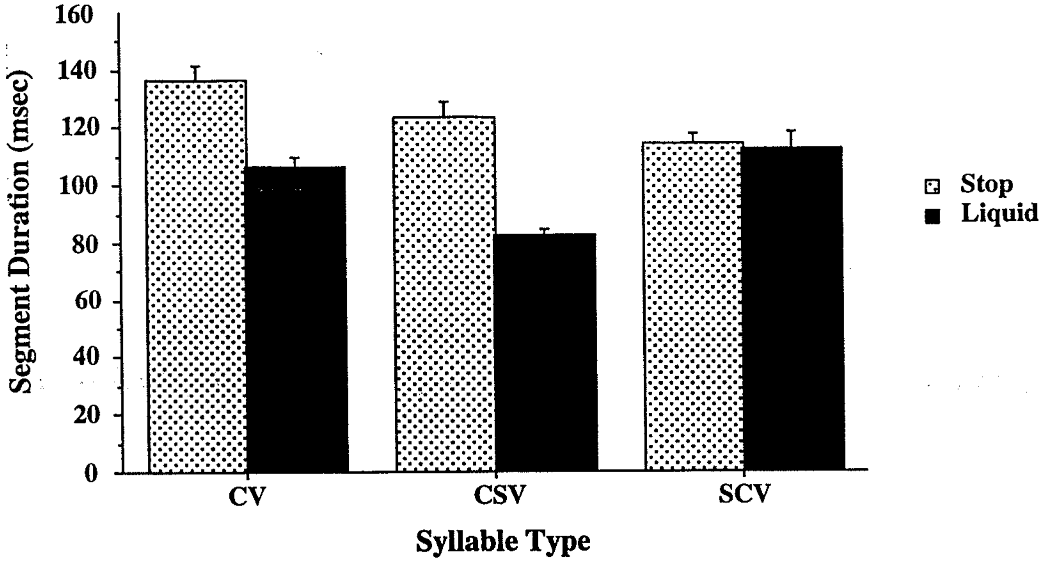


Figure 3.5: **Consonants duration as a function of syllable type.** Acoustic duration was measured (in milliseconds) for the stop and liquid consonants of the four syllable types. The four syllable types were the consonant-vowel (CV) syllables, the sonorant-vowel (SV) syllables, the syllables with normally-sequenced onset cluster (CSV), and syllables with reversed-sonority clusters (SCV).

Individual segments

The pattern of relative consonant duration was analyzed in a three-way (consonant type, syllable type, vowel type) ANOVA. Consonant duration differed as a function of consonant type ($F(1,11) = 15.715, p < 0.01$) and syllable type ($F(2,11) = 8.090, p < 0.01$), but not as a function of vowel nucleus. Stops were generally of greater duration than liquids, and consonants that occurred as single onsets were greater in duration than those that occurred as part of a cluster (CV+SV vs. CSV+SCV) ($F(1,22) = 13.500, p < 0.01$). The average duration of consonants in the reversed-sonority cluster was the same as in the normally-sequenced clusters. However, the interaction between consonant type and syllable type, shown in Figure 3.5, was significant ($F(2,22) = 4.815, p < 0.05$).

As can be seen in the figure, there is an interesting difference in the duration relationship of consonants in the two types of clusters. In the normally-sequenced cluster, C1, the stop consonant, is longer than C2 ($F(1,22) = 21.277, p < 0.01$). In contrast, C1, the liquid, and C2 of the reversed-sonority cluster are of equal duration. The liquid is of the same duration in C1 position as when it occurs as a single onset, but is shorter in duration when it occurs as in C2 position (liquid CSV vs. SV+SCV ($F(1,22) = 23.690, p < 0.01$). On the other hand, the stop consonants are of equal duration in C1 and C2 position, but longer when they occur as single onsets (stop CV vs. CSV+SCV ($F(1,22) = 23.658, p < 0.01$). Single onset stops are also of longer duration than single liquid onsets ($F(1,22) = 16.623, p < 0.01$)

3.4.3 Acoustic duration as a function of jaw openness

In a final set of analyses, measures of jaw movement were compared with measures of acoustic duration. As a first analysis, total opening displacement and syllable duration was correlated. The results indicated a relatively good and highly significant correlation between these two variables (Pearson's $r = 0.402, p < 0.01$). Surprisingly, the correlation between jaw height and duration for vowels was relatively low and not significant (Pearson's $r = 0.154, NS$). This latter result may be due to the fact that jaw height measures were taken at the acoustic midpoint of the vowel and, at least in syllables with onset clusters, the acoustic midpoint of the vowel occurred well after the midpoint of the cycle as defined by the point of maximum jaw openness.

In a second analysis, the relative acoustic duration of the C1 consonant was measured as a function of the point of maximum jaw closure, which is the point that corresponds to the onset of the jaw cycle for the syllable. The percentage of C1 articulated within the cycle (post-closure) was established by subtracting the temporal point corresponding to minimum jaw opening from the temporal point of acoustic offset for the segment. The difference was divided by the total acoustic duration of the segment and multiplied by 100. A two-way ANOVA (syllable type, vowel type) indicated that the percentage of C1 articulated within the jaw cycle of the syllable token differed as a function of syllable type ($F(3,33) = 5.129, p < 0.01$) and vowel type ($F(2,22) = 16.839, p < 0.01$). When consonants occurred as single onsets, almost their total duration was articulated within the cycle. When consonants occurred in C1 position of an onset cluster, most of their total duration was articulated within the cycle, but substantially more of the consonants began being articulated before maximum closure. The following figures show example acoustic and movement waveforms for a CV (Figure 3.6), CSV (Figure 3.7), and SCV (Figure 3.8) syllable type.

The figure shows how closure for the consonant occurs almost simultaneously with jaw closure. Mean comparisons indicated, however, that consonantal closure often occurred slightly prior to jaw closure in CSV and SCV syllables compared

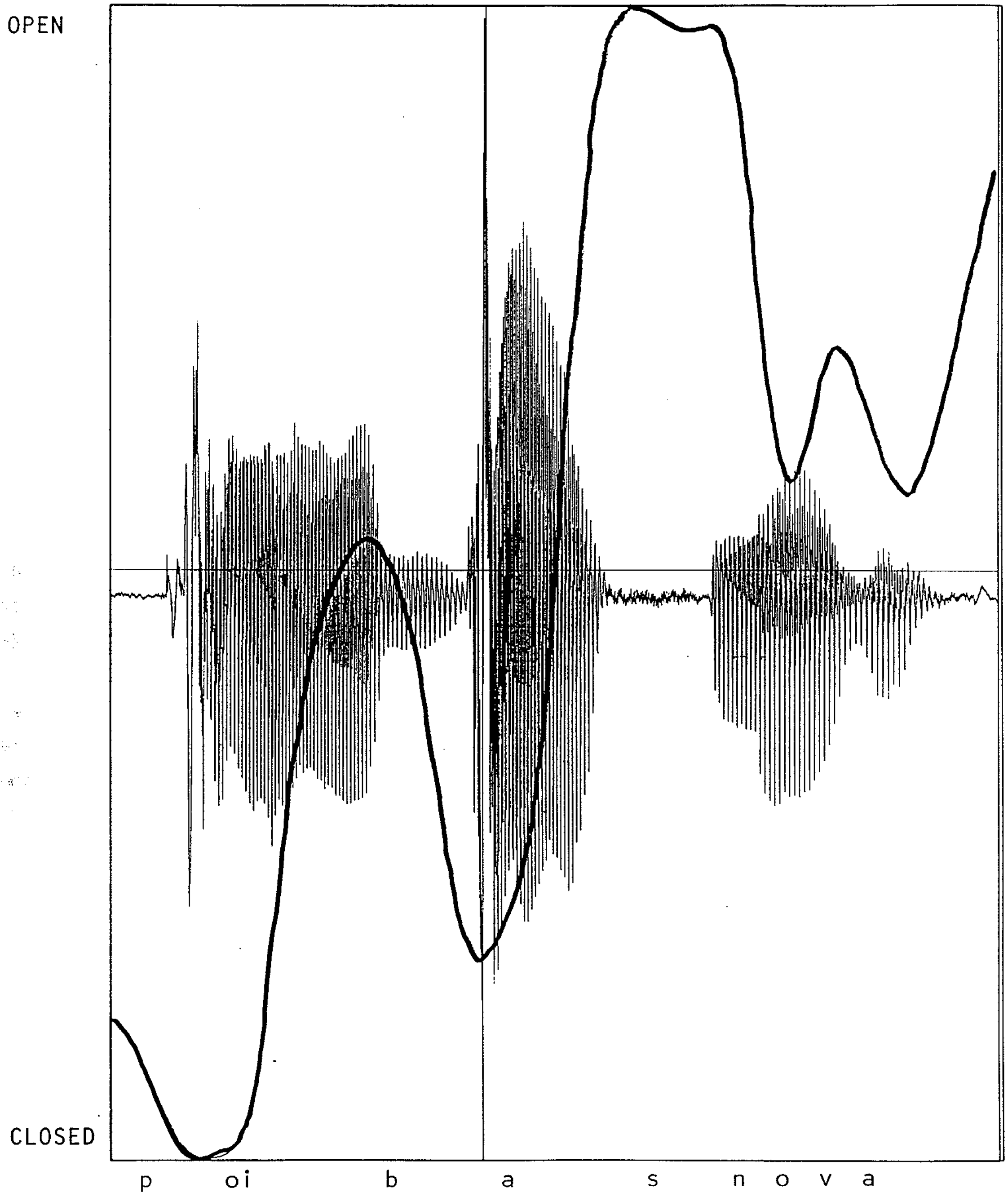


Figure 3.6: **Acoustic and movement waveforms for an example "ba" token in the frame sentence.** The onset of cycle for the token syllable is indicated at the first point of maximal jaw closure, which corresponds to the closure of the initial consonant.

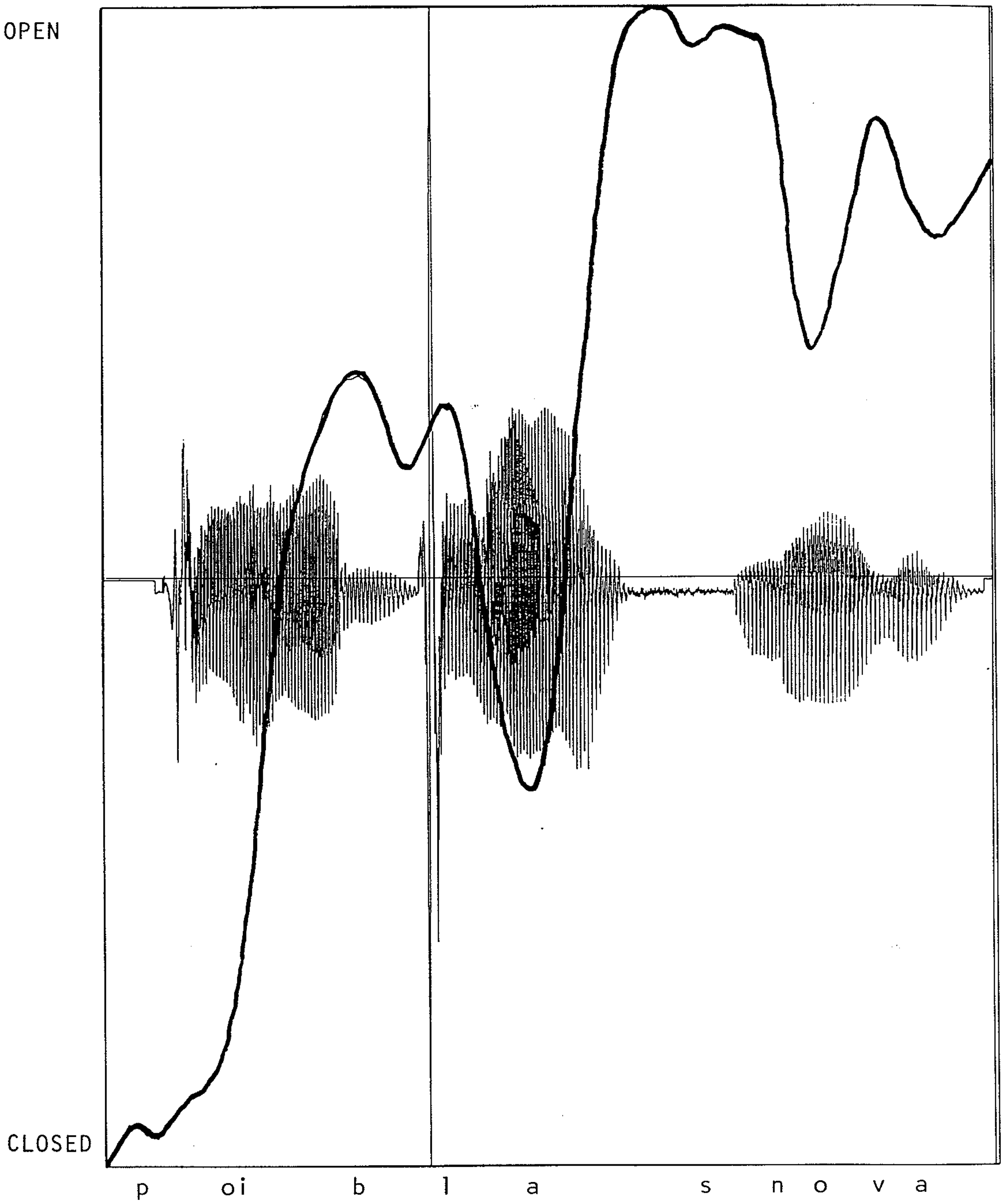


Figure 3.7: Acoustic and movement waveforms for an example "bla" token in the frame sentence. The onset of cycle for the token syllable is indicated at the first point of maximal jaw closure, which corresponds to the closure of the initial consonant.

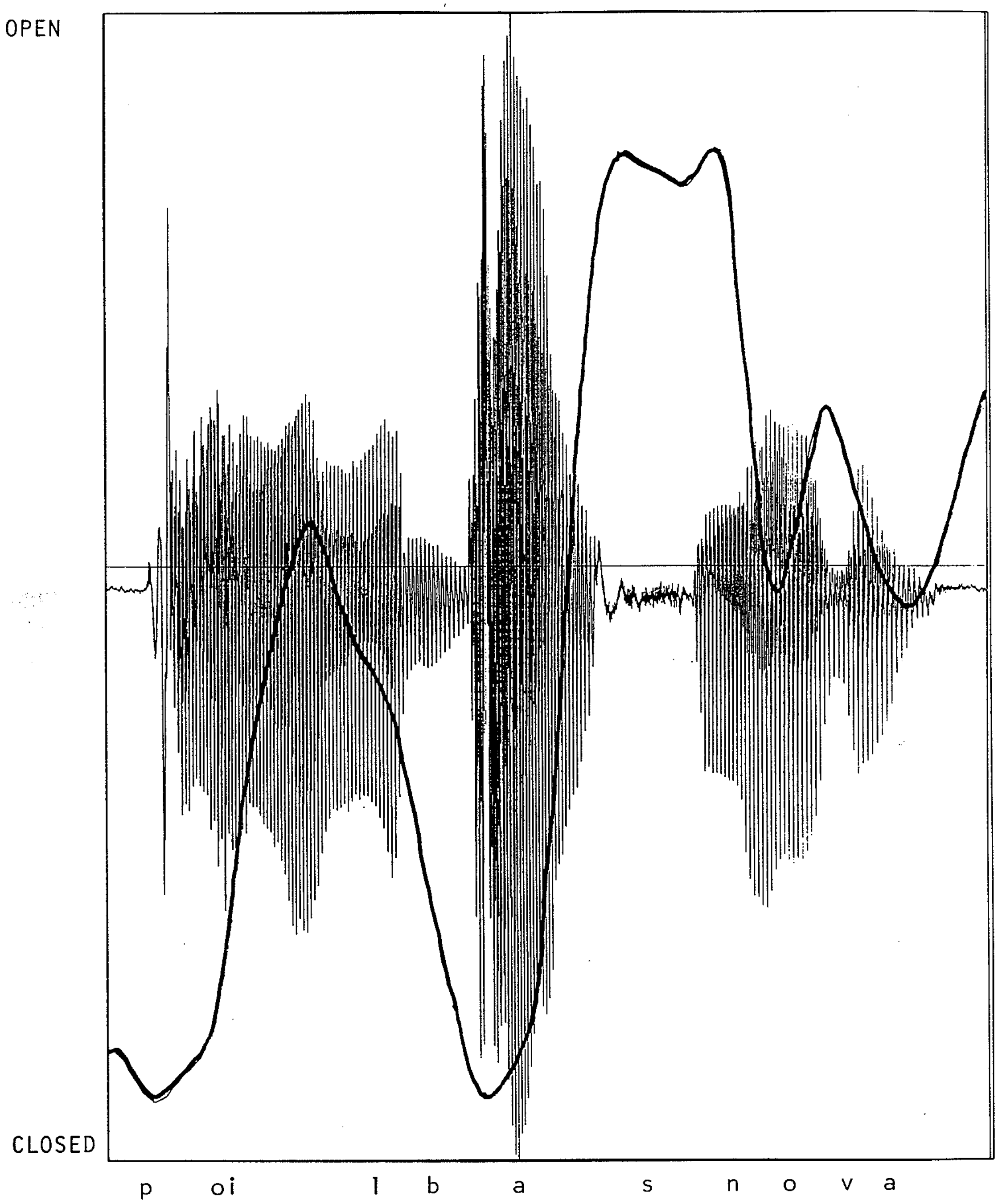


Figure 3.8: **Acoustic and movement waveforms for an example "lba" token in the frame sentence.** The onset of cycle for the token syllable is indicated at the first point of maximal jaw closure, which corresponds to the closure of the initial consonant.

with CV and SV syllables ($F(1,33) = 13.516, p < 0.01$).

The percentage of C1 articulated within the jaw cycle also varied as a function of vowel nucleus. When total jaw displacement for the opening portion of the cycle was smallest, as in the case of syllables with high-back vowel nuclei, more of C1 was articulated prior to maximum jaw closure, than when displacement was greater, as in the cases of syllables with low-central and high-front vowels. The effect of vowel type on percentage of consonant articulated pre- and post maximum jaw closure did not vary as a function of syllable type.

3.5 Discussion

The present study was conducted to establish whether jaw movement conforms to the sonority hierarchy or to its own basic cyclicity. In order to test between these two possibilities, jaw movement and acoustic data were recorded during the production of simple Russian syllables as well as during the production of Russian syllables with normally-sequenced and reversed-sonority onset clusters. Results from analyses of jaw movement and acoustic duration disconfirm the hypothesis that jaw movement is dictated by sonority characteristics of the segments. The highly sonorant consonant, [l], was articulated with less jaw opening than the less sonorant stops in the reversed-sonority onset clusters, [lbV] and [lgV]. Given the previously-discussed high correspondence between jaw opening and sonority, this result may at first appear counter-intuitive. If, however, one considers coarticulatory effects, the surprising result that reversed-sonority clusters obey the jaw cycle becomes less surprising. Keating, Lindblom, Lubker, and Kreiman (1994) measured jaw position during the articulation of English and Swedish inter-vocalic consonants as a function of differing vowel height. Their findings suggest that bilabial and velar stop consonants are more coarticulated with the preceding and following vowel than the lateral liquid. Thus, the present results from Russian are consistent with the Keating et al. results for English and Swedish. The difference in jaw position between C1 and C2 in the SCV syllables was due to the greater coarticulation of the stop with the following vowel. In contrast, when the liquid occurred in C2 position, it was produced with the same jaw height as when it was produced in C1 position. This result therefore indicates that segment type has some influence on jaw position. Overall, however, the jaw height measurements for the onset consonant(s) and vowel described a single jaw movement from a relatively closed position to a more open one. This overall result lends support to the hypothesis that the jaw provides a mechanical constraint on segmental articulation.

The jaw height results also suggest that the amplitude of the cycle may be determined by segment type. For instance, low vowels were articulated with more jaw opening than high vowels. Following Lindblom (1967), it was hypothesized that vowel driven changes in cycle amplitude would affect cycle duration due to the increased travel distance. This hypothesis was confirmed by a significant, positive correlation between syllable duration and maximum jaw opening. Such a result suggests that the jaw may also provide a temporal constraint on segment execution. In the same way that greater jaw opening leads to greater cycle duration, cycles of similar amplitude will be of similar duration. An increase in segment number within a single cycle should therefore lead to a decrease in segment duration not to a linear increase in cycle duration. This relationship between cycle amplitude, cycle duration, and segment duration received some support in this study. Stops were of greater duration when they occurred as single onsets than when they occurred as part of a cluster. Liquid consonants were of greater duration when they occurred as single onsets or in C1 position of clusters than when they occurred in C2 position of the cluster.

Thus far the results have been discussed with respect to the hypotheses on jaw movement. It has been noted that simple Russian syllables as well as those with normally-sequenced and reversed-sonority onset clusters roughly obey the mechanical and temporal constraint provided by the jaw cycle. But if all of these syllable types are similar in that they obey the jaw cycle, one might wonder why simple CV syllables are more common in languages than complex CSV or SCV syllables and why reversed-sonority (SCV) clusters are rare in Russian and in the world's languages. The first of these questions might be answered with the hypothesis proposed in the previous chapter, namely, that simple syllables are preferred over complex syllables because the former involve fewer articulations than the latter within a single jaw cycle. This hypothesis will be explored further in Chapters 4 and 5 when the relationship between acoustic duration and properties of the jaw cycle are examined in more detail. In order to answer the second question of why reversed-sonority clusters are rare, we may refer to the differences found in the present study between CSV and SCV syllables.

One major difference between the normally-sequenced and reversed-sonority clusters was in the duration relationship between the two consonants. In normally-sequenced clusters the second consonant was shorter in duration than the first. This type of relationship is consistent with what has been previously found for onset clusters (Haggard, 1973). In contrast, C1 and C2 were of equal duration in the reversed-sonority clusters. In the following chapter we will discuss evidence which suggests that a consonant sequence with this type of duration relationship would not be treated as an onset cluster by non-native listeners (see Chapter 4, section 4.5). Although it is unclear why C1 and C2 of a reversed-sonority cluster are of equal duration, if this different type of duration relationship has negative perceptual consequences, it would provide us with a reason for why these types of clusters are rare.

The fact that C1 and C2 are of equal duration in SCV syllables may shed some light on another difference observed between CSV and SCV syllables. When the vowel nucleus dictated a relatively shallower cycle the jaw movement associated with SCV syllables was generally of greater amplitude than the one associated with CSV syllables. The greater amplitude of SCV cycles may reflect the longer duration of the C2, stop consonant. Since C2 of the onset cluster is also articulated during the opening phase of the cycle, greater duration may translate into a longer opening movement, which would give rise to a larger cycle. If reversed-sonority syllables require greater jaw displacement than normally-sequenced syllables, it is possible that the preference for a certain sequence type might arise for reasons of articulatory ease. This relationship between displacement and articulatory ease is revisited in more detail in Chapter 5.

In summary, the data discussed in this chapter support the hypothesis that the jaw cycle functions as a mechanical and temporal constraint on segmental articulation. The goal of this dissertation is to show that this constraint may provide an articulatory basis for the syllable. Accordingly, the following chapters examine the effect of the jaw cycle constraint first in relation to syllable perception and then in relation to syllable structure.

Chapter 4

Syllable Production and Syllable Perception

Evidence was provided in the previous chapter to support the hypothesis that the jaw cycle functions as a mechanical and temporal constraint on segmental articulation. It was suggested that this constraint is realized in regular sound patterns that are associated with syllables. These sound patterns include the normal sequencing of phonemes within a syllable and patterns of relative segment duration. Although it may be clear from the previous exposition how the normal sequencing of phonemes may emerge from the constraint of the jaw cycle, it may be less clear how patterns of relative segment duration may emerge. In addition, the previous chapter assumed that these patterns relate to syllable perception without explaining how. In the present attempt to provide an articulatory basis for the syllable, it may be unnecessary to tackle the difficult question of how exactly syllable boundaries are perceived, but, as argued in Chapter 2, it is critical to show that the articulatory factor chosen relates to syllable perception and therefore to the perception of syllable boundaries. For this reason, this chapter is specifically aimed at establishing a relationship between the jaw cycle, the acoustic patterns defined in terms of syllables, and the perception of syllable boundaries.

4.1 Background

Segment duration is known to vary as a function of syllable position and/or syllable type (Lehiste, 1970; Klatt, 1976). For instance, initial consonants are longer than final consonants in monosyllables, provided that the syllable is not also phrase final (Hoard, 1966; Redford and Diehl, 1999). Consonants are also longer when they occur alone than as part of an onset cluster (Sigurd, 1973). Within an onset cluster, the first consonant tends to be longer than the second consonant (Haggard, 1973). Vowels also vary as a function of syllable position and syllable type. Vowels are longer in open syllables than in closed syllables (Maddieson, 1985), before a single consonant offset than before a consonant cluster (Lindblom and Rapp, 1973; Munhall, Fowler, Hawkins, Saltzman, 1992), in stressed than in unstressed syllables (Oller, 1973; Fant, Kruckenberg, Nord, 1991). The fact that segment duration varies in this manner suggests that relative segment duration could be used by listeners' to perceive syllables. Given that the duration of all segments varies as a function of position, it is possible that syllable perception is based on an analysis of duration relationships across multiple segments (Anderson and Port, 1994).

A question remains, however, as to why segment duration varies at all. One problem with answering this question, though, is that previous discussions of relative segment duration assume a syllable frame without defining one. For instance, the relative length of vowels before single consonants or consonant clusters has been attributed alternately to compensation (Kozhevnikov and Chistovich, 1965; Lehiste, 1977; Campbell and Isard, 1991) or to truncation (Munhall, Fowler, Hawkins, Saltzman, 1992; Harrington, Fletcher, Roberts, 1995). Compensation assumes that the overall duration of syllables is relatively constant and that this consistency may be achieved in part by control over the relative duration of each segment. Truncation assumes that the relative durations of consonants and vowels within a syllable emerge naturally from the temporal overlap of segmental gestures. While an undefined syllable is central to both views, the two views suggest different possibilities for defining the syllable. The argument that segments are subject to superordinate control might suggest that the syllable is an abstract, psychological unit of segment organization. The argument that coarticulatory effects influence the relative duration of segments may suggest an articulatory basis for defining a syllable frame. As before, this latter possibility is explored here with respect to the jaw cycle.

4.1.1 Segment duration and the jaw cycle

In order to begin to understand how the patterns of relative segment duration might emerge from the constraint of the jaw cycle, we can return to Lindblom's (1967) model of lip-mandible coordination. Lindblom hypothesized that the length of the opening phase in the jaw cycle, is a primary determinant of vowel duration in spite of the actions of the other articulators. This model was invoked in the previous chapter to explain the relationship between overall syllable duration and total jaw displacement during the opening phase of the cycle. The same model might explain differences in vowel duration that arise from differences in stress.

The fact that stressed syllables are longer than unstressed vowels (Oller, 1973; Fant, Kruckenberg, Nord, 1991) and articulated with more jaw opening (Stone, 1981; Erickson, Lenzo, Fujimura, 1994) suggests that a correlation between jaw opening and vowel duration could explain the different durations of stressed and unstressed vowels. Stress plays a significant role in syllabification. Stressed vowels "attract" more consonants in onset and offset position than unstressed vowels in the same environment (Treiman and Danis, 1988). This additional phonological/perceptual fact, though indirectly related to segment duration, may also arise from the interaction between the jaw cycle and segmental articulators. A larger cycle, specified by a stressed vowel, provides more physical and temporal "room" within which a consonant may be articulated. If the flanking syllables are unstressed, as they usually are, then adjacent cycles would be relatively shallow and would provide correspondingly less room for consonant articulation. Consonants may therefore be attracted to stressed vowels because they can be more easily articulated during their relatively longer and larger cycle.

The relative duration of consonants is also an important cue in syllable perception. Unlike for vowels, however, there does not seem to be any direct evidence linking the jaw cycle to duration for consonants. From the point of view of the jaw cycle as syllable frame, though, it is possible to understand how relative consonant duration may also be affected by the jaw cycle. To increase the number of consonants in a syllable is to increase the number of segments associated with the syllable. For example, CV, CVC, and CCVC syllables have two, three, and four segments respectively. If the overall amplitude and duration of the jaw cycle is constant across these different syllable types, then the different number of segments would need to be articulated in the same amount of time. This constraint would result in the relative compression of some, if not all, of the segments associated with a particular syllable. As previously noted, the idea of segment compression occurring to compensate for the relatively fixed duration of a syllable frame has been previously proposed (e.g., Kozhevnikov and Chistovich, 1965). Here, as elsewhere (Brodda, 1979, cited in Lindblom, 1983), compression is attributed to an articulatory constraint on segment production that is provided by the jaw cycle.

In addition to providing a constraint on segment duration, the jaw cycle also constrains the sequence of segmental articulation within the frame. This mechanical constraint was discussed at length in Chapter 3. Briefly, let us note that consonants that are usually internal members of a cluster (i.e., next to the vowel) are produced with relatively more jaw opening than consonants that are external members of a cluster. Accordingly, relative jaw height, though not an acoustic cue in itself, may help in establishing the location of syllable boundaries by signaling segment position, via segment type, within the frame.

4.1.2 Limits of the jaw cycle

Following from an articulatory definition of a syllable frame as a single jaw cycle, the beginning and ending of each syllable is defined by maximal jaw closure. Consequently, if the total duration of a particular consonant occurs immediately after maximal jaw closure, it should be in syllable-initial position, whereas if the total duration of the consonant occurs immediately before maximal jaw closure, it should be in syllable-final position. Nevertheless, since maximal jaw closure tends to occur only at one point in time and since segment articulation is achieved with articulators other than the jaw, there is no reason to believe that maximal jaw closure should, in itself, define an exact syllable boundary in perception.

One acoustic correlate of maximal jaw closure is a sharp decrease in the amplitude of a particular sound. This decrease can also be realized, however, by the action of the other articulators. Thus, the lips or tongue may achieve full closure for a stop consonant before or after the jaw does (Perkell, 1969; Gracco, 1988). In addition, the entire duration of a segment may need to be articulated with a relatively closed jaw, as in the case of /s/ (Keating, Lindblom, Lubker, Kreiman, 1994). In cases such as these, the duration of the consonant may be spread out across the boundary associated with maximal jaw closure since the jaw is relatively closed both before and after maximal jaw closure. This relative independence of the segmental articulators from the jaw cycle may provide a basis for ambiguity in the perception of syllable boundary location.

The inherent articulatory properties of the segments and the fact that the same acoustic effect may be achieved in various ways indicates that a one-to-one correspondence between syllable production and syllable perception may be impossible. But even if production does not map directly onto perception, it is certain that the two interact in speech and therefore syllable production should inform syllable perception. According to the present hypothesis that the articulatory frame provides a major constraint on segmental articulation and that this constraint is realized in acoustic patterns perceived as syllables, it should be possible to predict syllable perception, not only from acoustic patterns associated with the segments, but also from the jaw cycle.

4.2 Study outline

The present study uses stimuli in which segment content is manipulated at the boundary between two syllables to determine the effect of syllable production (as defined by the jaw cycle) on syllable perception. The stimuli were first created and used by Treiman, Gross, and Cwikiel-Glavin (1992) in their study of the syllabification of medial consonant sequences with initial /s/. Treiman et al. performed a series of syllable boundary judgment experiments using stimuli with medial /s/+stop and /s/+sonorant sequences in order to determine whether the Maximal Onset Principle (MOP) or the Sonority Sequencing Principle (SSP) could account for syllabification judgements in English. Since the MOP states that consonant sequences should be as long as possible in syllable-initial position, listeners might judge that both types of initial-/s/ sequences should remain intact. Since the SSP states that segments are organized within a syllable such that the most sonorous segments (liquids, glides, and nasals) are adjacent to the syllable nucleus (vowel) and the least sonorous segments are most peripheral to the nucleus (stops and fricatives), listeners might judge that only /s/+sonorant sequences should remain intact. Control stimuli were medial consonant sequences that formed legal English onset clusters (e.g., -dr-), which listeners should leave intact according to both the MOP and the SSP, and sequences that formed illegal English onset clusters (e.g., -gf-), which they should split according to the SSP. The results indicated that listeners were inclined to treat both types of initial-/s/ sequences in the same fashion, but in a manner that contradicted both the MOP and the SSP. Listeners tended to split both types of initial-/s/ sequences, as if they were medial clusters that formed illegal English onsets, while they left intact the sequences that formed legal onsets. Treiman et al. interpreted their results as support for the argument that pre-consonantal /s/ should be thought of as a phonological affix and therefore not really as part of an onset cluster (Kaye, Lowenstamm, Vergnaud, 1990).

To establish whether Treiman et al.'s (1992) results might be better explained by the effects of the jaw cycle on the articulation of segmental content, a replication of their basic findings was attempted. Treiman et al. used only one native English speaker per experiment, thus restricting the generality of their results. To allow for greater generality, the present study used four speakers. It was further reasoned that individual differences between speakers would contribute additional information on the effect of production on perception. Acoustic data and jaw movement data were recorded simultaneously for each of the speakers. Groups of participants were asked to listen to the nonsense words produced by each of the speakers and to break these words into syllables. Syllable boundary judgments were noted for the medial consonant sequences. Acoustic duration measures and measures of jaw openness were taken on the pre- and post-consonantal vowels as well as on each of the consonants of the sequence (i.e., V1, C1, C2, V2). Relative acoustic duration measures were examined for patterns that could be attributed to syllable perception. These measures also provided information on the inherent articulatory properties of the segments. Jaw openness measures were examined for patterns that could be related back to acoustic duration patterns. Since jaw openness measures are indicative of segment position within the frame, they also provided independent predictor variables of the effect of syllable production on syllable perception. The presence of a relationship between the variables and the degree to which they interacted with each other and with syllable perception was also tested. Overall, it was expected that the acoustic and articulatory correlates of the jaw cycle, rather than those associated with segments *per se*, would predict syllable perception as measured by syllable boundary judgments.

4.3 Method

4.3.1 Stimuli

The stimuli were 68 disyllabic nonsense words with medial, two-consonant sequences created and used by Treiman et al. (1992). The nonsense words consisted of four basic groups subdivided according to medial cluster type, three of which were legal onsets of English. In two of the four groups the first consonant of the sequence was /s/. In the first group /s/ was followed by a stop consonant (either /p/, /t/, or /k/) and in the second group /s/ was followed by a sonorant consonant (either /l/, /w/, /n/, or /m/). Examples of nonsense words with the /s/+stop medial consonants are "teskang" and "huspoit." Examples of nonsense words with /s/+sonorant medial consonants are "tesmang" and "husloit." Treiman et al.'s nomenclature for the various token types will also be used throughout this chapter. The tokens with /s/+stop medial consonants are referred to as Type 1 tokens, tokens with /s/+sonorant medial consonants as Type 2 tokens. The third group of nonsense words had medial consonant sequences in which the first consonant of the sequence was an obstruent consonant other than /s/ and the second consonant was a glide or liquid consonant. Examples of these types of nonsense words are "teblang" and "hudroit." These nonsense words are referred to as Type 3 tokens. The final group of nonsense words had medial consonants that were not legal onsets or offsets in English. Examples of these types of nonsense words include "chigfoon" and "vumroove." This last set is referred to as Type 5 tokens following Treiman et al.'s nomenclature. (Treiman et al.'s Type 4 tokens were not included.) A complete list of the nonsense words is provided in Table 4.1. The words are displayed in normal orthography because that is how they were presented to the speakers.

The speakers read the written form of the nonsense words in the frame sentence "Say ____ eight times." The order in which the nonsense words were read was randomized for each speaker. Since a primary aim of this study was to test the effect of production on perception, no instructions were given as to the correct pronunciation of the words. This meant that speakers were allowed to vary stress assignment according to their interpretation of the written word. Three speakers assigned stress to the second syllable regardless of token type. This pattern of stress assignment was expected due to the fact that the second syllable was usually associated with a dipthongized vowel nucleus and one or more final consonants. The fourth speaker (JH) unexpectedly assigned primary stress to the first syllable regardless of token type.

The sentences and the speakers' jaw movement during production were recorded.

Table 4.1: The nonsense word stimuli used in the present study borrowed from Treiman, Gross, and Cwikiel-Glavin (1992)

Type 1	Type 2	Type 3	Type 5
baspinge	baslinge	bapringe	chigfoon
chuskeem	chusleem	chupleem	dagmaste
daskeft	dasmeft	datreft	emlafe
destibe	deswibe	degribe	envorse
fiskanch	fismanch	fiklanch	gaktibe
geskint	geswelve	gefrelve	ganteeled
gespelve	geswint	gekrint	gethloove
huspoit	husloit	hudroit	hapkeet
jastoped	jasnoped	jadwoped	iptheen
kestibe	kesnibe	kethribe	lemgeeve
nuspeem	nuslange	nufleem	minveesh
nustange	nusweem	nuglange	monlave
teskang	tesmang	teblang	munleebs
vuspobe	vusluct	vuthwobe	obgorm
vuspuct	vusnobe	vutwuct	objenk
zeskib	zesmib	zegwib	vumroove
zusteeg	zusneeg	zushreeg	yadlorn

Jaw movement was sampled at 100 Hz. The sentences were recorded into a pentium PC using a waveform editor developed in the Speech Perception Laboratory with a Nakamichi CM700 microphone in conjunction with the waveform editor developed in the laboratory. The audio data was sampled at 11025 Hz. The root mean squared (RMS) amplitude of each stimulus sentence was normalized across talkers using the editor. RMS amplitude was calculated for the latter part of the frame sentence and then scaled so that the average amplitude of the entire utterance was consistent across all utterances.

4.3.2 Participants

Listeners were 4 groups of 12 (or 48) native American-English-speaking college students. They were instructed to listen to the nonsense words and asked to break them into syllables in writing on the provided response sheet. The listeners were given the example of the word "super" and asked how they would break it into syllables. The listeners inevitably responded that the syllables in "super" were "su" and "per." They were then instructed that if "super" were the nonsense word in question, they should write "su/per" on their response sheet.

In order to test the effect of production on perception and to avoid the possibility of listeners constructing their own word templates, four different groups of 12 participants each listened to the tokens produced by only one of the four speakers. All listeners were seated in a sound attenuated room and listened to the stimuli over earphones. The nonsense words were presented at normal hearing levels (around 65 dB) in a randomized sequence at interstimulus intervals of 3.5 seconds.

4.3.3 Measurements

Acoustic data

Segment durations were established for each of the consonants in the consonant sequence and for the vowel nuclei that preceded and followed the sequence using the visual display of the waveform editor as well as auditory judgments. The visual cues used to establish consonant boundaries were as follows. Fricative consonants were measured from the onset to the offset of fricative energy. Stop consonants were measured from the onset of closure, signified by a rapid amplitude decrease in the waveform, to the end of the release burst or aspiration. If a sequence of two stop consonants occurred and no release burst indicated the end of one and the beginning of another, the length of the closure duration for both consonants was established by dividing the entire duration of closure in half. The boundary between the more open consonants (nasals, liquids, and glides) and the vowels was signaled by changes in waveform characteristics as well as by a sudden drop (or rise) in amplitude. The demarcation of the boundary between these sonorant and glide consonants and the vowel were confirmed with auditory judgments. When surrounded by obstruents, vowels were measured from the onset to the offset of periodicity, evidenced by a sudden rise or fall in the amplitude of the waveform.

Jaw movement data

The speakers' jaw movement during production was recorded using two strain gauges attached to a depressor. The depressor was fixed under the speakers' chin by securing it to a light-weight head-mount, which speakers wore while producing the stimuli. The signal was sampled at 100 Hz. Movement calibration was achieved by recording the speaker with a clenched jaw and with a 1 cm spacer inserted between the premolars. The calibration recordings were made at the beginning and ending of each 15 minute recording sessions. Jaw openness data for the tokens was obtained by aligning the waveform display of jaw movement with the acoustic display of the stimulus token. Jaw openness measurements were then taken at the acoustic midpoint of the first and second vowel as well as at the acoustic midpoints of the medial consonants. The points of maximal jaw closure were established from the zero-crossing points on the velocity curve.

4.4 Results

Each of the data sets is first considered individually. Listeners' syllable boundary judgments on different token types for different speakers are presented. Acoustic duration patterns and patterns of jaw openness are also presented in relation to stimulus type and speaker. Next, the individual measurement sets are analyzed according to listeners' boundary judgments on the medial consonant sequences. The relationship between boundary judgments, relative segment duration, and relative position in the jaw cycle is then examined. The presence of a relationship between the two physical measurement sets is tested with correlations and then each of the physical measurement sets is used as independent predictor variables for boundary judgments. Finally, the interaction between the jaw cycle, segmental content, and syllable perception is examined by looking at the acoustic duration of C1 in relation to maximal jaw closure and as a function of syllable boundary judgments.

4.4.1 Syllable boundary judgments

Listeners' syllable boundary judgments were analyzed by examining how they treated the first consonant (C1) of the medial consonant sequences. If listeners judged that C1 belonged to the first syllable of the nonsense word, C1 was labeled "1". If C1 was judged to belong to the second syllable, it was labeled "2". A C1 consonant that was put into both syllables was labeled "1.5". The second consonant (C2) of the sequence was not analyzed because it almost always was judged to belong to the second. [1]

It was found that listeners syllabified the consonant sequences differently depending on token type. The first consonants in Type 1 and Type 2 tokens, the /s/+stop and /s/+sonorant sequences, were often associated with the first syllable of the nonsense word. The average judgments for the stimuli produced by the four speakers was that /s/ belonged to the first syllable 61.75% of the time in

---

[1] C2 was judged to belong to the second syllable 98% of the time. It was judged to belong to the first syllable 1% of the time and not perceived 1% of the time.

Table 4.2: Average judgment of C1's syllable association ("1" indicates syllable 1, "2" indicates syllable 2).

Token Types	Speaker JH	Speaker LR	Speaker MM	Speaker MT
1 (/s/+stop)	1.04	1.68	1.47	1.45
2 (/s/+son)	1.05	1.70	1.38	1.47
3 (legal onset)	1.22	1.91	1.71	1.77
5 (illegal onset)	1.01	1.04	1.02	1.04

Type 1 stimuli and 63.5% in Type 2 stimuli. These results can be compared with judgments concerning the other token types. The first consonant in Type 3 tokens, where consonant sequences formed legal English onsets (e.g., -dr-), were not usually associated with the first syllable (35.5%), whereas C1 in Type 5 tokens, where consonant sequences formed illegal English onsets (e.g., -gf-), were almost always associated with the first syllable (97.25%). These differences in syllable boundary judgments for the various token types are in agreement with the Treiman et al. (1992) results. As in Treiman et al., the differences were significant in a two-way (token type, speaker) analysis of variance ($F(3,45) = 72.46; p < 0.001$).

In addition to the main effect of token type on syllable boundary judgments, there was also a main effect of speaker ($F(3,45) = 75.77; p < 0.001$). Participants listening to different speakers made different judgments on the same stimuli regarding the association of C1 to the first syllable. Boundary judgments for different speakers also differed as a function of token type ($F(9,135) = 9.39, p < 0.001$). Table 4.2 displays average syllable association judgments for C1 as a function of the different token types. Scores closer to "2" indicate that the majority of 12 listeners judged that most of the examples of C1 in a particular medial consonant sequence belonged to the second syllable. Scores closer to "1" indicate that the examples of C1 were usually judged to belong to the first syllable. Scores closer to "1.5" indicate that either listeners did not agree on which syllable the examples of C1 should belong to or the same listener assigned the consonant to both syllables.

As can be seen from Table 4.2 the syllabification judgments elicited by speaker JH and LR were considerably different. Participants who listened to speaker JH almost always associated C1 with syllable 1 regardless of token type. In contrast, participants who Listened to speaker LR almost always associated C1 with the following syllable and not with syllable 1. Type 1 and Type 2 tokens produced by speakers MM and MT elicited the most ambiguous judgments by listeners. The ambiguity was largely due to listener disagreement as to where to assign C1 of the medial consonant sequences. Ambisyllabic judgments (where C1 was assigned to both the first and second syllable) accounted for 3% of the judgments for tokens produced by speaker MM and 17.6% of the judgments for tokens produced by speaker MT.

In the following section the phonetic basis for the different patterns of syllable boundary judgments will be considered.

4.4.2 Relative acoustic duration and jaw openness of the segments

A three-way (token type, speaker, segment) analysis of variance indicated that duration varied significantly as a function of token type ($F(3,45) = 3.14; p < 0.05$), speaker ($F(3,45) = 206.12; p < 0.001$), and segment type ($F(3,45) = 53.97; p < 0.001$). A similar analysis indicated that jaw height also differed significantly for each of the three main variables. Some main effects pointed to absolute differences that were probably not meaningfully related to syllable boundary judgments. For example, the overall duration for the four segments of Type 3 tokens was shorter than for the other token types. Other examples include the result that speakers JH and MM produced the longest and shortest tokens respectively, while speakers JH and LR produced tokens with more open jaw configurations than the other two speakers. Some obvious expectations were also confirmed. For example, vowels are articulated with a more open configuration than consonants.

Of greater relevance to the prediction of syllable boundary judgments are the significant interactions that occurred between the variables. These interactions are explored below for each of the two measurement types.

Acoustic duration patterns

In general, the first vowel, which corresponded to the vowel nucleus of the first syllable, was shorter in duration than the second vowel, which corresponded to the vowel nucleus of the second syllable. Conversely, the first consonant of the consonant sequence was generally of longer duration than the second consonant in the sequence. The relative duration of the various segments interacted, however, with token type ($F(9,135) = 4.97; p < 0.001$). Figure 4.1 displays this interaction.

One difference between token types emerges when the overall duration of the different consonant sequences is considered. If the durations of the two consonants in the sequence are summed, then it can be seen that sequences with initial /s/ were on average 28 milliseconds longer than the other sequences [Type 1 ($CC = 252.8$ *msec*),

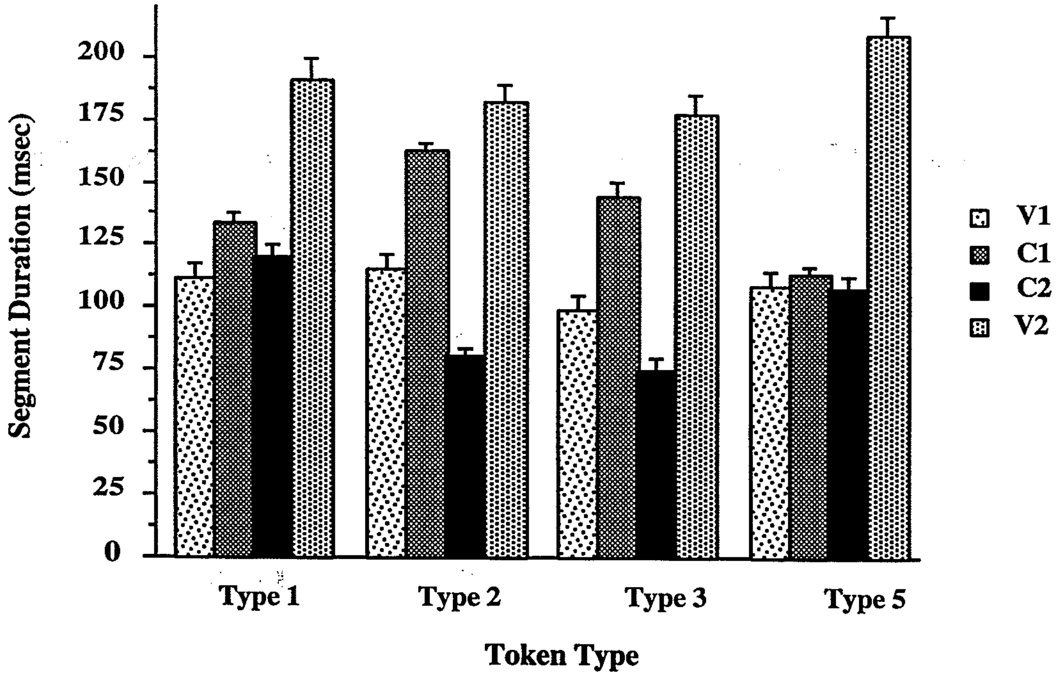


Figure 4.1: **Relative acoustic duration of the segments as a function of different token types.** Acoustic duration is shown in milliseconds for V1, C1, C2, and V2 for /s/+stop (Type 1), /s/+sonorant (Type 2), legal English onset (Type 3), and illegal English onset (Type 5) consonant sequences.

Type 2 ($CC = 241$ *msec*), Type 3 ($CC = 218.5$ *msec*), Type 5 ($CC = 219.5$ *msec*)]. Another difference between the token types was the relative duration of C1 and C2. In accordance with the general pattern of C1 and C2 duration, C1 was relatively longer than C2 in Type 2 tokens (medial /s/+sonorant sequences) and in Type 3 tokens (legal English onsets). In contrast, when the medial sequence formed an illegal English onset (Type 5) or when it was a /s/+stop sequence (Type 1), C1 and C2 were equal in duration (no significant differences were found).

Table 4.3: Ratio of the mean durations of the first to the second vowel and the first to the second consonant as a function of whether C1 was judged to belong to syllable 1 or syllable 2.

	C1 in Syllable 1			C1 in Syllable 2		
Token Types	V1 : V2	C1 : C2	N	V1 : V2	C1 : C2	N
1 (/s/+stop)	1 : 1.5	1 : 1.0	36	1 : 2.1	1 : 0.8	32
2 (/s/+son)	1 : 1.3	1 : 0.6	40	1 : 2.1	1 : 0.4	28
3 (legal onset)	1 : 1.3	1 : 0.9	21	1 : 2.3	1 : 0.4	47
5 (illegal onset)	1 : 1.9	1 : 1.0	68	na	na	0

If the relative duration of C1 and C2 are considered not only by token type, but also by listeners' syllable boundary judgments, then a principled pattern emerges for the relative duration of the various segment types. Table 4.3 displays the ratio of the mean durations of V1 to V2 and C1 to C2 as a function of token type and syllable boundary judgments.

Examination of Table 4.3 reveals a difference between tokens in which listeners split the medial consonant sequences and those in which listeners left them intact to form a coherent onset for the second syllable. When the consonants were judged to form an onset cluster, both the vowels and the consonants differed more in relative duration than when the medial consonants were split into single consonantal offset/onset by the listener. The direction of these differences corresponded with the relative duration pattern of the legal (Type 3) and illegal (Type 5) onsets. When the consonant sequence was perceived as an onset cluster, V2 was (much) longer than V1 and C2 was (much) shorter than C1.

The significant interactions between speaker and token type ($F(9,135) = 4.12; p < 0.001$) and speaker and segment type ($F(9,135) = 16.03; p < 0.001$) provided additional evidence for a connection between token production and syllable perception. Like the pattern of syllable boundary judgments, the pattern of relative segment duration was most different for speakers JH and LR. The pattern of relative vowel duration produced by speaker JH differed from the pattern produced by the other speakers. Speaker JH produced V1 and V2 with the same duration for Type 1 and 2 stimuli, whereas other speakers generally produced V1 with shorter duration than V2 for all stimuli types including Type 1 and 2. The stimuli produced by speakers JH and LR also differed in relative consonant duration. Individual paired, 2-tailed t-tests show that the medial consonants in Type 1, and 3 stimuli were pro-

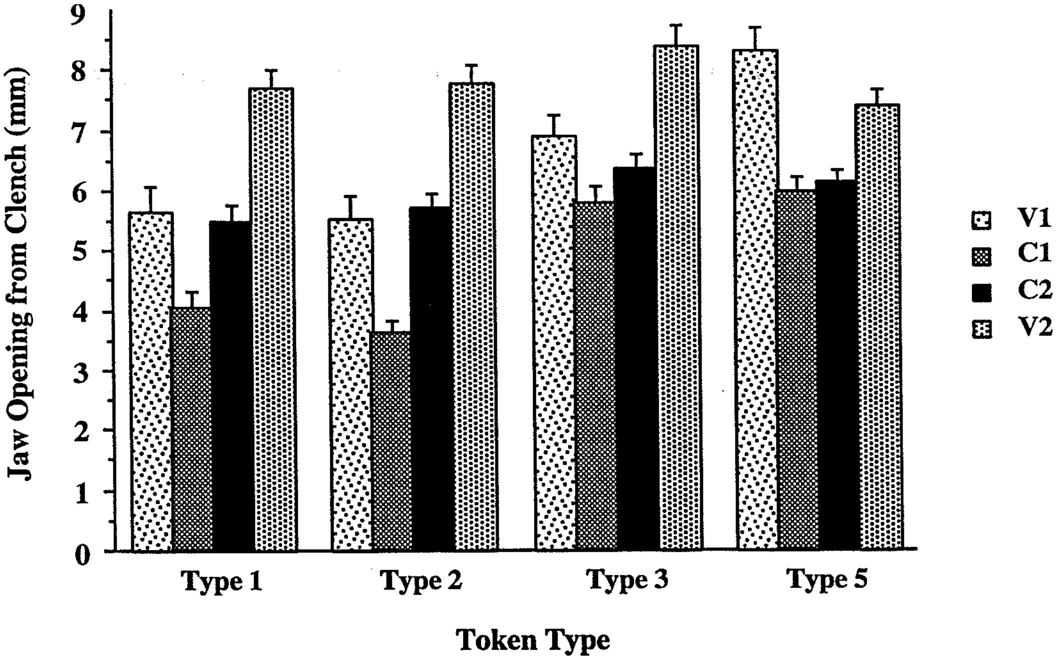


Figure 4.2: **Relative jaw openness of the segments as a function of different token types.** Jaw openness is shown in millimeters for V1, C1, C2, and V2 for /s/+stop (Type 1), /s/+sonorant (Type 2), legal English onset (Type 3), and illegal English onset (Type 5) consonant sequences.

duced with the same duration by speaker JH, whereas C1 was of significantly shorter duration than C2 in Type 1 and 3 stimuli produced by speaker LR. These differences between JH and LR are consistent with the fact that the medial clusters in most of tokens produced by JH were split, while most of those produced by LR were left intact.

Patterns of jaw openness

A significant interaction between token type and segment ($F(9, 144) = 14.51, p < 0.001$) revealed some interesting differences in the pattern of jaw displacement for the various token types. Figure 4.2 displays the different relative jaw openings at which the segments were articulated for different token types.

As shown in Figure 4.2, the vowel nucleus of the first syllable (V1) was articulated with less jaw opening than the vowel nucleus of the second syllable (V2) in Types 1 (/s/+stop), 2 (/s/+son), and 3 (legal onsets), but was articulated

Table 4.4: Ratio of the mean jaw openings of the first to the second vowel and the first to the second consonant as a function of whether C1 was judged to belong to syllable 1 or syllable 2.

	C1 in Syllable 1			C1 in Syllable 2		
Token Types	V1 : V2	C1 : C2	N	V1 : V2	C1 : C2	N
1 (/s/+stop)	1 : 1.1	1 : 1.3	36	1 : 1.8	1 : 1.5	32
2 (/s/+son)	1 : 1.2	1 : 1.5	40	1 : 1.8	1 : 1.7	28
3 (legal onset)	1 : 1.0	1 : 1.0	21	1 : 1.4	1 : 1.2	47
5 (illegal onset)	1 : 0.9	1 : 1.0	68	na	na	0

with more jaw opening in Type 5 (illegal onsets). Mean comparisons confirm these observations [Type 1 ($F(1,144) = 60.03, p < 0.001$); Type 2 ($F(1,144) = 72.87, p < 0.001$); Type 3 ($F(1,144) = 32.49, p < 0.001$); Type 5 ($F(1,144) = 12.28, p < 0.001$)]. The relative jaw opening related to consonant articulation also differed as a function of token type. C1 or /s/ was articulated with less jaw opening than C2 in stimuli Types 1 and 2 [Type 1 ($F(1,144) = 30.76, p < 0.001$); Type 2 ($F(1,144) = 65.74, p < 0.001$)]. C1 was also articulated with less jaw opening than C2 in Type 3 stimuli, but C1 and C2 did not differ in jaw height in Type 5 stimuli [Type 3 ($F(1,144) = 4.26, p < 0.05$); Type 5 ($F(1,144) = 0.216, NS$)].

As with acoustic duration, when the relative jaw opening for different segments is considered as a function of syllable boundary judgments, a principled pattern emerges. Table 4.4 shows the ratio of the mean displacements of V1 to V2 and C1 to C2 as a function of token type and syllable boundary judgment.

Table 4.4 shows that when the consonants were judged to form an onset cluster the relative difference in jaw opening between the first and second vowel was greater than when the consonants were split by the listeners. Similarly, the relative difference in jaw opening between the first and second consonants was greater for sequences that were judged to form an onset cluster. The direction of these differences was in accordance with the general pattern described above. V2 and C2 were produced with more jaw opening than V1 and C1.

A significant three-way interaction indicated that the general pattern of jaw openness described above was not consistent across speakers ($F(27,432) = 2.629, p < 0.001$). Speaker JH produced V1 with equal or greater jaw opening than V2 in every stimulus token regardless of type. This pattern of production is consistent with the fact that listeners almost always split the consonant sequences in the tokens produced by speaker JH.

4.4.3 Relationship between the measurement variables

Jaw displacement and acoustic duration

A Pearson correlation between jaw opening and acoustic duration revealed that these variables were significantly (in a 2-tailed test) and positively correlated for V1 ($r = 0.599, p < 0.001$), C2 ($r = 0.197, p < 0.001$), and V2 ($r = 0.125, p < 0.05$), but not for C1. Inter-correlations between the variables indicated that the relative jaw opening of V1 also predicted C2 acoustic duration ($r = 0.401, p < 0.001$), but not the acoustic duration of C1 or V2. This pattern of inter-correlation suggested that if the first vowel was longer (and therefore maybe perceived as stressed), the C1 of the medial sequence was produced as a final consonant of the first syllable and C2 was produced as a single onset to the second syllable. As a single onset, C2 would be longer than when it occurred as the second consonant of an onset cluster.

Acoustic duration and jaw openness as predictors of syllable boundary location

A statistical classification procedure was used to predict the average syllable boundary judgments for all the tokens on the basis of the different measurement sets. Two categories of syllable boundary judgments were designated. Category 1 encompassed all tokens in which C1 was assigned by the majority of listeners to the first syllable. Category 2 encompassed all other tokens. More specifically, if the average C1 judgment score for a particular token was below 1.5, the token was assigned to Category 1. Conversely, if the average C1 judgment score was 1.5 or above, the token was assigned to Category 2. Discriminant analyses using the duration values and the displacement values for V1, C1, C2, and V2 had similar success in accurately classifying the various tokens. Table 4.5 displays the standardized coefficients and correlations of the discriminant function.

Examination of the coefficients and correlations shown in Table 4.5 indicate that values associated with multiple segments were important in determining group membership. Discriminant analysis that used the duration values correctly classified C1 as belonging to either the first or second syllable 75.8% of the time. The discriminant analysis that used jaw openness values correctly classified the tokens 76.1% of the time.

Table 4.5: The standardized discriminant function coefficients and the correlations between the discriminating variables and the function are displayed for the two sets of variables.

Acoustic duration measures			Measures of jaw openness		
Segment	Coefficient	Correlation	Segment	Coefficient	Correlation
V1	0.626	0.674	V1	1.288	0.899
C1	-0.463	-0.390	C1	-0.148	0.512
C2	0.520	0.745	C2	-0.040	0.302
V2	0.104	0.099	V2	-0.452	0.153

Table 4.6: Mean percent of C1 articulated prior to complete jaw closure as a function of whether C1 was judged to belong to syllable 1 or syllable 2.

	C1 in Syllable 1		C1 in Syllable 2	
Token Types	Percent	N	Percent	N
1 (/s/+stop)	70.1	36	49.1	32
2 (/s/+son)	42.6	40	34.3	28
3 (legal onset)	100	21	62.3	47
5 (illegal onset)	96.9	68	na	0

Syllable boundary judgment, jaw openness, and acoustic duration

As noted in the introduction, the point of maximal jaw closure can be used as the boundary for the new jaw cycle. It might therefore be expected that syllable boundary judgments would correspond with whether a segment was produced prior to or after this point. To test this idea, the point of maximal jaw closure corresponding to the initial boundary of the second syllable (when the jaw cycle defines the articulatory syllable frame) was determined for each token. The relative duration pre- and post-closure was established for the first consonant in the consonant sequence. Table 4.6 shows the mean percentage of different C1s produced before maximal jaw closure as a function of syllable boundary judgment.

Table 4.6 shows that more of C1 was articulated prior to closure when C1 was judged to belong to the first syllable than when C1 was judged to belong to the second syllable. The relative percentage of consonant duration pre-maximal jaw closure for the different token types is informative from both an articulatory and a perceptual perspective. The relatively long duration of /s/ combined with the fact that this segment is articulated with a relatively closed jaw throughout its duration, may account for why it is not entirely articulated before or after maximal jaw closure. The perceptual salience of /s/ combined with its inherent acoustic duration and articulatory characteristics may also provide a reason for why boundary judgments are particularly ambiguous for consonant sequences beginning with /s/.

The presumed perceptual effect of the relative duration of /s/ pre- and post-maximal jaw closure can be contrasted with the presumed perceptual effect of relative duration pre- and post-maximal jaw closure for the C1s of the legal clusters. A large proportion of C1 of the intact legal onsets were articulated pre-maximal jaw closure. The reason that this may not have affected boundary judgments is because approximately half of these consonants were stop consonants whose predominant place-of-articulation cues – release bursts and transitions – would occur later into the opening phase of the jaw cycle. The assumption here is that the phonemes must be identified before the listener calculates a syllable boundary. The prediction this assumption makes is that the release burst of a stop must occur before maximal jaw closure in order for the phoneme to be perceived as a final consonant. This prediction is supported by the fact that C1 consonants, which were not /s/, were perceived as syllable final only when articulated in their entirety before maximal jaw closure (Table 4.6).

4.5 Discussion

The present study was motivated by the hypothesis that relative segment duration, an acoustic cue that informs syllable perception, emerges from the articulatory constraint of the jaw cycle on the movement of the segmental articulators. Given that jaw movement only constrains, and does not determine the movement of the other articulators, it was considered unlikely that syllable production, defined by the cyclic movement of the jaw, would map directly onto syllable perception. Nevertheless, it was reasoned that if jaw movement underlies syllable-linked patterns of relative acoustic duration, then measurements associated with the jaw cycle should predict syllable perception just as relative segment duration predicts syllable perception. The data from this study not only provided evidence for the existence of a relationship between segment duration and the jaw cycle, but also showed that syllable perception could be predicted equally well by either measurement set.

4.5.1 Syllable boundary judgments and the physical measures

On average the pattern of syllable boundary judgments obtained in the present study were similar to those of Treiman et al. (1992). Listeners split the medial sequence with initial /s/ significantly more often than the consonant sequences that formed legal English onsets (e.g., -dr-), but significantly less often than those that formed illegal English onsets (e.g., -gf-). Whereas Treiman et al. (1992) used only one speaker, the present study with multiple speakers. Although on average listeners' judgments in this study were found to match those of Treiman et al., large differences in syllabification behavior were observed when the same stimuli were produced by different speakers. It might be concluded from the acoustic and articulatory data presented in this study that almost all of the observed differences in syllabification behavior can be explained in terms of differences in syllable production. Such a conclusion would undermine the phonological explanation advanced by Treiman and colleagues for the different syllabifications of legal /s/ clusters in English compared with that of other legal English clusters. The specific differences in acoustic and articulatory patterns related to specific syllable boundary judgments are discussed below first in terms of relative acoustic duration then in terms of relative jaw opening.

Examination of the relative segment duration pattern for different token types revealed that different relative durations of the first and second syllable nucleus (V1 vs. V2) and of the first and second medial consonant (C1 vs. C2) were associated with different syllable boundary judgments. When the first consonant in the medial consonant sequence was treated as the final consonant of the first syllable, the relative duration of the two vowels and two consonants differed less than when the consonants were judged to form an onset cluster to the second syllable. When the consonant sequence was left intact, the duration of V2 was at least twice that of V1 and the duration of C2 was often only half that of C1. The emergence of these particular patterns in relation to syllable boundary judgments conforms to findings from previous research. For example, given the relationship between stress and vowel duration, the observed correspondence between relative vowel length and syllable boundary judgments is consistent with Treiman and Danis' (1988) finding that more consonants are associated with stressed vowels than with unstressed vowels in syllables. In addition, the fact that long-short consonant sequences are treated as onset clusters, while consonants that differ less in duration are split, is consistent with Haggard's (1973) finding that consonant durations of prevocalic clusters are negatively correlated.

The patterns that emerged from the jaw opening data mirror, in some respects, the patterns found for relative segment duration. The relative difference in jaw opening associated with the production of V1 compared to V2 and C1 compared to C2 was greater for sequences that were left intact by the listeners than for those that were split. The specifics of the pattern associated with the production of vowels appears to be directly related to the duration pattern for vowels. When the consonant sequence was judged to form an onset cluster to the second syllable, V2 was articulated with much greater jaw opening than V1. Jaw opening has been correlated with duration in vowels (Lindblom, 1967) and with stress (Stone, 1981; Erickson, Lenzo, Fujimura, 1994). In the present study, jaw opening was also found to correlate positively with vowel duration. This particular finding is consistent with the hypothesis that a larger cycle will attract more consonants since they are more easily articulated within its boundaries.

The jaw opening pattern for consonants was less obviously related to the duration pattern for consonants. The relative degree of jaw opening was greater for C2 compared with C1 when the medial consonant sequences were left intact than when they were split. The data from this study do, however, support a speculative link between the two patterns. The fact that C2 is articulated with more jaw opening relative to C1 in onset clusters may indicate that C2 is being articulated closer to the jaw target of the following vowel, or further into the jaw cycle, than when it functions as a single consonant onset. If C2 is articulated further into the jaw cycle it may be abbreviated because the jaw will reach its target configuration for the following vowel more quickly. This suggestion is supported by the analysis in which C1 duration was examined as a function of maximal jaw closure and syllable boundary judgment. It was found that when C1 was judged to belong to syllable 1, more of the consonant was articulated before maximal jaw closure than when C1 was judged to belong to syllable 2. Maximal jaw closure indicates the point at which the next opening phase begins. If C1 is being articulated during a greater portion of this phase, then C2 will have less room within the cycle, and consequently less time, in which to be articulated.

Overall, the results from this study suggest that the variation in segment duration associated with syllable perception may arise from an articulatory frame defined by the jaw cycle. Larger frames, associated with stressed vowel targets, are able to incorporate more segments than the smaller frames of unstressed vowel targets. Sequential consonant articulations must be articulated during the relatively more closed portions of the cycle. Accordingly, multiple consonants within a single frame will be compressed into these portions of the cycle. The duration of the vowel will also be impacted by the addition of multiple segments in the frame. A stressed vowel in a CV syllable can be articulated during almost the entire duration of the cycle. In contrast, a stressed vowel in a CVCC syllable will be necessarily be "truncated" by the articulation of the final consonant sequence.

4.5.2 On the relationship between the production and perception of syllables

A hypothesis that posits a relationship between production factors and perceptual judgments on syllables is not new. For example, both Boucher (1988) and Tuller and Kelso (1991) show that production factors can influence the acoustic pattern in such a way as to affect judgments of syllable type (CV vs. VC). What is new, however, is the formulation of a production-based definition of the syllable that is meant to account for the relative segment duration patterns identified with the syllable. In addition, a syllable definition based primarily on the jaw cycle has the virtue of being able to predict syllable boundary judgments *a priori*. The discriminant analyses results indicated that position within the cycle, as defined by relative jaw opening, was as good a predictor of syllable boundary location as relative segment duration. The statistical classifier was able to correctly classify C1 as belonging to the first or second syllable 75.8% of the time given the duration measures and 76.1% of the time given the relative jaw opening measures. In light of the disagreement between listeners regarding boundary judgments, particularly for consonant sequences with initial /s/, the predictability of the boundaries was remarkably high.

According to the present view, syllable boundary location is exactly specified in the jaw cycle by maximal jaw closure. It is in this more exacting interpretation of the syllable frame that the link between syllable perception and syllable production becomes more obviously indirect. The point of maximal jaw closure, as syllable boundary, cannot be linked to perception through acoustic duration cues because it has no real duration. Maximal jaw closure may however be linked to syllable perception by its effect on amplitude: maximal jaw closure would correspond to points of minimum amplitude in the waveform. But even though changes in amplitude inform syllabification behavior (Price, 1980; Ladefoged, 1993), it is clear that these changes may also be realized by the actions of the other articulators. In the extreme case of glottal articulations (as in [hahaha]), syllable perception will be completely decoupled from the jaw cycle. However, since most of speech is produced supraglottally, the decreases in amplitude wrought by lip and tongue configurations for consonants will still be related to the positioning of the jaw, albeit indirectly. It is also possible that the relative positioning of the jaw *per se* may be partially signaled in the waveform associated with consonant production. This possibility is supported in the present chapter by the relationship shown to exist between segment duration pre- and post-maximal jaw closure and syllable boundary judgments. Overall, though, the fact that segment duration often crosses the boundary of maximal jaw closure demonstrates not only the indirect link between syllable perception and production, but also provides a possible reason for why syllable boundary judgments frequently differ between subjects.

In sum, the data discussed in this chapter support the hypothesis that the jaw cycle constrains the articulation of segments in a manner that produces acoustic patterns associated with syllable perception. Measures of jaw height were found to correlate with measures of segment duration. Additionally, it was demonstrated that measures of jaw height were as highly predictive of syllable boundary judgments as measures of relative segment duration. Insofar as variation in syllable boundary judgments is one of the main problems associated with syllable definition, this latter finding is especially compelling evidence for the idea that the jaw cycle provides an articulatory basis for the syllable.

Chapter 5

Syllable Production and Syllable Structure

Thus far, evidence has been presented to support the hypothesis that the jaw cycle constrains segmental articulation and that this constraint results in acoustic patterns associated with syllable perception. Specifically, the previous two chapters have focused on explaining segment sequencing constraints, patterns of relative segment duration, and syllable boundary perception in terms of the mechanical and temporal constraint of the jaw cycle. Accordingly, we have gone a long way to defining a syllable in articulatory terms and in agreement with the three requirements for this type of definition given in Chapter 2. These requirements were that an articulatory definition of the syllable should (1) account for the sequential organization of segments into syllable-sized units, (2) relate to syllable perception, and (3) provide a basis for cross-language preferences in syllable structure. While the first two of these requirements were addressed in Chapter 3 and 4 respectively, the third requirement has yet to be addressed. Hence, the present chapter focuses on whether properties of the cycle may account for the notable cross-language preferences in syllable structure, namely, the preference for initial consonants over final consonants and single-consonant onsets and offsets over clusters.

5.1 Background

Three generalizations emerged from Bell and Hooper's (1978) summary of cross-language preferences in syllable systems (Chapter 2). The last of these had to do with the normal organization of segments within a syllable. It has been argued that this organization can be characterized by a jaw openness hierarchy and can be explained to emerge from the continuous open-close movement of the jaw during speech (Chapter 3). The first two generalizations were related to syllable structure. There is a cross-language preference for syllable-initial consonants over syllable-final consonants and for single consonants over consonant clusters. These two generalizations are also interrelated in that there exists a cross-language preference for initial clusters over final clusters. This overall phonological pattern is captured by the Maximal Onset Principle, which states that consonants should behave as onsets instead of offsets even if this means that clusters are formed.

One reason that initial consonants are preferred over final consonants may have to do with the relative perceptual distinctiveness of initial and final consonants. Redford and Diehl (1999) showed that there is a perceptual advantage for initial consonants over final consonants. Some evidence suggests that this advantage may stem from production factors. Syllable-initial consonants are less reduced and less variable on a variety of articulatory and acoustic measures than final consonants

(Byrd and Tan, 1996; Sussman, Bessell, Dalston, Majors, 1997; Redford and Diehl, 1999). The production of consonant clusters appears to be similarly affected by syllable position. Byrd (1996) examined articulatory overlap and variability in the onset cluster [sk-] and the offset clusters [-sk], [-ks] and [-gd]. She found that, with the exception of the offset cluster [-sk], articulatory overlap and variability was greater in final position than in initial position. While these studies provide evidence that consonants in syllable-initial and final position are produced differently, they do not provide any insight into why this is so.

From the point of view of sequential segmental articulation, it is difficult to understand why consonants that occur in different syllable positions should be produced differently especially if the immediate segmental environment is identical in both cases. For instance, it might be expected that the movement gestures necessary to go from one articulation to the next in the nonsense word *ta.tat* in the frame sentence "____ ought to go" would be the same for the second and third occurrence of the alveolar stop given that both occurrences are preceded and followed by the same low, central vowel, but they are not. One recurrent difference between the two syllable positions is that initial consonants are longer than final consonants in running speech (e.g., Hoard, 1966; Malecot, 1968; Henderson and Repp, 1982; Anderson and Port, 1992; Redford and Diehl, 1999). This difference is interesting because it parallels a similar difference in the phases of the jaw cycle. The opening (or initial) phase of the jaw cycle is generally longer than the closing (or final) phase (Kelso, Vatikiotis-Bateson, Saltzman, Kay, 1985). On the view that the jaw cycle provides a frame within which segmental content is articulated, the parallel difference between consonant duration and phase duration is meaningful because segments pertaining to a single syllable are usually articulated within the boundaries of a single open-close jaw cycle (Chapter 4). Initial consonants are articulated during the opening phase of the jaw cycle and final consonants during the closing phase (vowels span part of the opening and part of the closing phase).

Other, more consistent, differences between the opening and closing phase of the jaw cycle may provide an articulatory basis for why final consonants are more reduced and more variable than initial consonants. The peak velocity of the closing phase is greater than the opening phase (e.g., Sussman, MacNeilage, Hanson, 1973; Kuehn and Moll, 1976; Gracco, 1994), as is the slope of the relationship between peak velocity and peak displacement (Kelso, Vatikiotis-Bateson, Saltzman, Kay, 1985; Ostry and Munhall, 1985). Given the coordination between the movement paths of the segmental articulators and that of the jaw, it should not be surprising that increased jaw speed in a particular direction is related to a corresponding increase in the speed of movement of the segmental articulators (Gracco, 1988). Thus, the differences observed in the production of syllable-initial and syllable-final consonants may result from a trading relationship between articulatory speed and accuracy, where increases in speed lead to decreases in accuracy. This type of relationship has been previously postulated and empirically validated by Lindblom (1996; Moon and Lindblom, 1994) in reference to different speaking rates.

It is not clear, however, whether the differences between the opening and closing phases of the jaw cycle are due to dynamic properties of the cycle, as argued by Kelso et al. (1985) and Gracco (1994), or to differences in phonetic content associated with the frame. Studies that have reported differences between the two phases have based their observations on productions of open syllables, that is, syllables without final consonants [1]. It is possible that the opening phase of a CV syllable could be longer and slower than the closing phase because the consonant and part of the vowel are articulated during the opening phase, while only a part of the vowel is articulated during the closing phase. In order to test whether inherent differences in the jaw cycle may influence the articulation of consonants at different points in the cycle, it is necessary to compare the opening and closing phases of CVC syllables in identical segmental contexts. This comparison is made in the first part of the present study. A variety of segmental measurements are made and it was expected that these would show that initial and final consonants are produced differently, but within a single jaw cycle. Measurements on the opening and closing phases associated with the production of initial and final consonants were expected to show that differences in duration and velocity between the phases are movement properties of the jaw cycle and not attributable to differences in phonetic content.

In addition to comparing jaw movement associated with the production of simple consonantal onsets and offsets, the present study also compares jaw movement associated with the production of onset and offset clusters. Measurements of segment duration, segment jaw height, phase duration, phase velocity and phase displacement were made for CCVC and CVCC syllables. It was expected that if the movement properties of the jaw cycle influence the articulation of single consonants in initial and final syllable position, then these same properties should be evident during the articulation of consonant clusters. Specifically, the closing phase of the jaw cycle was expected to be of shorter duration and greater velocity than the opening phase in cases of identical phonetic content. The jaw cycle should, however, also reflect differences between simple and complex syllables since the absolute number of segments associated with a single frame differs in the two cases. In cases where

---

[1] An exception to this is Kuehn and Moll (1976), who examined closed, phrase-final syllables.

more segments are articulated within a single frame, the amplitude of the cycle may increase to accommodate the articulation of additional segments. The previously discussed correlation between jaw opening and segment duration (Lindblom, 1967; Munhall, Fowler, Hawkins, Saltzman, 1992, see also Chapters 3 and 4) suggests that if the amplitude of the cycle increases then so too should the duration. It was therefore expected that cycle duration would be greater during the articulation of complex syllables than simple syllables.

Finally, if a trading relationship between articulatory speed and accuracy exists and the articulatory measurements show that speed increases in final position relative to initial position and in cycles with clusters relative to those without, then articulatory accuracy should decrease in parallel. Since articulatory accuracy is thought to affect segment identifiability, then the relative identifiability of simple onsets should be greater than complex onsets and simple offsets and the relative identifiability of simple offsets should be greater than complex offsets. A perceptual confusion experiment is conducted in the second part of this study to test this possibility.

5.2 Methods

5.2.1 Speakers and recordings

The stimuli were single syllable nonsense words of the form CV[ʃ], [ʃ]VC, [s]CV[ʃ], and [ʃ]VC[s], where C was one of the three English voiceless stops [p, t, k] and V was one of the three point vowels [i, u, a]. Stimuli with [s] as a single initial or final consonant were also part of the set, but were not analyzed as they created an imbalance between the number of simple and complex stimuli. The stimuli set therefore consisted of mirror image tokens, for example, the mirror image of "poosh" was "shoop" and the mirror image of "spoosh" was "shoops."

Two male and two female American-English speakers produced each of the nonsense words twice in the frame sentence "Sign ___ nine times". The sentences and the speakers' jaw movement during production were recorded simultaneously using a program developed in the Speech Perception Laboratory at a sampling rate of 11025 Hz. The sentences were recorded with a Nakamichi CM700 microphone. The speaker's jaw movement was recorded using two strain gauges attached to a depressor. The depressor was fixed under the speaker's chin by securing it to a light-weight head-mount, which the speaker wore while producing the stimuli. The signal was sampled at 100 Hz. Movement calibration was achieved by recording the speaker with a clenched jaw and with a 1cm spacer inserted between the premolars.

Calibration recordings were made at 5 minute intervals.

5.2.2 Measurements

The acoustic and jaw movement data were analyzed using a waveform editor developed in the Speech Perception Laboratory.

Acoustic data. Segment durations were established for each of the stop consonants and for the vowel nuclei using the visual display of the waveform editor as well as auditory judgments. The visual cues used to establish consonant boundaries were as follows. Stop consonants were measured from the onset of closure, signified either by a rapid amplitude decrease in the waveform or the offset of fricative energy (associated with a preceding [s]), to the end of the release burst or aspiration or the beginning of fricative energy (associated with a following [s]).

Jaw movement data. Jaw openness data for the tokens was obtained by aligning the waveform display of jaw movement with the acoustic display of the stimulus token. Jaw openness measurements were then taken at the acoustic midpoint of the first and second vowel as well as at the acoustic midpoints of the medial consonants. The points of minimum and maximum jaw opening, which corresponded to the zero-crossing points on the velocity curve, were also recorded. Overall displacement values and duration values for the opening and closing phase were established from these points. Peak velocity was also measured for the opening and closing phase.

5.3 Results

To ensure that the onsets and offsets of the syllables in question were articulated within the limits of the cycle, the relative jaw opening of the consonants was analyzed in relation to the points of minimum and maximum jaw opening. Next duration differences were examined for the onsets and offset segments and then compared with differences in opening and closing phase duration. Opening and closing phase differences in displacement and peak velocity were also analyzed. A final section examined the relationships between the measurement variables.

5.3.1 Consonant position within the cycle

The opening and closing phases were delimited by the points of minimum and maximum jaw opening. The relative jaw height of the consonant segments were compared with these points. The top panel of Figure 5.1 shows relative jaw opening for the stop consonants in simple syllables in relation to the points of minimum and maximum opening. A 3-way (speaker, syllable position, measurement point) ANOVA was conducted to test for differences in jaw opening between the various points within the cycle. Mean comparisons indicated that the differences between stop consonant opening and minimum and maximum jaw opening were statistically significant [minimum opening vs. stop opening ($F(1,34) = 10.54, p < 0.01$), stop opening vs. maximum opening ($F(1,34) = 42.75, p < 0.01$)].

The bottom panel of Figure 5.1 shows relative jaw opening for each of the consonants in the onset or offset clusters in relation to the points of minimum and maximum opening. A 3-way (speaker, syllable position, measurement point) ANOVA was conducted to test for differences in jaw opening between the various points within the cycle. Mean comparisons indicated that the fricative consonant, always the outermost consonant of the cluster, was not significantly different in degree of opening from the point of minimum jaw opening. The stop consonant, or inner member of the cluster, was significantly more open than the fricative ($F(1,51) = 9.34, p < 0.01$), but significantly more closed than the point of maximum jaw opening ($F(1,51) = 42.75, p < 0.01$). As with simple syllables, there were no significant differences in jaw opening as a function of syllable position.

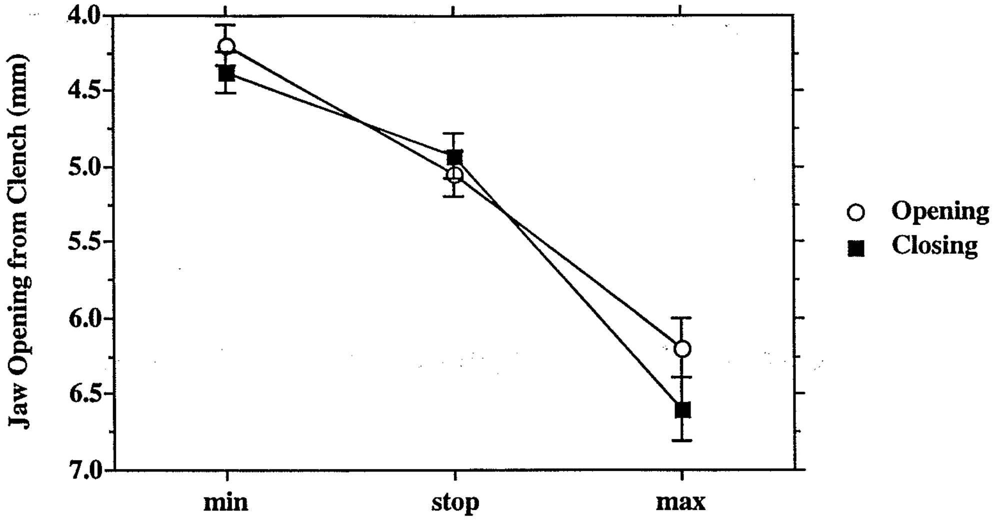
Although significant interactions occurred between syllable position and the measurement points in both syllable types, the relationship between the consonant(s) and the points of minimum and maximum jaw closure remained the same regardless of whether the consonant(s) occurred in syllable-initial or final position.

This section demonstrates that the onset and offset consonants of the token syllables were produced within the confines of the opening and closing phases of a cycle as defined by the points of minimum and maximum jaw opening.

5.3.2 Segment and phase duration

Segment duration differences were tested in a 4-way (speaker, syllable type, syllable position, segment type) ANOVA. The factor "segment type" included measures of total duration for the relevant onset or offset consonant(s) in addition to total duration of the vowel. Thus, in the CV[ʃ] or [ʃ]VC syllables, segment duration equaled C+V. In the CCV[ʃ] or [ʃ]VCC syllables, segment duration equaled (C+C)+V. The vowel was included in the segment duration measures since it is known to vary in duration as a function of offset complexity.

Significant main effects were in keeping with expectations. Onsets + vowel were longer than offsets + vowel ($F(1,17) = 95.92, p < 0.01$) and clusters + vowel were longer than single consonants + vowel ($F(1,17) = 463.89, p < 0.01$). The interaction between syllable position and syllable type, shown in Figure 5.2, was



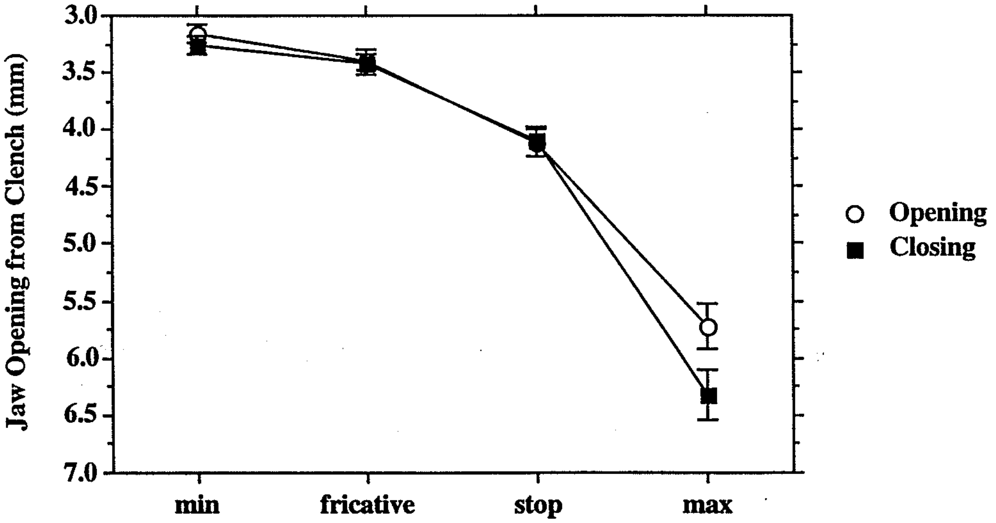


Figure 5.1: **Relative placement of the consonants in the opening or closing phase of the cycle.** The relative jaw opening (in millimeters) of the single stop onset and offset (top) and of the fricative-stop onset cluster and stop-fricative offset cluster (bottom) are shown in relation to the points of minimum and maximum jaw opening, which delimit the opening and closing phases of the cycle. These data are collapsed across the three vowel types.

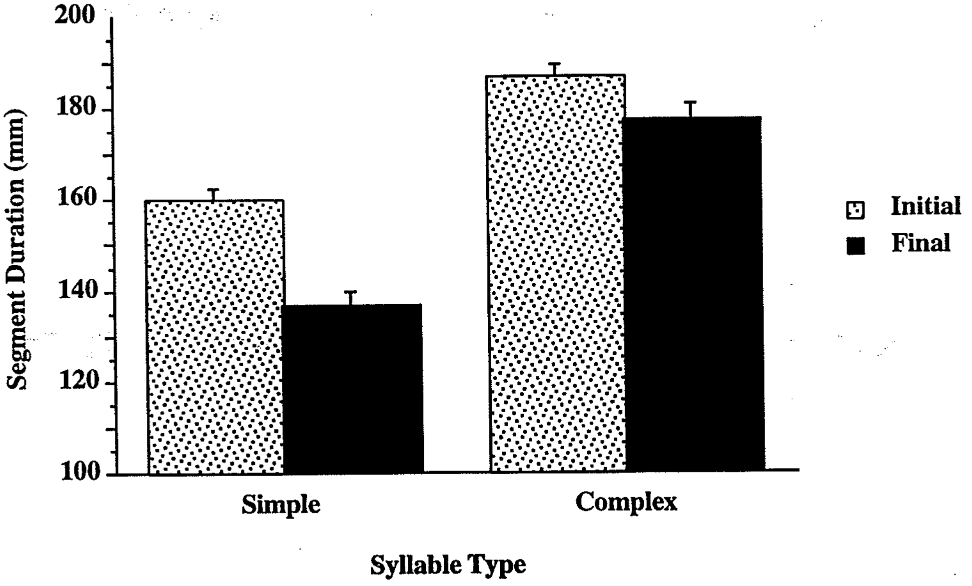


Figure 5.2: **Mean relative duration of initial and final segments as a function of syllable type.** The mean duration (in milliseconds) of the consonant(s) and vowel is shown for syllables with initial or final stop consonants (C) or initial and final fricative-stop/stop-fricative clusters (CC).

also significant ($F(1,17) = 15.28, p < 0.01$). In spite of this interaction, mean comparisons indicated that onset segments were longer than the offset segments in both simple and complex syllable types. The pattern of segment duration, shown in the figure, was true for 3 of the 4 speakers, but the onset and offset segments produced by speaker MM were equal in duration. This lack of difference in duration between onset and offset segments may have been due to the fact that speaker MM inserted an audible pause of 372 milliseconds (+/- 83 msec) between the offset of the stimulus token and the onset of the following word in the frame sentence. Thus, speaker MM produced syllable offsets that were also phrase-final and so were probably subjected to phrase-final lengthening.

The three-way interaction between syllable position, syllable type, and segment type was also significant ($F(1,17) = 59.49, p < 0.01$). Examination of the duration of different segments in different syllable positions and different syllable types provided a more detailed picture of the source of duration differences than what is shown in Figure 5.2. Mean comparisons indicated that the observed position difference in duration for simple syllables was due primarily to a reduction in duration of final consonants ($F(1,17) = 87.20, p < 0.01$) and secondarily to a reduction in duration of vowels associated with final consonants ($F(1,17) = 47.19, p < 0.01$). In contrast, the position difference in duration observed for clusters was entirely due to a reduction in duration of vowels associated with complex offsets ($F(1,17) = 95.74, p < 0.01$). The offset clusters were actually greater in duration than the onset clusters ($F(1,17) = 10.07, p < 0.01$).

It was expected that the differences in the opening and closing phases of the jaw cycle would parallel the differences found for segments. Specifically, the opening phase of the cycle was expected to be longer than the closing phase. A significant positive Pearson's correlation between segment duration and phase duration ($r = 0.335, p < 0.01$) and between consonant duration and phase duration ($r = 0.357, p < 0.01$) suggested that the anticipated relationship between the segment and phase variables was real. In a 3-way (speaker, syllable type, phase type) ANOVA, however, the difference between the opening and closing phases was not significant. A significant interaction between speaker and syllable position ($F(3,51) = 24.89, p < 0.01$) indicated, though, that the tokens produced by 2 speakers agreed with the expectation – for these speakers opening phases were longer than closing phases. Figure 5.3 displays the duration of the opening and closing phases of the jaw cycle as a function of the speaker for simple (top) and complex (bottom) syllables.

As shown in Figure 5.3, 3 of the 4 speakers produced opening and closing phases in a consistently different manner regardless of syllable type. Speakers MB and LR produced tokens in which the opening phase was significantly longer than the closing phase [MB ($F(1,51) = 21.96, p < 0.01$), LR ($F(1,51) = 9.57, p < 0.01$)]. Speaker MC produced tokens in which the phases were of equal duration. And, as before, the tokens produced by speaker MM were the most different from those produced by the other speakers: speaker MM produced tokens in which the closing phase was longer than the opening phase ($F(1,51) = 43.56, p < 0.01$). Even though the tokens produced by speaker MM are being considered here with the tokens produced by the other speakers, it is relevant to the discussion to point out, once again, that speaker MM produced offsets that were also phrase-final. Although the data from this speaker contradict those from the other speakers, they are internally consistent. Final consonants produced by speaker MM were longer than initial consonants.

Figure 5.3 also provides information regarding the significant difference be-

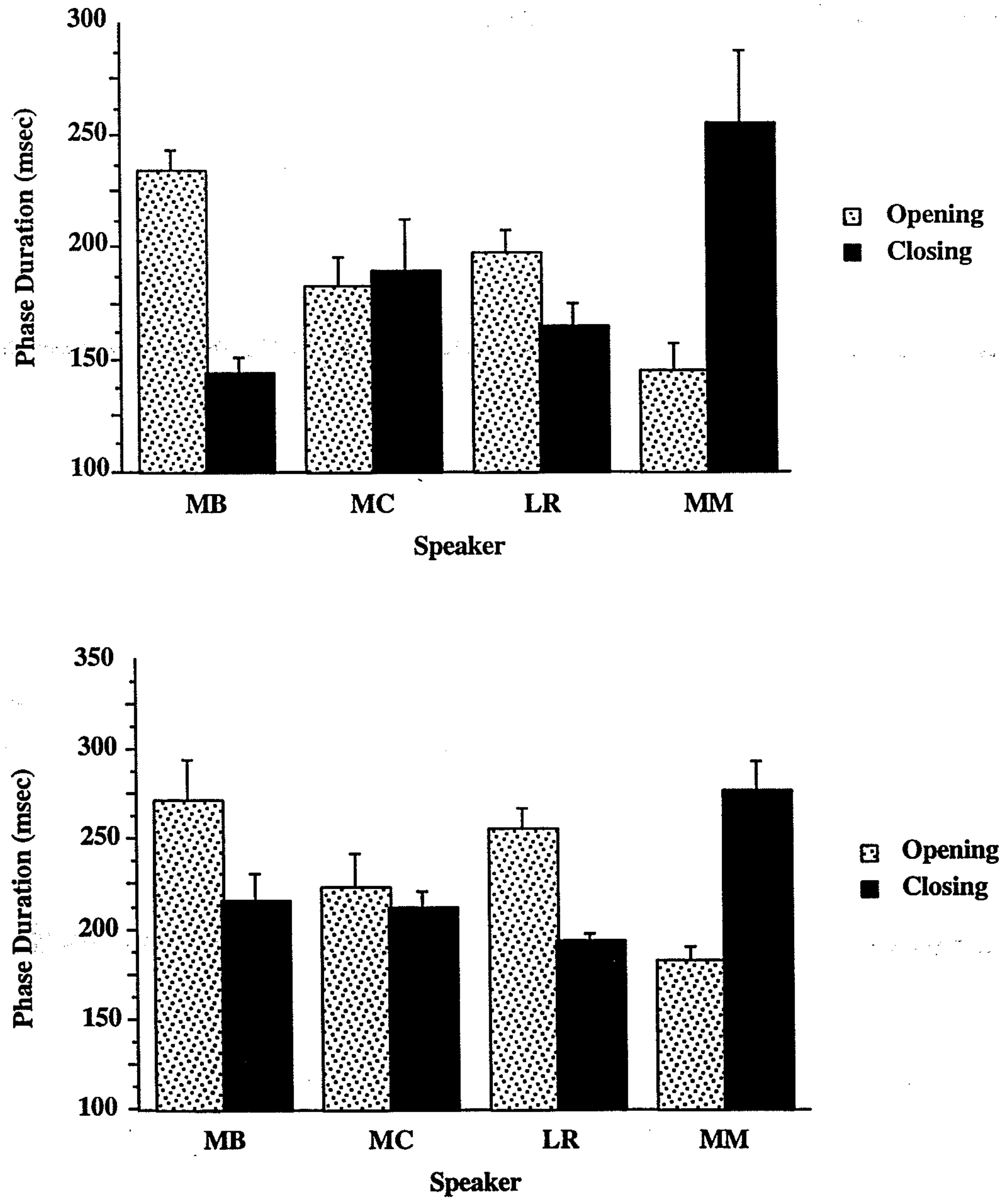


Figure 5.3: **Opening and closing phase duration as a function of speaker.** The total duration (in milliseconds) of the opening and closing phases is shown as a function of speaker for simple, CVC syllables (top) and complex, CCVC and CVCC syllables (bottom).

tween simple and complex syllables ($F(1, 17) = 20.75, p < 0.01$). Both phases of the jaw cycle were longer when a consonant cluster was articulated during the phase than when a single consonant was articulated. This result was expected and is consistent with the finding that the total duration of clusters+vowel was greater than the total duration of single consonant+vowel.

In sum, significant correlations between segment duration and phase duration as well as the individual duration results for segments and jaw movement confirmed that a relationship exists between segment duration and phase duration. Individual segments in syllable-final position are usually of shorter duration than syllable-initial segments just as the closing phase of the jaw cycle is often of shorter duration than the opening phase. The admittedly imperfect nature of the relationship between segment duration and phase duration may be due, in part, to the fact that, the three or four segments of the simple and complex syllables could not be neatly divided in half and mapped onto a single phase of the jaw cycle.

5.3.3 Phase displacement and peak velocity

Displacement

The distance between minimum and maximum opening for a particular phase was used as the displacement value for that phase. A 3-way (speaker, syllable type, phase type) ANOVA showed that displacement was greater in phases associated with complex syllables than in those associated with simple syllables ($F(1, 17) = 69.58, p < 0.01$) and that displacement was greater in closing phases (offsets) than in opening phases (onsets) ($F(1, 17) = 7.34, p < 0.05$). Even though there was no significant interaction between these two factors, mean comparisons indicated that the difference in opening and closing displacement was not statistically significant for phases associated with simple syllables. The differences between opening and closing phases as a function of syllable type can be seen in Figure 5.4.

Overall the extent of displacement was greater in closing phases than in opening phases during the articulation of offset and onset clusters respectively. However, as with opening and closing phase duration, speakers varied in the extent to which they produced the phases of their jaw cycle with different degrees of displacement. Mean comparisons indicated that 3 of the 4 speakers followed a similar pattern for complex syllables. The closing phases of syllables with offset clusters were produced with more displacement than the opening phases of syllables with onset clusters in these speakers. Speaker MM also produced closing phases of simple syllables with

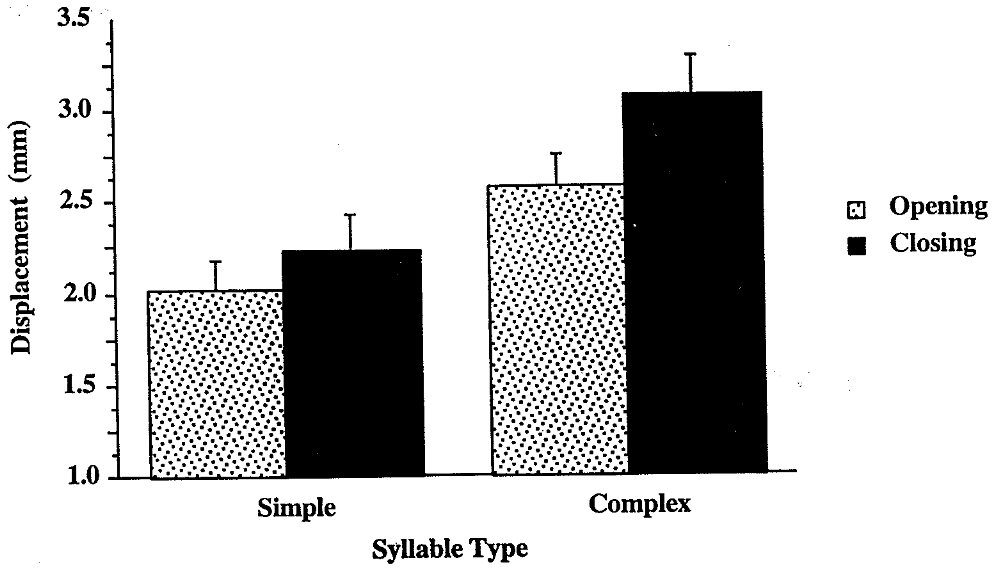


Figure 5.4: **Displacement of the opening and closing phases as a function of syllable type.** The total displacement (in millimeters) of the opening and closing phases is shown as a function of syllable type. The two types are simple (CVC) and complex (CCVC or CVCC).

more displacement than the corresponding opening phases.

Peak velocity

The peak velocity of the phases also differed as a function of phase type and syllable type. A 3-way (speaker, syllable type, phase type) ANOVA showed that peak velocity was greater in phases associated with complex syllables than in those associated with simple syllables ($F(1,17) = 56.48, p < 0.01$) and that peak velocity was greater during the closing phases (for final consonants) than during the opening phases (for initial consonants) ($F(1,17) = 8.71, p < 0.01$). The interaction between these two factors, though not significant, is seen in Figure 5.5.

As with all the articulatory measures discussed thus far, speakers varied in the extent to which they produced the phases of the cycle with different peak velocities. Overall, however, all speakers produced closing phases with greater velocity than opening phases and phases associated with a complex onset/offset with greater velocity than those associated with a simple onset/offset.

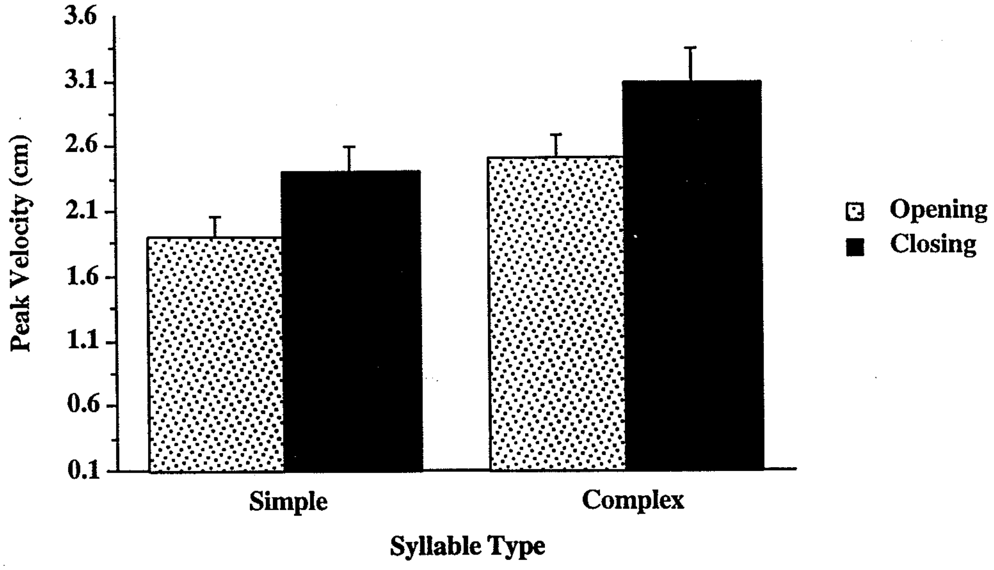


Figure 5.5: **Peak velocity of the opening and closing phases as a function of syllable type.** The peak velocity (in centimeters squared) of the opening and closing phases is shown as a function of syllable type. The two types are simple (CVC) and complex (CCVC or CVCC).

This section shows a parallel between displacement and peak velocity such that the jaw cycle generally travels further and faster in closing phases relative to opening phases and in phases associated with clusters relative to those associated with single consonants.

5.3.4 Relationship between duration, displacement, and velocity

Previous work suggests that segment and cycle duration is positively correlated with jaw displacement (Lindblom, 1967; Munhall, Fowler, Hawkins, Saltzman, 1992; see also Chapters 3 and 4 of this dissertation). The results from the previous sections of the present study indicate that phase duration and phase displacement may also be positively correlated, but only for syllable type. Phases associated with the production of clusters are articulated with greater duration and displacement than those associated with the production of single consonants. In contrast, given the result that opening phases were generally produced with greater duration, but less

Table 5.1: Correlation coefficients for the relationship between phase duration and displacement for the 4 speakers (N = 18 in each case).

	simple syllables		complex syllables	
Speakers	opening	closing	opening	closing
MB	0.23	0.60	-0.08	-0.30
MC	0.17	0.36	0.19	-0.51
LR	0.76	0.43	0.68	-0.53
MM	0.45	0.54	-0.03	0.31

displacement than closing phases, it was not expected that phase duration and displacement would be positively correlated for syllable position. The relationship between the two measurement variables was formally tested here with respect to phases of the jaw cycle. Table 5.1 shows the coefficients of the correlations between opening/closing phase duration and displacement as a function of syllable type and speaker.

Surprisingly, the correlations between phase duration and displacement varied more as a function of simple and complex syllable types than as a function of opening or closing. The positive correlations between the variables for phases associated with the production of simple onsets and offsets were in line with expectations given previous findings, but not in line with intuition given the results from previous sections of this study. In contrast, the mostly negative and relatively weak correlations between the variables for phases associated with the production of clusters were unexpected. Especially during the closing phase of the cycle, cluster production was associated with either an increase in phase displacement and a corresponding decrease in phase duration or vice-versa.

While the overall correlation between phase duration and displacement was surprisingly inconsistent, the correlation between phase displacement and peak velocity was very strong. This difference in strength is illustrated when the data are plotted for displacement as a function of duration (Figure 5.6) and displacement as a function of peak velocity (Figure 5.7).

According to Kelso, Vatikiotis-Bateson, Saltzman, and Kay (1985) and Ostry and Munhall (1985), the ratio of displacement to peak velocity is a measure of muscle stiffness. These researchers argue that the degree of stiffness is indicated by the slope of the correlation and greater slopes indicate greater stiffness. Table 5.2 provides the slopes and regression coefficients for the relationship between distance and peak

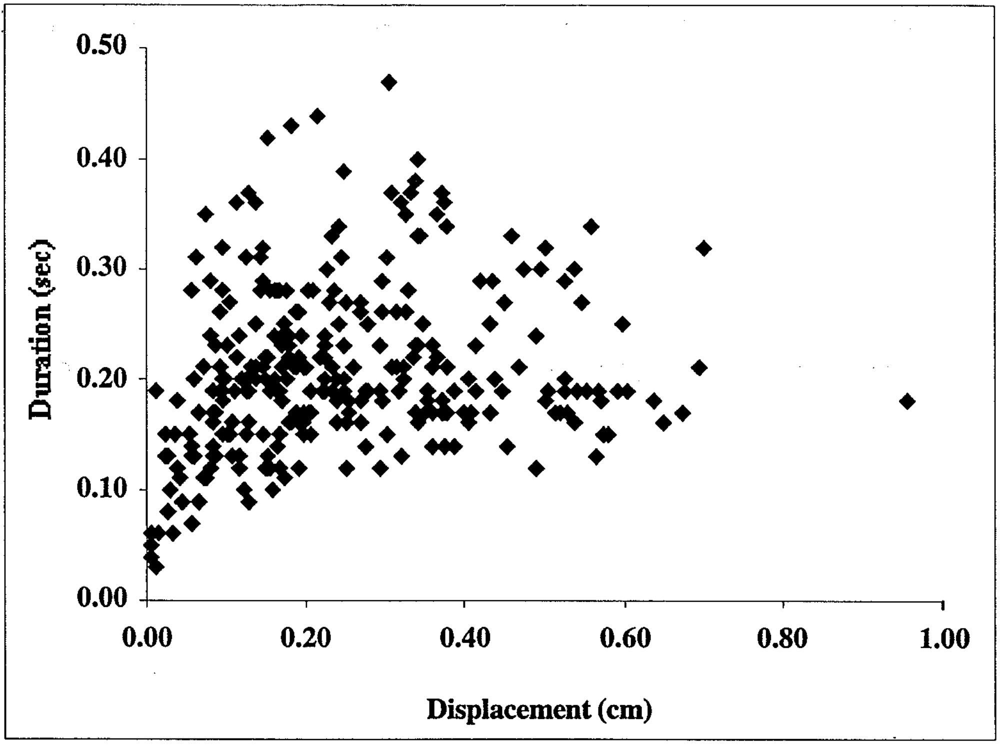


Figure 5.6: **Correlation between displacement and duration of the phase.** The scatterplot shows phase displacement as a function of phase duration for both phases, all syllable types, and all speakers.

velocity during the opening and closing phases associated with simple and complex onsets and offsets respectively.

Table 5.2 shows that for every speaker, (with the exception once again of speaker MM), the Kelso et al. (1985) and Ostry and Munhall (1985) measure of stiffness is greater during the closing phase than during the opening phase. Stiffness also tends to be greater during the articulation of clusters than during the articulation of single consonants, particularly if the clusters are syllable final.

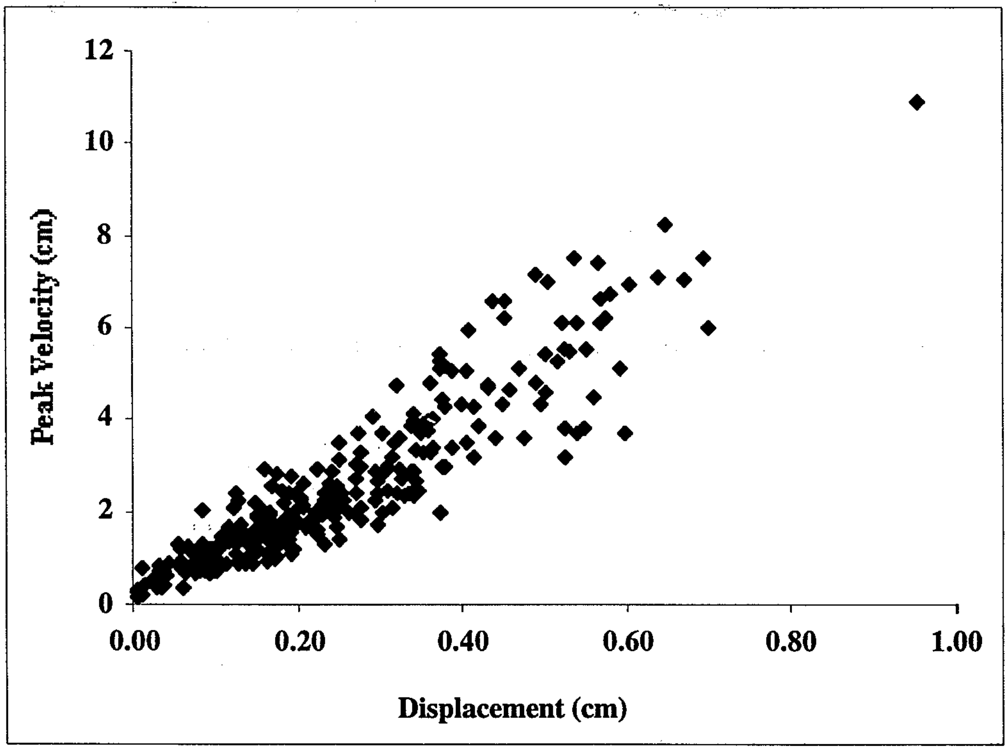


Figure 5.7: **Correlation between displacement and peak velocity of the phase.** The scatterplot shows phase displacement as a function of peak velocity for both phases, all syllable types, and all speakers.

Table 5.2: Slope and R^2 (in parentheses) of the relationship between phase displacement and peak velocity for the 4 speakers (N = 18 in each case).

	simple syllables		complex syllables	
Speakers	opening	closing	opening	closing
MB	6.79 (0.82)	10.81 (0.95)	9.51 (0.78)	12.12 (0.96)
MC	9.28 (0.85)	11.85 (0.91)	12.80 (0.93)	12.34 (0.95)
LR	7.93 (0.93)	9.28 (0.92)	7.06 (0.88)	12.10 (0.96)
MM	10.76 (0.78)	3.86 (0.24)	13.21 (0.84)	11.39 (0.63)

5.4 Discussion

Four speakers produced mirror-image syllables embedded in frame sentences while acoustic and jaw movement recordings were made. The resulting acoustic and jaw movement data were used to test the hypothesis that inherent properties of the jaw cycle may contribute to the increased articulatory and acoustic reduction and variability of syllable-final consonants as compared with syllable-initial consonants. Two types of syllables were examined: those with simple (single consonant) onsets and offsets and those with complex (consonant cluster) onsets or offsets. As expected, syllable type influenced the jaw cycle in that as syllables with more segments required a larger frame than syllables with fewer segments. Accordingly, duration, displacement, and peak velocity all tended to be greater for phases associated with the production of clusters than for those associated with the production of single consonants. Apart from these specific differences associated with the production of clusters, differences between the opening and closing phases of the cycle were consistent across syllable type. Given that the syllables produced in the study were constructed as mirror images of one another, the overall result could not have been driven by the phonological content of the syllable. Instead, it must be concluded that jaw movement associated with an opening gesture is slower, shorter, and less stiff than movement associated with a closing gesture due to inherent properties of the cycle.

5.4.1 Segments within the cycle

In order to establish a relationship between the syllable onsets/offsets and the jaw cycle, measurements of jaw height and acoustic duration were made for the initial and final consonants, the vowel nucleus and for the opening and closing phases of the jaw cycle. The opening and closing phases of the cycle were delimited by consecutive points of minimum and maximum jaw opening. Measurements of jaw height during the articulation of syllable-initial and syllable-final consonant(s) relative to the points of minimum and maximum opening provided a rough indicator of where these consonants were articulated during the cycle. Regardless of syllable type, initial consonants were articulated further into the cycle, that is, closer to maximum opening, than final consonants. Stop consonants that functioned as the interior segment of initial or final clusters were articulated closer to maximum opening or further into the cycle than single stop consonants. These results were consistent with the view that the segments of a single syllable are articulated within a single jaw cycle.

Measurements of segment duration indicated that, regardless of syllable type, segments associated with syllable onsets were of greater duration than those associated with syllable offsets. The placement and duration of segments within the cycle also differed in the two syllable types. The overall duration of consonant clusters was greater than for single consonants. These results therefore replicated known and expected differences in segment duration as a function of syllable type. The remainder of the analyses were conducted to provide an articulatory basis for the difference observed in the production of initial and final consonants.

5.4.2 Properties of the jaw cycle

The movement properties examined for the jaw cycle included phase duration, displacement, and velocity. It was found that the opening phase of the jaw cycle was usually associated with greater duration than the closing phase. Phases associated with the articulation of complex onsets or offsets were longer than those associated with simple onsets or offsets. Hence, the duration differences between the phases of the jaw cycle correlated with the duration differences found for syllable-initial and syllable-final segments.

Like duration, jaw displacement was found to vary as a function of the opening or closing movement. Closing phases were articulated with greater displacement than opening phases in complex syllables, but both phases of complex syllables were articulated with greater displacement than the phases of simple syllables. These differences in displacement paralleled those found for velocity. Closing phases were articulated with greater peak velocities than opening phases. These measurements were in agreement with previously reported results for open (CV) syllables and phrase-final, closed syllables (e.g., Sussman, MacNeilage, Hanson, 1973; Kuehn and Moll, 1976; Kelso, Vatikiotis-Bateson, Saltzman, Kay, 1985; Gracco, 1994). As with displacement, peak velocity was effect by syllable type in that complex syllables were articulated with more velocity than simple syllables.

In addition to analyses of individual measures, the relationship between the various measures was examined. Correlations between the duration and displacement variables suggested that one influenced the other, at least in cycles associated with the production of simple syllables. In other parts of this dissertation it has been argued that amplitude of the cycle is specified by the phonological content of the syllable, but that the duration of the cycle is due to movement properties of the jaw. In the present study, the lack of a similar relationship between phase duration and displacement for complex syllables can only mean that movement velocity is especially increased during the production of these types of syllables.

The relationship between jaw displacement and velocity was also examined. As elsewhere (Kelso, Vatikiotis-Bateson, Saltzman, Kay, 1985; Gracco, 1994), displacement and peak velocity were very strongly correlated. Kelso et al. (1985) and Ostry and Munhall (1985) have argued that the relationship between these two variables is an indicator of articulatory stiffness. According to these researchers, the degree of stiffness is indicated by the slope of the correlation between the variables – bigger slopes indicate greater stiffness. In the present study, the slopes of the correlations between distance and velocity were smaller for opening phases than for closing phases and for simple syllables than complex syllables. The first of these results is consistent with the findings of the above-mentioned researchers, whereas the second of these results is new. It is not clear, however, what the results actually mean. Kelso et al. (1985) and Ostry and Munhall (1985) base their analysis of muscle stiffness on a spring model of articulator movement. Their model does not, however, incorporate articulator mass and therefore may not accurately reflect anything about the relative stiffness of the underlying musculature.

5.4.3 Relationship between production and perception

A trading relationship between articulatory speed and accuracy was hypothesized, where increases in jaw speed lead to corresponding decreases in production accuracy. The results from the first part of this study indicate that the speed of jaw movement increases from simple onsets to simple offsets and from complex onsets to complex offsets. If segmental articulatory accuracy decreases in parallel, onsets should be more identifiable than offsets and single consonants should be more identifiable than consonant sequences. This hypothesis is tested in the next phase of the present study. A perceptual confusion study is conducted where listeners are asked to identify the previously analyzed tokens from our 4 speakers.

5.5 Methods

5.5.1 Stimuli

The stimuli for the perceptual confusion task were the previously recorded sentences with their embedded tokens and included the CV[ʃ] and [ʃ]VC tokens with initial and final [s]. The waveform editor was used to calculate the RMS amplitude for the first word, (i.e., "sign"), of the frame sentence and the entire utterance was then scaled so that the average amplitude of the first word was consistent across all utterances. In this way the stimuli were normalized across speakers, but the inherent amplitude differences of the different target syllables were preserved.

5.5.2 Listeners

Listeners' were 14 undergraduate Introductory Psychology students from the University of Texas; all had normal hearing and spoke English as a first language. The listeners were told that they would be listening to a variety of single-syllable nonsense words embedded in a frame sentence. They were told what consonant and vowel types they would hear, but not about the different syllable types, and were instructed to write the nonsense word in normal orthography on the provided response sheet.

Listeners were seated in a sound attenuated room and the sentences were presented in noise (+15 S/N) and at normal hearing levels (around 65 dB) over earphones in a randomized sequence at interstimulus intervals of 3.5 seconds.

5.6 Results

Misperception errors included the omission or addition of consonants as well as the more common misperceptions of place and/or manner. Voicing errors were exceedingly rare and are therefore ignored in this discussion. Table 5.3 and 5.4 show the types and number of errors associated with the consonants in different syllable positions and those associated with single consonants versus consonant clusters. The two types of onsets and offsets are placed together as they were often confused with one another.

The confusion matrices indicate that place-of-articulation errors were more common than manner-of-articulation errors. Misperceptions involving an omission of at least one consonant were more common for final clusters than for initial clusters. Of the total number of errors for final clusters, 50.09% involved the omission of at least one consonant, whereas 32.08% of the total number of errors for initial clusters were errors of omission. The greater number of omission errors in final position is consistent with the idea final consonant have less "room" to be articulated within the cycle than initial consonants since the closing phase of the cycle is executed more quickly than the opening phase.

It is also clear from the two tables that listeners misperceived offsets more frequently than onsets and they misperceived clusters more often than single consonants. This latter finding was confirmed in a 5-way (speaker, syllable position, syllable type, vowel nucleus, stop consonant type) ANOVA. Initial consonants were found to be identified at a statistically significantly higher level than final

Table 5.3: Perceptual confusion matrix for simple and complex onsets. Target onsets are shown on the horizontal and responses are presented on the vertical. A total of 288 responses were made for each target.

	Simple Onsets			Complex Onsets		
	p	t	k	sp	st	sk
p	**229**	72	32	46	17	19
t	12	**143**	21		5	3
k	43	70	**225**	21	28	37
f	1	1	4	24	14	2
s				3	24	6
S			1	2	12	6
sp		1		**179**	27	8
st			1	8	**99**	9
sk	1	1	2	2	55	**189**
none			1			2
other	2		1	3	7	7

Table 5.4: Perceptual confusion matrix for simple and complex offsets. Target offsets are shown on the horizontal and responses are presented on the vertical. A total of 288 responses were made for each target.

	Simple Offsets			Complex Offsets		
	p	t	k	sp	st	sk
p	**161**	54	78	59	36	15
t	10	**123**	8	12	16	9
k	37	21	**110**	18	17	55
f	6	4	4	5	3	2
s	4	10	9	3	12	6
S	2	5	5	3	5	6
sp	21	16	8	**131**	44	29
st	27	35	17	21	**124**	31
sk	7	7	35	19	19	**129**
none	6	6	2	6	2	4
other	7	7	12	11	10	5

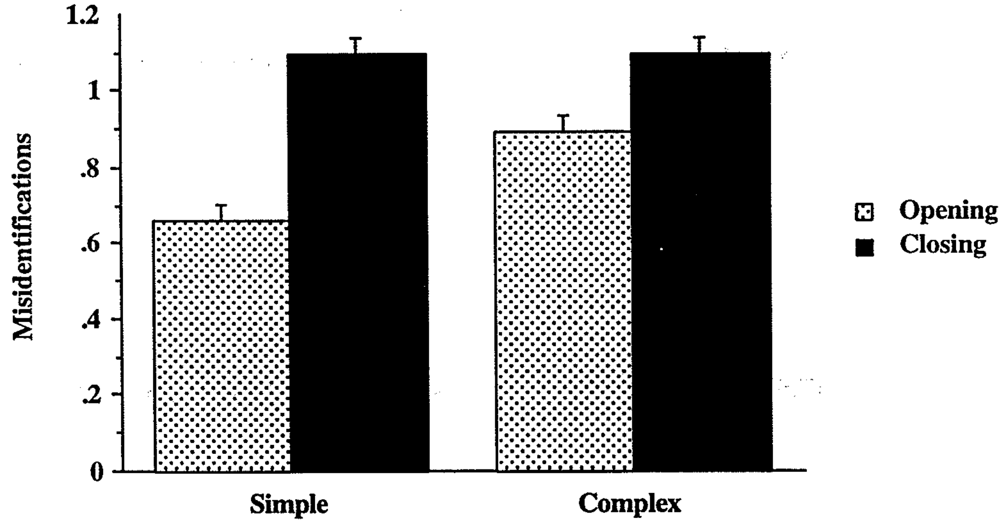


Figure 5.8: **Identification errors as a function of syllable position and type.** Since there were two repetitions of each stimuli, the maximum number of errors possible per cell was two.

consonants ($F(1,11) = 59.277, p < 0.01$). The difference in level of identification between single consonants and clusters did not reach statistical significance, but there was a significant interaction between syllable position and syllable type ($F(1,11) = 5.290, p < 0.05$). This interaction is displayed in Figure 5.8.

Mean comparisons showed that single consonant onsets were more perceptible than cluster onsets, but offsets were misidentified equally often regardless of syllable type. Figure 5.8 also shows that offsets were misidentified more often than onsets in simple and complex syllables (initial C vs. final C ($F(1,11) = 37.397, p < 0.01$); initial CC vs. final CC ($F(1,11) = 8.194, p < 0.05$)).

Other main effects were found for speaker, vowel nucleus, and consonant type. Each of the interactions between each of these conditions and syllable position and syllable type was significant. Mean comparisons indicated, however, that the previously mentioned trends held with greater or less strength across different conditions. For instance, final consonants were misidentified more often than initial consonants in 4 of the 4 speakers, 2 of the 3 vowels, and 2 of the 3 consonants. There were no conditions in which initial consonants were misidentified more than final consonants. In addition, clusters were misidentified more often than single consonants in 2 of the 4 speakers, 1 of the 3 vowels, and 1 of the 3 consonants. There were no conditions in which single consonants were misidentified more often than clusters. Thus, in spite of the significant interactions between the factors, the overall pattern of perceptibility remained relatively constant: consonants in syllable-initial position were more perceptible than those in final position, simple onsets were more perceptible than complex onsets, but simple offsets and complex offsets were misidentified equally often.

5.7 General Discussion

The results of the production experiment coupled with those of the perceptual confusion experiment provide an articulatory and perceptual basis for the cross-language tendency formalized by the Maximal Onset Principle of phonology, that is the preference for initial consonants over final consonants and the preference for single onsets/offsets over complex onsets/offsets and for complex onsets over complex offsets. The articulatory factor examined in this study was jaw movement because the interaction between the jaw cycle and segmental articulation is hypothesized to yield suprasegmental patterns associated with syllables. The results from the perceptual confusion experiment were assumed to reflect underlying articulatory differences in the production of single consonants and consonant clusters in different syllable positions.

The data discussed in this chapter suggest that segments may be articulated differently in different syllable positions due to inherent differences in the phases of the jaw cycle. It was found that the closing phase of the jaw cycle is usually produced with shorter duration, greater displacement, peak velocity, and stiffness, than the opening phase regardless of syllable type. Similarly, phases associated with the articulation of complex onsets/offsets were usually also produced with greater displacement and peak velocity than phases associated with the articulation of simple onsets/offsets. The differences between the phases of the jaw cycle are hypothesized to lead to a trading relationship between articulatory speed and accuracy. Thus, initial consonants may be articulated more distinctively and with less variability than final consonants (Byrd, 1996; Byrd and Tan, 1996; Sussman, Bessell, Dalston, Majors, 1997; Redford and Diehl, 1999) because these are articulated during the first half of the cycle when the jaw is moving more slowly. The relatively slower movement of the jaw gives the segmental articulators more time in which to position themselves for consonant articulation. Single consonants may be articulated more distinctively and with less variability than clusters for the same reason. In addition, clusters also require that the segmental articulators move more than once, hence more quickly, during the relatively closed portions of the cycle.

Because articulatory accuracy is hypothesized to affect perceptibility, a perceptual confusion experiment was conducted to establish the relative perceptibility of simple and complex onsets and offsets. The results indicated that initial consonants are more perceptible than final consonants and that single consonants are more perceptible than consonants in consonant clusters. The syllable position effect holds for both types of onsets and offsets so that complex onsets, though less perceptible than simple onsets, are more perceptible than complex offsets.

If one does not accept the proposed connection between the inherent differences in the phases of the jaw cycle and their presumed perceptual effects, then the combined results from the production and perceptual confusion study suggest that two distinct constraints operate against the final consonants and consonant clusters. The finding that final consonants and clusters are articulated during more rapid movement of the jaw than initial consonants and single onsets/offsets indicates that the observed cross-language preferences in syllable structure may be influenced by articulatory ease. In addition, the results from the perceptual confusion experiment suggest that a perceptual distinctiveness constraint may play a role in the preference for syllable-initial consonants over syllable-final consonants and single consonants over clusters. Nevertheless, given the necessary relationship between perception and production in speech and the parallelism between the production and perception results from this study, a more probable scenario is one in which the constraints of articulatory ease and perceptual distinctiveness work together to define these specific cross-language preferences.

Chapter 6

Implications and Conclusions

A major implication of the ideas presented in this dissertation is that we may be able to ground the syllable in phonetic fact after all. If it is possible to provide a unified phonetic explanation for the cross-language phonological patterns characterized by formal principles, such as the Sonority Sequencing Principle and the Maximal Onset Principle, then it becomes unlikely that these patterns are the result of an innately specified, formal concept. In addition, if syllable perception can be linked to articulatory factors, which are in turn linked to the acquisition of speech, then it may be possible to explain the formation of the concept "syllable" in embodied terms.[1] At this point, however, the evidence is suggestive rather than conclusive and future work is required to solidify and extend the basic idea that the jaw cycle is the defining articulatory factor in syllable production. Hence, in this final chapter, the specific hypotheses presented in this dissertation are revisited with an eye to their limitations and implications. Avenues of future research are also considered.

6.1 The jaw cycle and syllable production

The hypothesis that the jaw cycle provides an articulatory basis for syllable production is dependent on the assumption that the jaw cycle constrains the movements of the segmental articulators and that this constraint is reflected in sound patterns perceived as syllables. Though in keeping with a view developed by MacNeilage (1998) to explain the development of speech production, this view contradicts the dominant view that sees jaw movement as entrained to movements of other supraglottal articulators (e.g., lips and tongue), but otherwise largely irrelevant to speech production. The difference between these two views is most evident in the alternative explanations they provide for the phonological and phonetic patterns associated with syllables.

6.1.1 Mechanical constraint

The normal sequential organization of segments within syllables is described by linguists in terms of a manner hierarchy (see Bell and Hooper, 1978) usually referred

---

[1] Infants may come to recognize the syllable as a unit of speech via their initial experience in producing speech-like sounds. In these initial productions, infants are not controlling the production and concatenation of individual consonants and vowels, rather they are producing consonant-vowel units merely by the action of opening and closing their mouth during phonation (Davis and MacNeilage, 1995; MacNeilage, 1998). Thus, infants may come to identify consonant-vowel units as real units of speech prior to recognizing that speech also consists of segmental units that can be independently manipulated.

to as the sonority hierarchy. The manner hierarchy can in turn be described in articulatory terms as an openness hierarchy. Stops and fricatives or obstruent consonants, which are the segments articulated with the most vocal tract constriction, usually occur on the edges of syllables. Liquids and glides, segments articulated with less vocal tract constriction than the other consonants, usually occur adjacent to the syllable nucleus. Vowels, the segments articulated with the least vocal tract constriction, usually function as syllable nuclei. Lindblom (1983) noted a strong correspondence between the degree of vocal tract constriction and jaw height during segment articulation. He showed that the sonority hierarchy described by linguists can be correlated with a jaw openness hierarchy: obstruents tend to be articulated during maximal jaw closure, liquids and glides during a more open portion of the jaw cycle, and vowels during maximal jaw opening. This correlation between sonority and jaw height is important, but it does not indicate which variable is derived from the other. Two interpretations are possible, each of which corresponds to one of the two views of jaw movement in speech. On the view that the jaw moves in service of segmental articulation, the jaw cycle may be derived from a cognitive template provided by the Sonority Sequencing Principle of phonology. On the view proposed in this dissertation, namely, that the lips and tongue tend to conform to the jaw cycle, the sonority hierarchy emerges naturally from the mechanical constraint of the cycle on segmental articulation.

Evidence

In Chapter 3, these two views were distinguished by examining jaw movement during the production of legal Russian monosyllables that either conformed to or violated the Sonority Sequencing Principle. The organization of segments in the two complex syllable types could be described from the segmental point of view as "closed, more open, open" (e.g., [bla]) or as "more open, closed, open" (e.g., [lba]). If the jaw moves in service of segment articulation, it was predicted that the jaw movement associated with the different syllable types should follow the segmental pattern. If segmental articulations conform to jaw movement, it was predicted that both syllable types should be described by a single jaw cycle. Although the results clearly disconfirmed the hypothesis that the jaw moves in service of the segmental articulators, the strong version of the alternative hypothesis – the jaw cycle as a mechanical constraint – was not upheld. The onsets of both syllable types could be described by a single jaw movement, which proceeded from a relatively closed position to a more open one, but there also appeared to be some compromise on the part of jaw for segmental articulation. Instead of finding that the first consonant of the cluster was always articulated with more jaw closure than the second, it was found that, in certain cases, both segments were articulated at the same relative jaw height. The evidence therefore supports a view in which the jaw provides a relatively flexible mechanical constraint on segment articulation.

A relatively flexible constraint may, in fact, allow for the emergence of more natural patterns than an inviolable mechanical constraint. The constraint would be sufficient to account for the emergence of the regular pattern of segment sequencing observed across languages and described by the Sonority Sequencing Principle. In addition, it would allow for those less frequent patterns of segment sequencing in which segments of a cluster are of equal 'sonority' (e.g., the English onsets [sp-], [st-], and [sk-]) or for those very rare patterns where segments of relatively high 'sonority' occur as external members of a cluster (e.g., the Russian onsets [lb-] and [lg-]).

Future work

In subsequent chapters of this dissertation, the mechanical constraint of jaw cycle on segment articulation was assumed. In Chapter 5, this assumption led to the proposal that inherent differences in the phases of the jaw cycle affected segmental articulation to produce the well-documented production differences observed for syllable-initial and syllable-final consonants. It was argued that greater articulator velocity during the closing phase of the jaw cycle adversely affects the accuracy of segmental articulation during the production of final consonants. Reasoning of this sort should be further supported by direct evidence of the action of the segmental articulators during the different phases of jaw movement. This evidence could be procured by measuring lip and tongue movements as well as jaw movements during the production of consonants in different syllable positions.

More generally a model needs to be developed to explain precisely the manner in which jaw movement affects segmental articulation. My working hypothesis would be based on the notion of articulatory ease as defined by Willerman (1994) for tongue movement in consonantal articulations. According to Willerman, articulatory ease decreases as deviations from a neutral position increase. Similarly, one might argue, as Lindblom (1983) has, that different consonants and vowels are produced during different portions of the jaw cycle in order to minimize travel distance for segmental articulators. This hypothesis could be tested by obtaining measures of the preferred jaw heights at which different consonant and vowel types are articulated (as in Keating, Lindblom, Lubker, Kreiman, 1994). The jaw could then be immobilized at different heights and the relative displacement and velocity of the articulators could be measured during the production of the same segments. A comparison of these measurement sets should indicate the extent to which tongue configurations and jaw height correlate during production. The second set of measurements would also provide an estimate of the relative cost of producing different segment types at different points during the jaw cycle. A computational model simulating the emergence of sound patterns could use these measures of relative cost to constrain the emergence of segment sequencing patterns. It should be noted, however, that such a constraint would mostly give rise to patterns that best conform to the jaw cycle, even though these are not the only patterns that exist in languages (e.g., [lba]). The fact that unexpected patterns of segment sequencing occur in languages provides evidence for the view that sound patterns emerge in response to multiple functional constraints, only one of which is articulatory.

6.1.2 Temporal constraint

Another pattern associated with syllables is the temporal pattern of different relative segment duration. This pattern is well documented in the phonetic literature and has been shown to play a role in syllable perception (Boucher, 1988; Tuller and Kelso, 1991; Anderson and Port, 1994). It is often assumed that segment duration variation results either from unspecified principles of neural organization in the articulatory program (e.g., Kozhevnikov and Chistovich, 1965; Lehiste, 1977) or from unspecified principles of gestural timing (e.g., Munhall, Fowler, Hawkins, Saltzman, 1992; Harrington, Fletcher, Roberts, 1995). The view that attributes segment duration variation to a unit of neural organization in the articulatory program is similar to the phonological view: both propose that the syllable is an abstract unit of representation. If the hierarchical organization of the supraglottal vocal tract is ignored or thought to be inconsequential to speech, then the view that attributes segment duration variation to principles of relative gestural timing is also consistent with the phonological view since relative timing must then be explained further upstream. In contrast, if the jaw cycle is thought to constrain segmental articulation, a gestural timing or coarticulatory explanation of segment duration variation does not need to posit the abstract representation of the syllable in speech production. Instead syllable-related segment duration variation can be explained to emerge as a by-product of the interaction between segmental articulation and jaw movement. This latter view was adopted in this dissertation.

The specific hypothesis developed in this dissertation regarding the emergence of syllable-related segment duration patterns assumes that the jaw cycle acts as a sort of "receptacle" within which segments are articulated. A primary assumption was that one jaw cycle, designated by two sequential points of maximal jaw closure, describes one syllable. A large cycle is "wider" (i.e., longer) and "deeper" (i.e., more open) than a smaller cycle. The larger the cycle, the more room for segment articulation. This view provides a different understanding for why stressed vowels are longer in duration than unstressed vowels and why they are perceived to attract consonantal onsets and offsets: a stressed vowel is associated with more jaw opening than an unstressed vowel and, thus, with a larger overall cycle.

Another assumption behind the hypothesis of the jaw cycle as receptacle is that depth (i.e., cycle amplitude) is specified by segment type, whereas width (i.e., cycle duration) is specified by inherent movement properties of the jaw. This assumption extends Lindblom's (1967) explanation of duration differences associated with different vowel types to the duration patterns associated with all segment types articulated within a single cycle. Lindblom showed that vowel duration was positively correlated with jaw opening. Low vowels, articulated with more jaw opening, are greater in duration than high vowels, articulated with less jaw opening. Lindblom argued that more open vowels are longer because the increase in jaw velocity that accompanies greater jaw opening is not sufficient to compensate in time for the greater jaw displacement. If we can assume that the duration of the cycle is set by its depth, then it is possible to understand why increasing the number of segments associated with a single cycle of a fixed depth would produce a corresponding decrease in the duration of the consonants within that cycle. This view explains, for example, why vowel duration decreases as the number of final consonants increase. It also explains why segments that participate in a cluster are shorter in duration than when they occur singly. This view does not, however, explain why decreases in segment duration are unevenly spread out across all segments of a syllable. Hence, the fact that interior members of a cluster are shorter than exterior members is not explained by the receptacle hypothesis.

Evidence

The results from this dissertation provided evidence for the view that a single syllable corresponded to a single jaw cycle, defined by two consecutive points of maximal jaw closure. This was most clearly shown in Chapter 4 where subjects were asked to make syllable boundary judgments on di-syllabic tokens with medial consonant sequences. It was found that subjects divided or left intact the consonant sequence depending on whether the first consonant of the sequence was primarily articulated during the first or second jaw cycle (see Table 4.5). For instance, if the [s] of *destibe* was primarily articulated during the closing phase of the first jaw cycle associated with the word, then subjects syllabified the word as *des.tibe.* On the other hand, if the [s] of *destibe* was primarily articulated during the opening phase of the second jaw cycle, then subjects syllabified the word as *de.stibe.*

Other evidence that segments of a single syllable are articulated within a single cycle was provided with measures of jaw height during segment articulation. As previously indicated, the jaw height measures of the Russian onset clusters were consistent with the view that these were articulated during the opening gesture of a single cycle. In Chapter 5, the jaw height measures for the various consonantal onsets and offsets were compared with the points of minimum and maximum jaw opening, which delimit a single phase of the cycle (Figure 5.1). This comparison, though made in space and not in time, convincingly showed that the consonants of simple and complex onsets and offsets were articulated during the opening and closing phase of the cycle, respectively.

Evidence that segment duration variation emerges from the interaction of the segmental articulators and the jaw was provided in the form of correlations and parallels between jaw movement and acoustic patterns of segment duration. In Chapter 3, overall syllable duration was positively correlated with maximal jaw height. In Chapter 4, vowel duration was positively correlated with jaw height. In addition, data from Chapter 4 supported the receptacle hypothesis. The duration of the second consonant (C2) in the sequence of word-medial consonants was predicted by the duration (and therefore height) of the first vowel (or cycle). If the first vowel was long in duration, the first cycle was larger. The first consonant (C1) of the medial sequence was therefore articulated during the closing phase of this larger first cycle. This means that C2 was articulated, by itself, during the opening phase of the second cycle, which would have allowed C2, now a single consonant onset to the second syllable, to be longer than if it was the second segment articulated during the same opening phase.

The hypothesis that segment duration variation reflects the temporal constraint of the jaw cycle on segment articulation was also supported in Chapter 5. A correlation was found to exist between segment duration and phase duration in the tokens produced by all the speakers. In the tokens of most speakers, segments associated with syllable offsets were shorter than those associated with onsets just as the closing phases of the jaw cycle were shorter than opening phases. In one case (speaker MM), however, this pattern was reversed and phrase-final lengthening was observed for segments associated with syllable offsets as well as for the closing phase of the cycle.

Future work

Although this dissertation provided evidence to support the hypothesis of the jaw cycle as temporal constraint, the link between the jaw cycle and acoustic patterns was only indirectly established through correlations and parallels. Establishing a link between jaw movement and acoustic patterns associated with syllable perception is critical to the hypothesis that the jaw cycle provides an articulatory basis for the syllable. On the view that the sound patterns of language are shaped by an interaction between speakers and listeners, syllables only gain their status as speech units once they are perceived. A number of different experiments could be conducted to establish this link more directly.

One test of the hypothesis that the temporal constraint of the jaw cycle gives rise to the different patterns of segment duration associated with syllables would be to hold jaw movement constant during speech production and test whether the micro-structure of the segment duration patterns remains the same. Jaw movement can be held constant with bite-blocks. Subjects could be asked to produce different syllable types in frame sentences with or without bite-blocks. Relative segment duration could then be analyzed and compared across conditions. The expectation is that when jaw movement is held constant the normal pattern of relative segment duration would be disrupted and syllable boundaries would be less predictable from the duration patterns.

A similar test could be made with natural speech that does or does not involve jaw movement. For instance, it is likely that pharyngeal consonants, such as those that occur in Arabic, are produced outside of the jaw cycle (though this would need to be established through measurement). In some dialects of Arabic (e.g., Cairene) pharyngealization spreads to all the segments of the syllable (Broselow, 1979). A native Arabic speaker could therefore produce pharyngealized syllables in addition to syllables that would require normal jaw movement. Segment duration could then be measured as a function of syllable position. Follow-up perceptual judgment studies could be conducted with native and non-native Arabic speakers to determine whether the duration differences, or lack thereof, appropriately signal syllable boundaries.

6.2 Syllable structure

The hypothesis of a jaw cycle constraint on segmental articulation is meant to explain the patterns of segment sequencing and relative segment duration that are associated with syllables across languages. But this constraint may also be successful at explaining cross-language preferences associated with syllable structure. The aspect of syllable structure treated in this dissertation was the cross-language preference for consonantal onsets over offsets and for single consonants over clusters. It was argued that the preference for syllable-initial consonants over syllable-final consonants is derived from inherent asymmetries in the jaw cycle. Since the closing phase of the cycle is executed more rapidly than the opening phase, segmental articulation during the closing phase of the cycle would also need to be executed more rapidly. It was hypothesized that the more rapidly the articulators move, the more likely target articulatory configurations will be undershot or variable. Although this hypothesis will require further work to be substantiated, it provides a basis for the observed articulatory and perceptual disadvantages associated with final consonants compared with initial consonants.

The preference for single consonants over consonant clusters was also explained to emerge from inherent properties of the cycle. Here, however, the explanation rested on the assumption that articulatory effort increases with each additional segmental gesture within a cycle. Since cycle duration is assumed to be relatively stable, an increase in the number of segmental gestures associated with a single syllable implies a corresponding increase in the rate at which these gestures will be executed. The previously mentioned patterns of relative segment duration support the hypothesis that the articulators must move more rapidly from one configuration to the next when the number of segments is increased within the syllable. The duration of consonants in clusters is shorter than the duration of consonants that occur as single onsets or offsets. Similarly, vowel duration is negatively effected by increases in the number of syllable-final consonants.

In spite of decreases in segment duration, the total duration of complex syllables is usually greater than the duration of simple syllables. This corresponds with the relatively longer duration of cycles associated with complex syllables. Results from Chapter 5 suggest that the increased duration of the jaw cycle may be accomplished by increasing the amplitude of the cycle. Phases associated with the production of clusters were greater in duration and displacement than those associated with the production of single consonants. Although a larger cycle provides more room within which segments may be articulated, the observed corresponding increase in articulator displacement and velocity suggests that increases in amplitude may also increase articulatory costs associated with syllable production.

The assumption that articulatory costs increase with greater cycle amplitude or when a greater number of segments are articulated within a single cycle, combined with the hypothesized effect of the cycle's asymmetry on segmental articulation predicts certain patterns of cross-language syllable structure. One prediction, discussed in Chapter 5, is that consonants should occur preferentially in syllable-initial position and that simple onsets/offsets should occur more frequently than complex onsets/offsets. And, as previously discussed, this is, in fact, the pattern documented in the typological data. If this reasoning is continued, however, a generalization emerges: additional consonants should be added where none occur or at the beginning of the cycle rather than at the end. Syllable types should therefore become more complex in an iterative fashion. For instance, the basic CV syllable type may be first modified to form a CVC or a CCV syllable. The next expected modification would yield a CCVC syllable, the next a CCCVC or CCVCC syllable, and so on. Since each modification would increase articulatory costs, we might expect that more complex forms would be less well represented across languages and within a single language than more basic forms. Similar hypotheses regarding the effect of articulatory complexity on sound systems have been supported for phoneme inventories (Lindblom and Maddieson, 1985) and for the relative use of different sounds within a single inventory (Willerman, 1994).

The typological data are fairly consistent with an "iterative principle of syllable structure." Blevins (1994) notes, for example, that "if clusters of n Cs are possible syllable-initially, then clusters of (n - 1) Cs are also possible syllable initially (217)." She notes that the same relation holds in syllable-final position. And, as previously mentioned, clusters are preferred in syllable-initial position relative to syllable-final position (Bell and Hooper, 1978). The hypothesis could be tested further with relative frequency data on syllable types in a diverse set of languages. An initial sampling of this sort, shown in Table 6.1, suggests that this hypothesis may receive support within and across languages.

One hundred words were randomly selected from different language dictionaries. The words were then syllabified according either to information given in the dictionary or, if no information was given, according to the Sonority Sequencing Principle and the Maximal Onset Principle. Table 6.1 shows that CV syllables are the most common syllable type in each of these languages (with the one exception of Efik). Other highly frequent syllable types include, as hypothesized, the CVC and CCV syllables. More complex syllables are under-represented. Table 6.1 also suggests one limitation to the iterative principle; namely, syllables consisting of a single V are relatively frequent, though usually less frequent than CVC or CCV syllables.

The problem of syllables that lack a consonantal onset was not addressed in this dissertation and will need to be explored in future research. These syllable

Table 6.1: The relative frequency of different syllables types are displayed for a diverse group of languages. Syllable types were derived from a random sample of 100 words per language.

Languages	CV	CVC	CCV	V	VC	CVV	CCVC	CCVV	other
Czech	136	56	42	9	5	4	13		10 (N=4)
English	77	51	8	7	8	7	9	2	37 (N=10)
Spanish	195	65	20	12	15	25	5	2	11 (N=2)
Alabama	181	86		20	23	50			5 (N=2)
Dakota	194	56	31	54	3		17		2 (N=1)
Efik	42	81	4	71	14	2	6		16 (N=2)
Luganda	214	29	25	2		29	2	1	3 (N=1)
Mansaka	112	92	3	14	11				
Hawaiian	178			28		57			9 (N=9)
Japanese	198	30	5	10	1	22			1 (N=1)

types, like the others, conform to a single jaw cycle, since vowels are articulated with an open jaw configuration and the jaw rests in a relatively closed position. If a syllable consists of a single initial vowel, however, then voicing must be discontinued during the opening trajectory of the jaw, otherwise a consonant-like sound will result during the more closed portions of the cycle. It is perhaps for this reason that glottal stops are often inserted before stressed V syllables (Hoard, 1966; Redford and Diehl, 1999). This example illustrates how the idea that the jaw cycle provides the major articulatory basis for the syllable may also be useful in explaining the phonetic and phonological facts associated with syllables not addressed in this dissertation.

6.3 Conclusion

The present studies were undertaken in order to establish a phonetic basis for the syllable, a unit of speech which is currently thought by many to lack such a basis. In this dissertation it was assumed that syllables are psychologically real to language speakers because they exist in the speech stream. Although phonologists and phoneticians have identified sound patterns in speech that are associated with syllable perception, it has usually been assumed that these patterns are inserted into the speech stream because of our innate concept of a syllable. In this dissertation, however, it was assumed that sound patterns associated with syllables emerge from articulatory factors. Specifically, it was hypothesized that the regular open-close motion of the jaw constrains segmental articulation in such a way as to yield specific phonological and phonetic patterns that form the basis of syllable perception.

In this dissertation, evidence was presented in support of the hypothesis that the jaw cycle provides a mechanical and temporal constraint on segmental articulation. It was argued that this constraint accounts for the patterns of segment sequencing and relative segment duration normally associated with syllables. In order to make this argument, the potential acoustic effects of articulatory dynamics were explored in depth. This exploration produced new and detailed hypotheses concerning the interaction between jaw movement and the movement of the lips and tongue.

Evidence was also presented in this dissertation in support of the idea that the constraint of the jaw cycle influences syllable perception, as defined by syllable boundary judgments. The evidence suggested that a production-based approach to the problem of syllable perception is preferable to a phonological approach since syllables are perceived as a function of how they are produced. Thus, variability in syllable boundary perception, which constitutes one of the major obstacles to a linguistic theory of the syllable, is accounted for in an approach that sees the syllable emerging from the interaction between the supraglottal articulators.

Finally, it was shown that the hypothesized constraint of the jaw cycle may also influence syllable structure. Phonologists have identified the most and least frequent syllable types across languages. Although phonologists attempt to explain them, the cross-language patterns have usually been ignored by phoneticians. One strength of the present phonetic account of the syllable is that it may provide insight into why certain syllable types are more frequent than others in the world's languages.

Bibliography

Ainsworth, W.A. (1986). Pitch change as a cue to syllabification. *Journal of Phonetics*, **14**, 257-264.

Anderson, S. and Port, R. (1992). Evidence for the syllable structure, stress and juncture from segmental durations. *Journal of Phonetics*, **22**, 283-315.

Bell, A. and Hooper, J. Bybee (1978). Issues and evidence in syllabic phonology. In *Syllables and Segments*, Bell and Hooper (eds.), Amsterdam: North Holland Publishing, pp. 3-22.

Blevins, J. (1995). The syllable in phonological theory. In *The handbook of phonology*, Goldsmith (ed.), Cambridge: Blackwell, pp 206-244.

Boucher, V.J. (1988). A parameter of syllabification for VstopV and relative-timing invariance, *Journal of Phonetics*, **16**, 299-326.

Brodda, B. (1979). Nagot om de svenska ordens fonotax och morfotax, *Papers form the Institute of Linguistics, University of Stockholm*, **38**.

Broselow (1979). Cairene Arabic syllable structure, *Linguistic Analysis*, **5**, 542-582.

Butt, M. (1992). Sonority and the explanation of syllable structure. *Linguistische Berichte*, **137**, 45-67.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, **24**, 209-244.

Byrd, D. and Tan, C.C. (1996). Saying consonant clusters quickly. *Journal of Phonetics*, **24**, 263-282.

Campbell, W.N., Isard, S.D. (1991). Segment duration in a syllable frame, *Journal of Phonetics*, **19**, 37-47.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English.* Harper Row: New York.

Clements, G.N. (1990). The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I: Between the Grammar and the Physics of speech*, Kingston, J. and Beckman, M. (eds.), Cambridge: University Press, pp. 283-333.

Clements, G.N. and Keyser, S.J. (1983). *CV Phonology: A Generative Theory of the Syllable.* MIT Press: Cambridge.

Davis, B.L. and MacNeilage, P.F. (1995). The articulatory basis of babbling, *Journal of Speech and Hearing Research*, **38**, 1199-1211.

Davis, B.L., MacNeilage, P.F., Matyear, C.L. (1999). *Acquisition of serial complexity in speech production: phonetic patterns in first words*, (submitted).

Derwing, B.L. (1992). A 'pause-break' task for eliciting syllable boundary judgments from literate and illiterate speakers: preliminary results for five diverse languages. *Language and Speech*, **35**, 219-235.

Erickson, D., Lenzo, K., Fujimura, O. (1994). Manifestations of contrastive emphasis in jaw movement, *Journal of the Acoustical Society of America*, **95**:2822.

Fant, G., Kruckenberg, A., Nord, L. (1991). Durational Correlates of Stress and Swedish, French, and English, *Journal of Phonetics*, **19**, 351-365.

Fromkin, V.A. (1968). Speculations on performance models. *Journal of Linguistics*, **4**, 47-68.

Fry, D.B. (1964). The functions of the syllable. *Zeitschrift fur Phonetik, Sprachwissenschaft und Kommunikationsforschung*, **17**, 215-237.

Gracco, V.L. (1988). Timing factors in the coordination for speech motor activity, *Journal of Neuroscience*, **8**, 4628-4634.

Gracco, V.L. (1994). Some organizational characteristics of speech movement control, *Journal of Speech and Hearing Research*, **37**, 4-27.

Greenberg, S. (1997) On the origins of speech intelligibility in the real world, *Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels*, Pont-a-Mousson, France, pp. 23-32.

Haggard, M. (1973). Correlations between successive segment durations: values in clusters, *Journal of Phonetics*, **1**, 111-116.

Harrington, J., Fletcher, J., Roberts, C. (1995). Coarticulation and the accented / unaccented distinction: evidence from jaw movement data, *Journal of Phonetics*, **23**, 305-322.

Harris, R. (1991). *Reading Saussure: A critical commentary on the "Cours de linguistique générale"*. La Salle, IL: Open Court.

Hauser, M.D. and Fowler, C.A. (1992). A fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates, *Journal of the Acoustical Society of America*, **91**, 363-369.

Henderson, J.B. and Repp, B. (1982). Is a stop consonant released when followed by another stop consonant? *Phonetica*, **39**, 71-82.

Hoard, J.E. (1966). Juncture and syllable structure in English, *Phonetica*, **15**, 96-109.

Hooper, J. B. (1976). *Introduction to Natural Generative Phonology.* Academic Press: New York.

Jakobson, R. and Waugh, L.R. (1979/1987). *The Sound Shape of Language.* New York: Mouton de Gruyter.

Kaye, J., Lowenstamm, J., and Vergnaud, J.-R. (1990). Constituent structure and government in phonology, *Phonology*, **7**, 193-231.

Keating, P.A., Lindblom, B., Lubker, J., Kreiman, J. (1994). Variability in jaw height for segments in English and Swedish VCVs, *Journal of Phonetics*, **22**, 407-422.

Kelso, J.A.S., Vatiokiotis-Bateson, E., Saltzman, E.L., 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.

Kenstowicz, M. (1994). *Phonology in generative grammar.* Cambridge, MA: Blackwell.

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.

Kozhevnikov, V.A. and Chistovich, L.A. (1965). *Speech articulation and perception.* Joint Publications Research, 30, 543.

Kuehn, D.P. and Moll, K.L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, **4**, 303-320.

Ladefoged, P. (1967). *Three areas of experimental phonetics.* London: Oxford.

Ladefoged, P. (1993). *A course in phonetics.* New York: Jarcourt Brace Javanovich.

Ladefoged, P., Draper, M.H., Whitteridge, D. (1958). Syllables and stress, *Miscellanea Phonetica*, **3**, 1-14.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Lehiste, I. (1977). Isochrony reconsidered, *Journal of Phonetics*, **5**, 253-265.

Lindblom, B. (1967). Vowel duration and a model of lip-mandible coordination, *Speech Transmission Lab. Prog. Status Report*, **4/1967**, 1-29, Royal Institute of Technology, Stockholm.

Lindblom, B. (1983). Economy of speech gestures. In *The Production of Speech* MacNeilage (ed), New York: Springer-Verlag, pp. 217-246.

Lindblom, B. (1996). Role of articulation in speech perception: clues from production. *Journal of the Acoustical Society of America*, **99**, 1683-1692.

Lindblom, B. and Maddieson, I. (1988). Phonetic universals in consonant systems. In *Language, speech and mind*, Hyman and Li (eds.). Routledge.

Lindblom, B. and Rapp, K. (1973). Some temporal regularities of spoken Swedish, *Publ. No. 21, Institute of Linguistics, University of Stockholm.*

MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production, *Brain and Behavioral Science*, **21**, 499-546.

MacNeilage, P.F. and Davis, B.L. (1990). Acquisition of speech production: Frames then content. In *Attention and Performance XIII: Motor Representations and Control*, Jeannerod, (ed.), New Jersey: Erlbaum.

MacNeilage, P.F., Davis, B.L., and Matyear, C.L. (1997). Babbling and first words: Phonetic similarities and differences. *Speech Communication*, **22**, 269-277.

MacNeilage, P.F., Davis, B.L., Kinney, A., and Matyear, C.L. (1999). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development*, in press.

Maddieson, I. (1985). Phonetic cues to syllabification, In *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, Fromkin (ed.), Orlando: Academic Press, pp. 203-221.

Malecot, A. (1968). The force of articulation of American stops and fricatives as a function of position. *Phonetica*, **18**, 95-102.

Moon, S. and Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowel. *Journal of the Acoustical Society of America*, **96**, 40-55.

Munhall, K., Fowler, C., Hawkins, S., Saltzman, E. (1992). 'Compensatory shortening' in monosyllables of spoken English, *Journal of Phonetics*, **20**, 225-239.

Munhall, K.G., Ostry, D.J., Flanagan, J.R. (1991). Coordinate spaces in speech planning, *Journal of Phonetics*, **19**, 293-307.

Nelson, W., Perkell, J. and Westbury, J. (1984). Mandible movements during increasingly rapid articulations of single syllables: preliminary observations. *Journal of the Acoustical Society of America*, **75**, 945-951.

Ohala, J. (1975). The temporal regulation of speech. In *Auditory Analysis and Perception of Speech*, Fant and Tatham (eds.), London: Academic Press, pp. 431-453.

Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English, *Journal of the Acoustical Society of America*, **54**, 1235-1247.

Ostry, D.J. and Munhall, K.G. (1985). Control rate and duration of speech movements. *Journal of the Acoustical Society of America*, **77**, 640-648.

Perkell, J.S. (1969). *Physiology of speech production.* Cambridge, MA: MIT Press.

Price, P.J. (1980). Sonority and syllabicity: acoustic correlates of perception, *Phonetica*, **37**, 327-343.

Redford, M.A. and Diehl, R.L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables, *Journal of the Acoustical Society of America*, in press.

Saussure, F. (1959). *Course in general linguististics.* New York: Philosophical Library.

Sigurd, B. (1973). Maximum rate and minimal durations of repeated syllables, *Language and Speech*, **16**, 373-395.

Stetson, R.H. (1951). Motor phonetics: A study of speech movements in action. Amsterdam: North Holland Publishing.

Stone, M. (1981). Evidence for a rhythm pattern in speech production: observations of jaw movement, *Journal of Phonetics*, **9**, 109-120.

Stone, M. and Vatikiotis-Bateson, E. (1995). Trade-offs in tongue, jaw, and palate contributions to speech production, *Journal of Phonetics*, **23**, 81-100.

Straka, G. (1979). *Les sons et les mots.* Strasbourg: Klincksieck.

Sussman, H.M., Bessell, N., Dalston, E., and Majors, T. (1997). An investigation of stop place articulation as a function of syllable position: a locus equation perspective. *Journal of the Acoustical Society of America*, **101**, 2826-2838.

Sussman, H.M., MacNeilage, P.F., Hanson, (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, **16**, 397- 420.

Treiman, R. and Danis, C. (1988). Syllabification of intervocalic consonants, *Journal of Memory and Language*, **27**, 87-104.

Treiman, R., Gross, J., Cwikiel-Glavin, A. (1992). The syllabification of /s/ clusters in English, *Journal of Phonetics*, **20**, 383-402.

Tuller, B. and Kelso, J.A.S. (1991). The production and perception of syllable structure, *Journal of Speech and Hearing Research*, **34**, 501-508.

Vennemann, T. (1972) On the theory of syllabic phonology. *Linguistic Berichte*, **18**, 1-18.

Willerman, R. (1994). *The phonetics of pronouns: articulatory bases of markedness.* Ph.D. Thesis, University of Texas, Austin.

Vita

Melissa Annette Redford was born to Hamish Stephen Redford and Denise Marie Lauzon Redford in Vancouver, British Columbia on December 15th 1969. After completing her primary and secondary school education in San Diego, California, she entered the University of California at Berkeley and completed a B.A. in Anthropology in 1992. She began graduate school at the University of Texas in Austin in 1993 and received a M.A. in Linguistics in 1995. In September 1995, she began work on a Ph.D. in Psychology, which was completed in December 1999.

Permanent Address: 4006 Avenue G, Apt. A
Austin, Texas 78712
USA
`mredford@utxsvs.cc.utexas.edu`

This dissertation was typeset with LaTeX 2_ε[2] by the author.

---

[2] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin.