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:

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

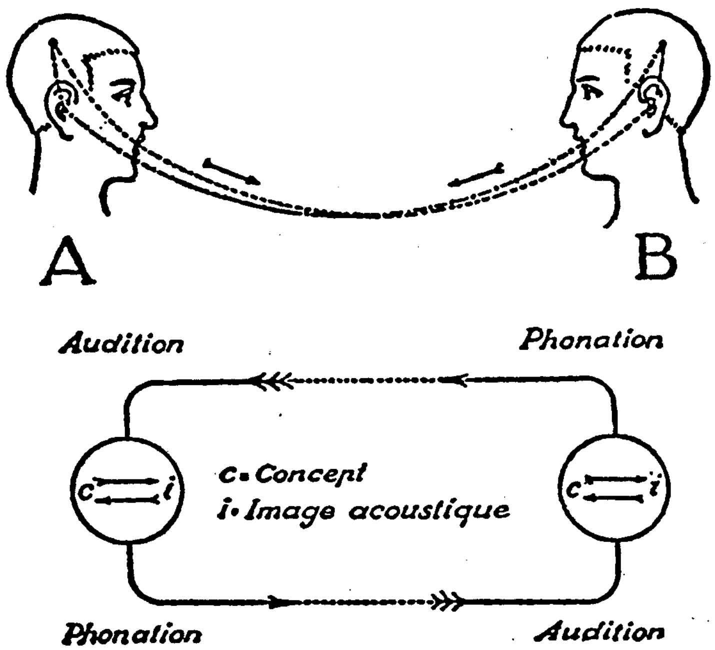

Figure 2.1: **Saussure's circuit of speech.** Language is shown to exist in the space between a speaker and listener. Though knowledge of the language resides in the heads of the interlocutors, language is also always tied to the production and perception systems.

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

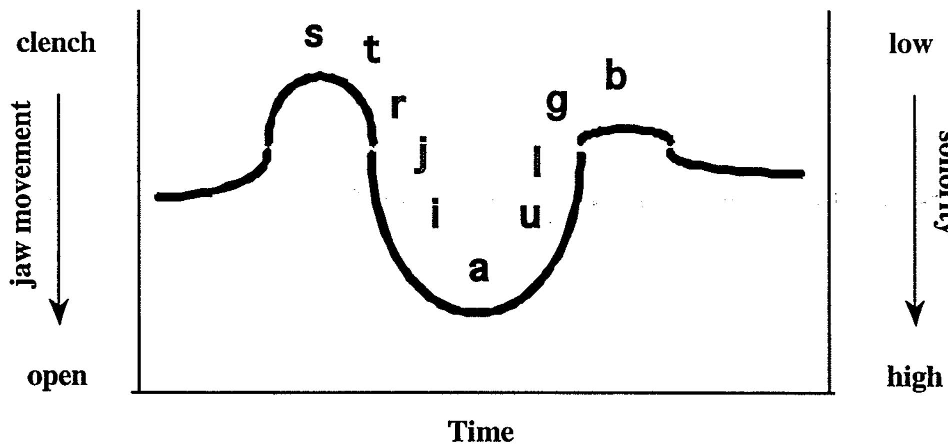

Figure 3.1: **Correspondence between jaw openness and sonority scale.** Schematic of Lindblom's (1983) results, which show a relationship between the sonority hierarchy and relative jaw opening.

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.

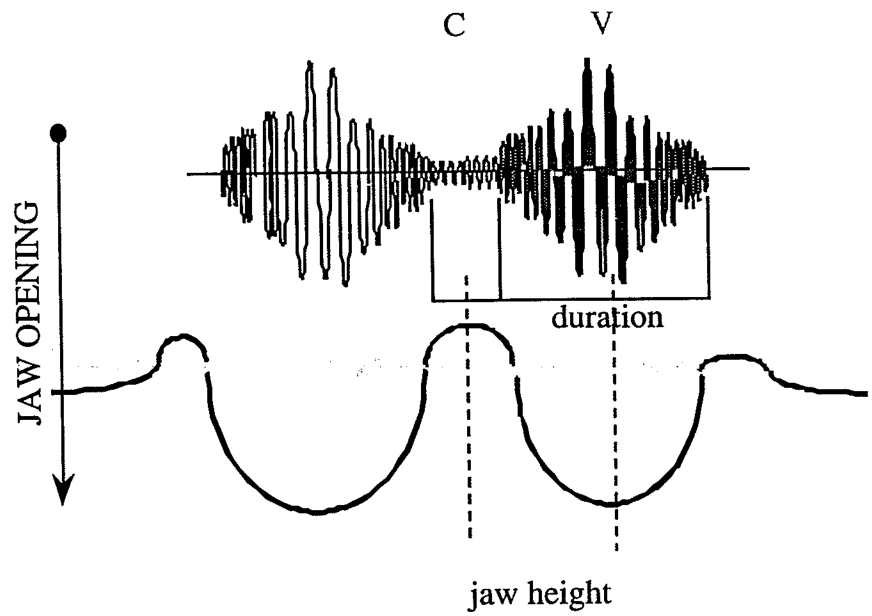

Figure 3.2: **Illustration of measurement locations.** Acoustic and movement waveforms were aligned. Jaw openness measurements were taken at the absolute midpoint of the acoustic segment.

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.

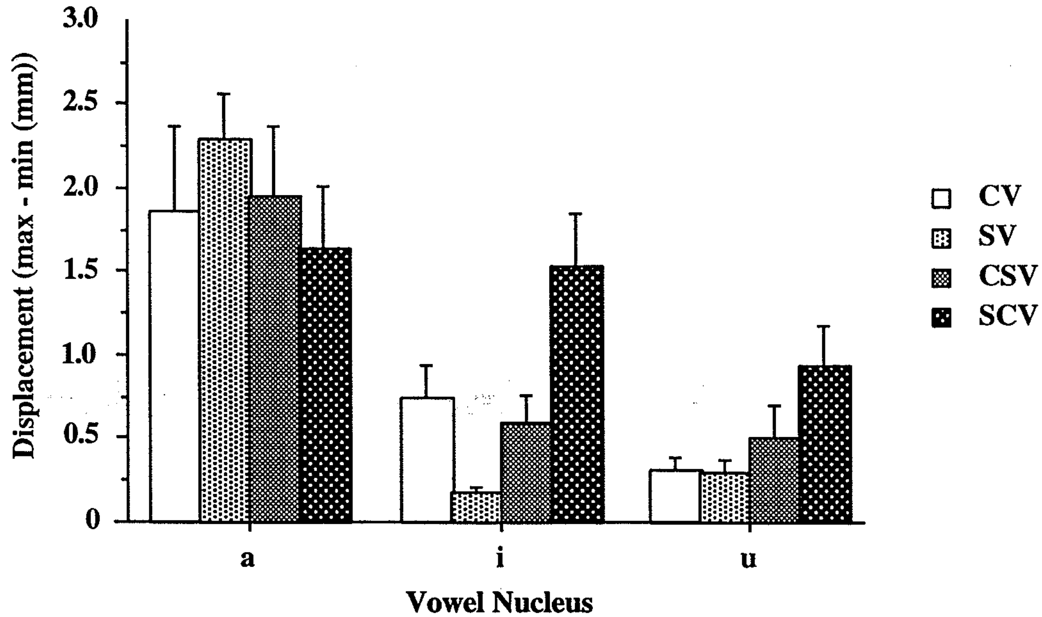

Figure 3.3: **Displacement as a function of syllable type and vowel nucleus..** Displacement was measured during the opening phase of the jaw cycle as the distance (in millimeters) between minimum and maximum jaw opening. The four syllable types were the consonant-vowel (CV) syllables, the sonorant-vowel (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.

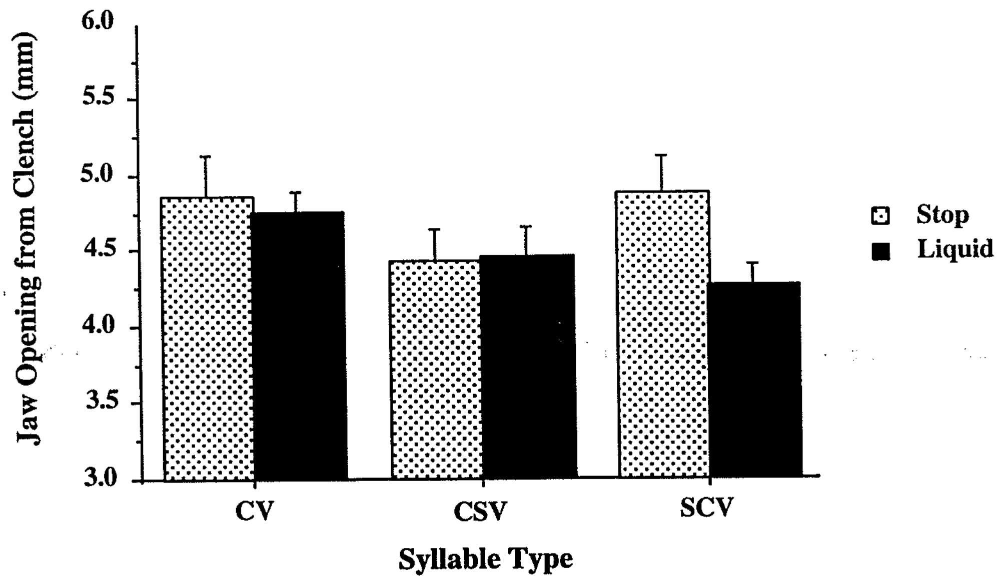

Figure 3.4: **Jaw opening for consonants as a function of syllable type..** Jaw opening was measured from clench (in millimeters) for the stop and liquid consonants of the four syllable types. The four syllable types were the consonant-vowel (CV) syllables, the sonorant-vowel (SV) syllables, the syllables 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.

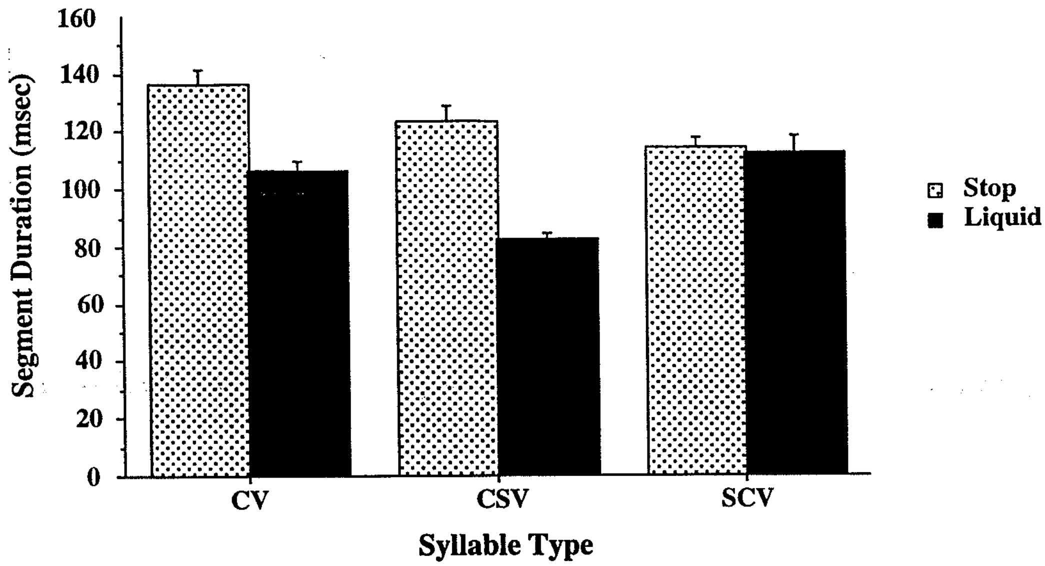

Figure 3.5: **Consonants duration as a function of syllable type.** Acoustic duration was measured (in milliseconds) for the stop and liquid consonants of the four syllable types. The four syllable types were the consonant-vowel (CV) syllables, the sonorant-vowel (SV) syllables, the syllables with normally-sequenced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