INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 8HR
AN ACOUSTIC CORRELATE OF SYLLABICITY IN ENGLISH

DISSERTATION

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

By
Holly Semiloff, B.A., M.A.

*****

The Ohio State University
1977

Reading Committee: Approved By
Catherine A. Callaghan Ilse Lehiste
David L. Stampe Arnold M. Zwicky

Adviser
Department of Linguistics
ACKNOWLEDGEMENTS

Although it sounds like a cliche, this dissertation would not have been possible without the help, both technical and advisory, of a large number of people. First, I would like to thank my subjects for the production experiment who sat for hours in an echo-free chamber reading paragraphs and making up stories: Robert Kantor and Gregg Eiden at Ohio State and Robert Ewald at Indiana University. I also am grateful to Dr. J.R. Mundie of the Aerospace Medical Research Laboratory at Wright Patterson Airforce Base in Dayton, Ohio, who allowed me to use his Rockland Digital Speech Synthesizer; and to Zinny Bond (currently at Ohio University) who patiently guided me through the intricacies of the elusive Rockland. Thanks also to Dr. A.M. Liberman who made the synthesizer facilities at Haskins Laboratories available to me and in particular, to Gary Kuhn, who in fact made the final stimuli. I am also grateful to Dr. David Pisoni and James Sawusch at Indiana University who allowed me to use the Psychology Department's subject pool for my identification experiment; to Dr. Rita Naremore of the Speech and Hearing Department who made the listening room
and recording facilities available for running subjects; and to Dr. Donald Robinson who provided much needed help with statistics. Thanks also to Dr. June Shoup of the Speech Communications Research Laboratory in Santa Barbara, California who made available computer facilities for preparing my discrimination tapes and in general provided a most congenial place to work. Particular thanks to Hideki Kasuya for help with the discrimination tapes. Thanks to my discrimination subjects, Don Bollinger, Judy Cherry and Mary Mines, of SCRL, who patiently put up with some very boring listening sessions. I also express gratitude to Victoria Fromkin, Peter Ladefoged, Breyna Moskowitz and the students at the UCLA Phonetics Laboratory for two hours of extremely stimulating discussion of my research. Finally, my dissertation committee — Catherine Callaghan, David Stampe and Arnold Zwicky — asked me probing, long-term questions and my goal, now that I have finished my degree, is to answer them. My thanks to my advisor, Dr. Ilse Lehiste, I have saved for last. Her support, encouragement and belief in the value of scientific research kept me going when my instincts told me to chuck it all.

This work was supported by NSF S0075-10043 as well as the Advanced Research Project Agency and the Office of Naval Research N00014-73-C-0221.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. REVIEW OF THE LITERATURE ON SYLLABLES AND SYLLABICITY</td>
<td>5</td>
</tr>
<tr>
<td>0. Introduction</td>
<td>5</td>
</tr>
<tr>
<td>1. Phonetic Syllabicity</td>
<td>7</td>
</tr>
<tr>
<td>1.1 Sonority</td>
<td>7</td>
</tr>
<tr>
<td>1.2 Chest Pulse Theory</td>
<td>10</td>
</tr>
<tr>
<td>1.3 Degrees of Opening</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Characteristics of CV Sequences</td>
<td>14</td>
</tr>
<tr>
<td>1.5 Syllabic Consonants</td>
<td>17</td>
</tr>
<tr>
<td>1.6 Summary</td>
<td>21</td>
</tr>
<tr>
<td>2. Phonetic Syllable Boundaries</td>
<td>22</td>
</tr>
<tr>
<td>2.1 Acoustic Evidence for Syllable Boundaries</td>
<td>22</td>
</tr>
<tr>
<td>2.2 Articulatory Evidence for Syllable Boundaries</td>
<td>25</td>
</tr>
<tr>
<td>2.2.1 Implosive versus Explosive Boundaries</td>
<td>25</td>
</tr>
<tr>
<td>2.2.2 Extrinsic Allophones</td>
<td>28</td>
</tr>
<tr>
<td>3. Phonological Syllabicity</td>
<td>36</td>
</tr>
<tr>
<td>4. Phonological Syllable Boundaries</td>
<td>38</td>
</tr>
<tr>
<td>4.1 Distributional Constraints</td>
<td>45</td>
</tr>
<tr>
<td>4.2 Phonological Rules and Syllable Boundaries</td>
<td>50</td>
</tr>
<tr>
<td>5. Rhythmic Syllabicity</td>
<td>55</td>
</tr>
<tr>
<td>6. Summary</td>
<td></td>
</tr>
<tr>
<td>II. SYLLABICITY IN ENGLISH: PRODUCTION</td>
<td>57</td>
</tr>
<tr>
<td>0. Introduction</td>
<td>57</td>
</tr>
<tr>
<td>1. Experimental Design</td>
<td>57</td>
</tr>
<tr>
<td>2. Choice of Key Words</td>
<td>60</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1  Duration (in msec) of Key Words in the Four Speech Styles ............. 63
Table 2.2  Average Duration (in msec) of Segments in the Key Words .................. 66
Table 2.3  Confusion Matrix, All Responses, All Subjects ................................. 77
Table 3.1  Average Crossover Durations for the Key Words ................................. 91
Table 3.2  Stimulus Differences for the Key Words ........................................ 92
Table 3.3  Speech Style Differences for the Key Words ..................................... 93
Table 3.4  Trial Differences for the Key Words ............................................. 101
Table 3.5  Average Percent Correct and Standard Deviations for the 3 Subjects, One-Step Discrimination ............................. 116
Table 3.6  Analysis of Variance, One-Step Discrimination ................................. 117
Table 3.7  Average Percent Correct and Standard Deviations for the 3 Subjects, Two-Step Discrimination ............................ 125
Table 3.8  Analysis of Variance, Two-Step Discrimination ................................. 126
Table 3.9  Average Percent Correct and Standard Deviations for the 3 Subjects, Three-Step Discrimination ............................ 131
Table 3.10  Analysis of Variance, Three-Step Discrimination .............................. 132
Table 4.1  Average Crossovers (in msec), Five Speech Styles .............................. 141
Table 4.2  Speech Style Guessing Averages (in msec) .................................... 146
Table 4.3  F₂ Onset Frequencies for Liquids (In Hz) ...................................... 148
Table 4.4  Average Crossover Durations (in msec), 3 Subjects ............................. 156
Table 4.5  Average Crossover Durations (in msec) for the 6 Key Words, All Groups ............. 167
Table 4.6  Analysis of Variance Among the 6 Key Words ................................. 168
<table>
<thead>
<tr>
<th>Figure 2.1</th>
<th>Percent Correct for Style and Rate</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.2</td>
<td>Percent Correct for Style-Rate Combinations</td>
<td>73</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Crossover, SPORT-SUPPORT, All Groups</td>
<td>88</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Crossover, BLOW-BELOW, All Groups</td>
<td>94</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Crossover, PLIGHT-POLITE, All Groups</td>
<td>95</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Crossover, BROKE-BAROQUE, All Groups</td>
<td>96</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Crossover, DRESS-DURESS, All Groups</td>
<td>97</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Crossover, CREST-CARESSED, All Groups</td>
<td>98</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>Crossover, PRAVED-PARADE, All Groups</td>
<td>99</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>Trial Differences, BLOW-BELOW</td>
<td>102</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>Trial Differences, PLIGHT-POLITE</td>
<td>103</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>Trial Differences, BROKE-BAROQUE</td>
<td>104</td>
</tr>
<tr>
<td>Figure 3.11</td>
<td>Trial Differences, DRESS-DURESS</td>
<td>105</td>
</tr>
<tr>
<td>Figure 3.12</td>
<td>Trial Differences, CREST-CARESSED</td>
<td>106</td>
</tr>
<tr>
<td>Figure 3.13</td>
<td>Trial Differences, PRAVED-PARADE</td>
<td>107</td>
</tr>
<tr>
<td>Figure 3.14</td>
<td>ABA One-Step Discrimination, All Subjects</td>
<td>112</td>
</tr>
<tr>
<td>Figure 3.15</td>
<td>ABB One-Step Discrimination, All Subjects</td>
<td>113</td>
</tr>
<tr>
<td>Figure 3.16</td>
<td>BAA One-Step Discrimination, All Subjects</td>
<td>114</td>
</tr>
<tr>
<td>Figure 3.17</td>
<td>BAB One-Step Discrimination, All Subjects</td>
<td>115</td>
</tr>
<tr>
<td>Figure 3.18</td>
<td>ABA Two-Step Discrimination, All Subjects</td>
<td>120</td>
</tr>
<tr>
<td>Figure 3.19</td>
<td>ABB Two-Step Discrimination, All Subjects</td>
<td>121</td>
</tr>
<tr>
<td>Figure 3.20</td>
<td>BAA Two-Step Discrimination, All Subjects</td>
<td>122</td>
</tr>
<tr>
<td>Figure 3.21</td>
<td>BAB Two-Step Discrimination, All Subjects</td>
<td>123</td>
</tr>
<tr>
<td>Figure 3.22</td>
<td>ABA Three-Step Discrimination, All Subjects</td>
<td>127</td>
</tr>
<tr>
<td>Figure 3.23</td>
<td>ABB Three-Step Discrimination, All Subjects</td>
<td>128</td>
</tr>
<tr>
<td>Figure 3.24</td>
<td>BAA Three-Step Discrimination, All Subjects</td>
<td>129</td>
</tr>
<tr>
<td>Figure 3.25</td>
<td>BAB Three-Step Discrimination, All Subjects</td>
<td>130</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Speech Style Acceptance Table</td>
<td>144</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>ABA One-Step Discrimination, Predicted and Obtained Values, All Subjects</td>
<td>157</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>ABA Two-Step Discrimination, Predicted and Obtained Values, All Subjects</td>
<td>158</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>ABA Three-Step Discrimination, Predicted and Obtained Values, All Subjects</td>
<td>159</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Differences Among the Key Words After Duncan's Multiple Range Test</td>
<td>169</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Voicing of Preceding Stops and Crossover Durations</td>
<td>169</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Place of Articulation of Preceding Stops and Crossover Durations</td>
<td>170</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Amplitude of 3LOW-BELOW, Short /l/</td>
<td>174</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Amplitude of 3LOW-BELOW, Long /l/</td>
<td>175</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Amplitude of CREST-CARESSED, Short /r/</td>
<td>176</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Amplitude of CREST-CARESSED, Long /r/</td>
<td>177</td>
</tr>
</tbody>
</table>
INTRODUCTION

The goals of this dissertation arose from three areas of interest: (1) the place of the syllable in phonological theory, (2) the phonetic nature of syllabic ity, and (3) the categorical perception model (Liberman, et al., 1957). From these concerns, three hypotheses emerged; and they form the bases for the experimental research described below. (1) The suprasegmental feature, duration, is a cue to the perception of syllabic ity. (2) The use of a precursor phrase meant to suggest a particular speech style can shift listeners' perception from a monosyllabic to a disyllabic word (and vice versa). (3) The perception of syllabic ity is categorical.

A brief design of the research is the following: I chose seven minimal pairs, words that differed only in the number of syllables, such as BLOW-BELOW and BROKE-BAROQUE. I synthesized the monosyllabic member of each pair and then lengthened the steady state portion of the /r/ or /l/ in each word, in increments of 10 msec. I ran a forced choice identification test using a large number of subjects to determine whether the labeling responses would show an abrupt crossover from the monosyllabic to the disyllabic
member as the duration of the liquid increased. In addition, a precursor phrase suggesting a particular speech style was inserted before each stimulus. The styles were Formal Slow, Formal Fast, Casual Slow and Casual Fast. There were five different labeling tapes — one for each speech style and one to be used as a control, with no precursor at all. Because crossovers were obtained, indicating that syllabic ity is perceived categorically, an ABX discrimination test was run using stimuli one, two, and three steps apart along the acoustic continuum. But discrimination results did not show peaks of discriminability at the 'syllabic ity boundary' as suggested by the labeling test and it was concluded that the perception of syllabic ity is not categorical, but more or less continuous. Thus hypothesis (3) was not accepted. Hypothesis (2) was also rejected, since there were found to be no significant differences among the five speech style groups. But the first hypothesis, that duration is a cue to syllabic ity, was accepted based on the perceptual crossovers that were obtained on the identification test.

Chapter I is a general review of some of the most important literature on syllables and syllabic ity. The review is hardly definitive, but in one way or another the research and theories described have all served in the most fruitful intellectual capacity — that of stimulating more research and revised theories. I have divided the review
into five categories as a general taxonomy: phonetic syllabiccity, phonetic syllable boundaries, phonological syllabiccity, phonological syllable boundaries and rhythmic syllabiccity.

In Chapter II, I describe a production experiment, designed to elicit the minimal pairs in the four speech styles mentioned above. The purpose was to investigate if and how speakers differentiate the speech styles in production. In addition, I describe a perception experiment whose purpose was to test whether listeners could perceive differences among the four speech styles. I concluded that the Formal Slow and Casual Fast styles are readily perceived, but the mixed styles, Formal Fast and Casual Slow, are considerably more difficult. Finally, differences in rate (slow-fast) are easier to perceive than differences in style (formal-casual).

Chapter III includes the design and method of preparing the synthetic stimuli and setting up the identification and discrimination experiments. The results of the experiments are stated and a limited number of figures and tables describe the results graphically.

Chapter IV consists of an interpretation of the results of the labeling and discrimination tests. First, I state why there were no differences among the five speech style groups; essentially I argue that the task was too difficult. Second, I attempt to account for the discrepancy between the
labeling and discrimination tests, where I argue that the labeling results are not unambiguously categorical to begin with, and that the discrimination paradigm is a more potent measure of categorical perception than the labeling paradigm. I also provide a tentative reason for the differences in crossover durations among the key words based on some inherent differences between /r/ and /l/ perception. Third, I discuss the implications of this research for phonological theory and theories of speech perception. Although duration does seem to be a necessary and sufficient cue in the perception of syllabic liquids, I do not believe that syllabicity is phonemic in English. The reason is that there are phonological reasons for deriving syllabic liquids from underlying vowel + liquid. With regard to theories of speech perception, I do not think this research provides any real evidence for the syllable as the most basic unit of speech perception, as opposed to the segment. The reader is referred to a distinction drawn by Studdert-Kennedy (1976) between an acoustic and a linguistic unit of speech perception. I propose that the syllable is an example of the former, but not necessarily the latter.

In the last section of Chapter IV, I suggest additional research on the perception of syllabicity. This dissertation has raised many questions about ways to test hypotheses, experimental designs and methods, and most important, new questions to answer and old questions to re-answer.
Chapter I: REVIEW OF THE LITERATURE ON SYLLABLES AND SYLLABICITY

0. INTRODUCTION

The literature on syllables in phonetics and phonology is extensive and covers a wide range of topics, from the physical characteristics of syllabic ity to the place of the syllable in linguistic theory. From reading some of that literature, it seems that a five-way classification system emerges, and I will divide this review of the literature accordingly. Topics dealing with the syllable fall into the following categories: (1) phonetic syllabic ity -- the acoustic characteristics and articulatory movements that are required to produce syllable nuclei; (2) phonetic syllable boundaries -- the acoustic and articulatory characteristics that constitute syllable divisions; (3) phonological syllabic ity -- those syllabic units which are necessary to describe the prosodic features of language; (4) phonological syllable boundaries -- the interaction of syllable boundaries and phonological rules, and both universal and language-specific distributional constraints; and (5) rhythmic syllabic ity -- the syllable as a higher level linguistic unit, a unit of neurological programming.
The distinction between the phonetic and the phonological syllable has been emphasized by many researchers. Notions like 'every syllable must contain a vowel or vowel-like segment' characterize phonetics; to say that the syllable is a structural unit that is necessary in order to state distributional constraints on consonant clusters and which is required in order to state generalizations about phonological rules characterizes the phonological viewpoint. Less often has the distinction between syllabic ity and syllable boundaries been made. But surely it is possible to hold that while a vowel or vowel-like segment is necessary for syllabic ity, it is not always possible to determine phonetic syllable divisions phonetically. On the phonological side, it has been argued that syllable boundaries are crucial in stating some phonological rules but that the syllable is not needed for describing prosodic features like tone and quantity. Actually all combinations of opinions occur in the literature. In this chapter I will present some of the major contributions to studies of the syllable in accord with the five-way classification presented above. In so doing, it is hoped that different treatments of the syllable can be seen as reflecting different purposes of the various authors.
1. Phonetic Syllabicity
1.1 Sonority.

Defining phonetic syllabicity in terms of the presence of sonority amounts to saying that it is the relative loudness of a segment which makes it syllabic. Vowels are known to be inherently louder than consonants. Jones (1949) says sonority is an inherent quality of a sound and is its 'carrying power' or loudness. Sonority is distinguished from prominence, which refers to the combination of a sound's inherent quality plus some amount of length, stress and intonation. Jones argues that phonetic syllabicity is actually a peak of prominence. "In theory a syllable consists of a sequence of sounds containing one peak of prominence. In practice it is often impossible to define the limits of a syllable because there is no means of fixing any exact points of minimum prominence." (p. 55) (Jones is not alone in denying the existence of phonetic syllable boundaries; Jespersen (1904) also says that syllabicity is determined by degrees of sonority but it is not possible to precisely locate the boundaries. Likewise for Bloomfield (1933) and Pike (1943). Jones feels that some minimum of prominence is needed to separate syllable peaks. He argues that when two vowels are adjacent, a glide is inserted between them to act as a minimum of prominence. (This sounds as if there is a perceptual reason for inserting a glide between two vowels; it is also likely that the reason
has to do with articulatory timing: when two adjacent vowels are not tautosyllabic, the time taken to change directions can result in a glide of short duration.) Jones also says, "More sonorous sounds may become less prominent, and therefore more consonant-like, by diminishing length or stress, and... sounds of relatively small sonority may be made prominent by increasing length or stress." (p. 24) This implies that in a word like stop the /s/, which is relatively nonsonorous, can become more prominent by increasing its length. But in English it will never be prominent enough to become syllabic, so prominence alone cannot insure syllabicity. It seems there is an overriding phonological constraint in English--initial /sCl clusters always belong to the same syllable regardless of what one does to the /s/ phonetically. On the other hand when followed by a vowel, liquids and nasals do become syllabic when lengthened. (In fact this is the major hypothesis of the experiment to be described in subsequent chapters of this thesis.) Bloomfield (1933) also talks about syllabic resonants, but he attributes their syllabicicy to increased stress instead of duration. Although stress is probably best defined as degree of respiratory effort (Jones 1949, Lehiste 1970), an increase of such effort can have the secondary effects of increased loudness and greater

^Throughout the dissertation I will use // to enclose phonemic symbols and use [ ] only when allophonic information is included in the transcription.
duration. Bloomfield also deals with the /s/ problem although not very satisfactorily. He says that while the /s/ in text /tekst/ is more sonorous than the surrounding stops, the sonority on the vowel is so much greater that it 'drowns out' the sonority of the /s/. But since vowels are more sonorous than any other sound, why doesn't the first vowel in member [məmbər] 'drown out' the syllabicity of the [ə]?

Others who have identified sonority as the defining characteristic of phonetic syllabicity include deGroot (1926), Pike (1947), Hála (1960,1964) and Kurath (1964). According to Pike, phonetic syllables are "...units of one or more segments during which there is a single chest pulse and a single peak of sonority or prominence." (p. 60) (On the chest pulse theory, see below.) deGroot claimed that syllable nuclei have greater intensity than the margins, and that the presence of sonority always signals a syllable, although he actually contradicts himself by further claiming that some syllables may have more than one peak of sonority. If this is the case, sonority is necessary but not sufficient for identifying phonetic syllabicity. Furthermore, deGroot does not actually define sonority.

Definitions based on sonority are not universally accepted. Lebrun (1966) criticizes appeals to a 'sommet de sonorité.' His own experiments showed that listeners can't always tell how many peaks of sonority an utterance really
has, and instrumental analyses can't always demarcate an acoustic syllable because the transitions cannot always be analyzed in terms of peaks versus troughs. Therefore the phonetic syllable as a peak of sonority doesn't elucidate anything in linguistic studies. Lebrun also refers to the /s/ problem in stop and mentions Hála (1961), who says that /s/ isn't more sonorous than the /t/ -- it is only more perceptible. But this seems to beg the question since it is precisely the presence of sonority that leads to heightened perceptibility. Actually Lebrun's major complaint is that sonority has not been adequately defined. Its vagueness would allow final aspirated stops in French to be considered sonorous, and therefore syllabic, although no one would want to claim that they actually are syllabic.

Saussure also argues against sonority as the basis for syllabicity. He cites the vowels /i,u/ as having as much sonority as /y,w/, yet the glides are not syllabic. Likewise, the /s/ in /pst/ is syllabic, and it has no sonority to speak of; and finally /w/ is more sonorous than /l/ but in Indo-European *wlkos, 'wolf', it is the /l/ which is syllabic, not /w/.

1.2. Chest Pulse Theory

The person credited with proposing the chest pulse theory of phonetic syllabicity is R. H. Stetson (1951). His experiments led him to conclude that there is a single
chest pulse, meaning a contraction of the intercostal muscles, for every syllable in a word. But subsequent research by Ladefoged and others (1958, 1967) has led to a rejection of that theory. According to Ladefoged, "...there is no simple correlation between intercostal activity and syllables..." (1967, p. 20) He found that Stetson came to the wrong conclusions on several counts. (1) Stetson considered the action of the intercostal muscles to be ballistic; they either contracted or they didn't. But in fact, Ladefoged found that their tension varied in the amount and rate of change. For example there could be a single increase in tension by the intercostals over two-syllable words like pity and around, and there may be two bursts of activity in monosyllabic words like sport and stay, the first burst associated with the /s/. But it was found that muscular activity does correlate with the production of syllables that receive sentence stress. (2) Stetson said that in open syllables, the collapse of the lungs is checked by an active inspiratory effort by the external intercostals. But Ladefoged found that in normal conversational utterances there was no activity of the external intercostals in open syllables. (3) Stetson said

\footnote{Inferential evidence against the chest pulse theory is cited by Lehiste (1970). Apparently normal syllabication patterns occur in the speech of some patients who have a complete paralysis of the respiratory musculature and who breathe through an iron lung. Thus at least the strong version of the chest pulse theory, that a chest pulse occurs on every syllable, can't be correct.}
that in stressed syllables the action of the internal intercostals is reinforced by the abdominal muscles led by the rectus abdominis. But Ladefoged found this to be true only in very emphatic stressing and at the end of very long sentences. The main conclusion then is that the chest pulse theory only accounts for sentence stress. It is generally believed that Stetson's rather unsophisticated measurement techniques were responsible for most of his overzealous conclusions.

Other authors have been less quick to reject Stetson in theory if not in practice. Allen (1973) voices basic acceptance of the chest pulse theory even though in its specifics it was found to be wrong. Nonetheless the place to look for some physical correlate of the syllable is in the respiratory muscles. Likewise Pike (1967) argues that Ladefoged and his colleagues make the mistake of rejecting the chest pulse theory just because they can't find a chest pulse for every syllable. Like Allen, Pike defends Stetson by saying his theory is probably wrong in detail, but right in its insistence that there is a physiological basis for phonetic syllabicity. Pike says, "I feel we must assume a physiological basis for long vowels and for other perceived contrasts in syllable types." (p. 367n) Hála (1960) also emphasizes a phonatory definition. He combines the physical and acoustic parameters of syllabicity in saying that the nature of the syllable consists solely in the presence of
voice (sonority), which is controlled by the transition from the closure of the speech organs to their release1.

1.3. Degrees of Opening

Phonetic syllabicility has also been defined according to degrees of articulatory opening and closure. Grammont (1946), in discussing syllable boundaries, talks of 'degrés d'aperture' of various segment types and says that the 'point vocalique' of a syllable is the point at which the aperture changes from increasing to decreasing. He denies that the presence of a vowel or syllabic resonant is essential. For example, in the interjection /pst/, it is the point between the /s/ and the /t/ which constitutes the 'point vocalique'; in other words that is the place a vowel would occur if there were one. If a vowel were inserted in /pst/, the result would be /psit/, not */pist/. This is somewhat surprising since French does not allow initial /ps-/ clusters, but Grammont apparently does not want to consider distributional criteria, only purely phonetic.

---

1Hála and Rosetti engaged in an interesting polemic on syllabicility (Hála 1960, 1964, Rosetti 1959, 1961). Basically, Rosetti considers the presence of air as the defining feature, and Hála, the presence of voicing. Rosetti attempts to explain whispered speech (no voicing) and interjections like pst by saying that the passage of air is responsible for their syllabicility. But Hála discounts both whisper and interjections as representing a different level of speech (!) and subject to their own rules. Hála's argument is not very convincing since whispered speech has all the characteristics of phonated speech except vocal cord vibration, and surely it has syllables as well.
Another author who emphasizes the aperture theory is Sommerfelt (1931), who is in complete agreement with Grammont. In listing the five essential characteristics of the phonetic syllable, deGroot (1926) also says there must be closing followed by opening of the articulators, although it is hard to see how closure occurs in syllable-initial vowels. Pike (1967) uses the terms crest and trough for the open and closed portions of the syllable, respectively. The trough is the region of the heaviest air pressure, which is partly conditioned by the degree to which the air stream is closed off. Segments with the heaviest air pressure and which close off the air stream are nonsyllabic.

1.4. Characteristics of CV Sequences

Syllabicity has also been approached from the viewpoint of looking at the contrast between consonants and vowels. That is, syllabicity is defined by the inherent differences between these classes of segments. Jakobson and Halle (1956) say, "The pivotal principle of syllable structure is the contrast of successive features within the syllable." (p. 20) It is the relative contrast between optimal consonantality (a decrease of energy) plus optimal vocalicity (an increase of energy). Pushed to its logical conclusion, any sound in isolation would be syllabic because it is bounded by silence, which has no energy at all. This would account for the /ʃ/ in sh! and the /s/ is pstl
functioning as syllabic nuclei. But it doesn't explain why /s/ in stop isn't syllabic since it has more energy than the silence preceding and the /t/ following. Jakobson and Halle do say that if the syllable margins contain an inherently louder sound in less loud surroundings, its loudness is reduced to preserve the unity of the syllable. Presumably that is how they would explain the non-syllabic nature of /s/ in stop. (This recalls the remarks of Bloomfield, mentioned above, that the vowel in text 'drowns out' the sonority of the /s/.) Acoustically then, the nucleus exceeds the margins in intensity and often in increased fundamental frequency. While the crest is usually a vowel, it may also be a liquid, nasal, and occasionally a fricative. (See below on syllabic sonsonants.)

Some interesting research has been done by Bondarko (1969) relating to the primacy of CV over VC syllables. She argues that CV syllables are more common because they are not symmetrical with VC syllables articulatorily or acoustically, and that there is a greater assimilation of elements in CV types. Specifically, the coarticulation (e.g. rounding) of C and V is due to specific activity of the articulatory organs, but what assimilation there is of a V on a following C is due to inertia of the articulators. Acoustically, where formant patterns show up on consonants,

---

1What I have called the syllable nucleus both Pike and Jakobson and Halle call the syllable crest. Where I have used margin, Pike uses trough and Jakobson and Halle use slope.
particularly sonorants, they often coincide with the higher formants of the following vowel, but there is little dependence of the formant structure of a consonant on a preceding vowel. (An exception would have to be made for vowel-nasal sequences where there typically is nasalization on the vowel.) She also claims that there is a duration dependency such that syllable-initial consonants are shorter before open vowels and longer before close vowels, but that the duration of syllable-final consonants is not affected by the preceding vowel. Voicing differences in consonants are also weakened in VC syllables. A voiceless consonant can be voiced for up to 60% of its duration and still be perceived as voiceless, indicating that voicing differences are not maximized as they are in CV syllables. This fact would seem to show an assimilation effect of a vowel on the following consonant since the voicing of the vowel often spreads to the consonant. This of course doesn't happen in CV syllables. Finally Bondarko says that the actual articulatory movements for CV syllables involve only closure + release, but the movements for VC syllables involves

---

Bondarko's work is the only place I am aware of where consonantal length is said to be determined by the quality of a following vowel, although she presents experimental evidence to support that claim. Related research on the perception of stop-voicing and duration (Summerfield, 1976) showed that lengthening the duration of the vowel following voice onset of the preceding stop increased the probability of the stop being perceived as voiced. That is, vowel duration influenced the interpretation of the consonant voicing cue, the VOT. But that study referred to vowel duration, not open versus close vowels.
closure + release + closure + release. This analysis is somewhat surprising since 'closure' usually means approximation by the oral articulators, and is not used to describe vowel articulations. Also, 'release' implies a previous closure, and in the transition from vowel to consonant one does not typically speak of articulatory closure or release. Regarding this relationship, Öhman (1966) has shown that the gross movements of the tongue are all toward vowel positions and that consonant articulations are actually just super-imposed on those gross movements.

Looking at syllabic from a perceptual standpoint, Whitney (1867) had the following to say, "This property, syllabic, the foundation of which is in the ear of the hearer rather than in the mouth of the speaker, depends upon the antithesis of the opener and closer sounds composing the word: the comparatively open and resonant vowels strike the ear as the prominent and principal constituents of the series, while the closer consonants appear as their adjuncts, separating at the same time that they connect them." (p. 89) (underlining mine, HS)

1.5. Syllabic Consonants

That glides, liquids, and nasals can function as syllable nuclei is well-known. All three groups are like vowels in that their articulation allows the relatively free passage of air through one or more resonance cavities,
orally for glides and liquids, nasally for nasals. (When liquids and glides are nasalized, both cavities are used.) They also exhibit a more-or-less clearly defined formant structure, like vowels. Syllabic obstruents on the other hand pose the greatest threat to traditional theories of phonetic syllabic structure - they do not typically exhibit the sonority, articulatory opening, or formant structure associated with syllable nuclei. Bell (1969, 1970a) has described several languages that apparently do have syllabic obstruents. Some of these are the Central Sudanic language Lendu, the Amerindian languages Navaho and Acoma, South Kyuushuu Japanese, Luganda, the Berber language Shilba, and several Chinese dialects. There seems to be an implicational hierarchy for syllabic obstruents such that if a language has syllabic stops or affricates, it also has syllabic fricatives, and there is an overriding preference for /s/ to be syllabic among fricatives¹. Among stops, if a language has voiceless syllabic stops, it also has voiced syllabic stops. Voiceless stops are the most extreme nonsyllabic segments. Their articulation consists of a period of silence followed by a release, and the acoustic cue to their place of articulation lies in the frequency

¹Bell has other things to add about syllabic obstruents. There is a tendency for them to occur in grammatical particles and affixes rather than in stems and in unstressed position. An example of a syllabic voiced stop is the English word probably, which in the speech of some is [præbbli], and of a voiceless stop, Daniel Jones' example of [ˌkuː], thank you.
range of the burst and the transition to the following sound. In order to maintain voicing for a voiced stop, some supraglottal adjustment must take place, such as expanding the pharyngeal walls or lowering the larynx, both of which enlarge the supraglottal cavity and decrease the air pressure within it. Since syllabic identity requires at the very least some period of time, if the supraglottal adjustment occurs immediately upon closure, a voiced stop can be maintained for some time, and thus, be syllabic. Voiceless stops on the other hand involve no such adjustment and their syllabic identity can only be found in their release. In languages like Bella Coola, said to have syllables without vowels (Newman 1947), there is always a relatively long period of release, or 'transitional vocoid' (Bell 1970b) present. Syllabic fricatives, being continuants, need only to extend the duration of the fricative to meet the requirement of syllabic identity. It should be added though that duration itself is not sufficient for an obstruent to be syllabic, as the example of /s/ in stop indicated. Ultimately phonetic syllabic identity may have to be defined from two articulatory perspectives: true consonants, where duration is the necessary and sufficient factor, and sonorants and vowels, for which duration is sufficient but not necessary. (A more detailed investigation into the relationship between duration and the perception of syllabic identity is the subject of the experiment to be described
Writing about syllabic consonants in Polish, Rubach (1974) argues that one way of getting around the problems associated with sonority and degree of aperture as criteria for phonetic syllabic is to say that a consonant must be both lengthened and more sonorous than one of its environments in order to be syllabic. (Rubach does not specify whether by 'lengthened' he means a segment must be longer than it normally is, or longer than an adjacent segment.) But he adds, "...still the condition of one less sonorous context is far from being satisfactory on purely phonetic grounds." (p. 110) This would explain why the /l/ in bottling is syllabic (it is more sonorous than /t/ and long), and why the /l/ in wholly is not syllabic (it is long, but not more sonorous than its environment). Rubach admits though that his criteria do not explain why /s/ in stop is not syllabic even though it is more sonorous than /t/ and can be lengthened.

In an acoustic study of syllabic in Danish, Rischel (1967) compared pairs of words like /bon(ə)nə/ and /bonnə/.

1Ilse Lehiste has raised the question regarding Rubach's work that if the /l/ in bottling is syllabic, why isn't the /l/ in stripling? There is no phonetic reason for the difference (and for some New York dialects, bottling does not have a syllabic /l/). It is probable that the underlying representation of bottle is /batəl/ but strip is /stripl/. In the uninflected form, the /l/ in strip is syllabic because English syllables cannot end in -pl. But in stripling, the /l/ can be syllable-initial, so no adjustment is needed.
The former word need have no [ə], and when [ə] is absent, the word is most appropriately transcribed /bonnhə/. The acoustic characteristics associated with [bonnhə] are (1) /n/ is longer than /n/, (2) a steep intensity peak coincides with a peak in the fundamental frequency contour for /n/, and (3) the fundamental frequency peaks about two-thirds of the way into /bonnhə/, as opposed to almost the end of the word in /bonnə/. It appears that all three cues signal syllabic ity, although it is not known which cues are necessary and/or sufficient\(^1\). (Also see Sommerfelt (1931) on similar pairs of words in Norwegian.)

1.6. Summary

To summarize the literature on phonetic syllabic ity, the most widely held theories say that sonority (meaning loudness) and some durational factor are the primary acoustic cues to syllabic ity, relative degree of opening is the supraglottal articulatory correlate, and some sort of muscular activity of the respiratory organs is the subglottal articulatory correlate. That there are counterexamples to all of these criteria is readily

\(^1\)Rischel was working with natural speech on Danish, but a rather simple perception experiment using synthetic speech should answer the question of which acoustic cues are necessary and/or sufficient. The word pairs could be synthesized and the individual parameters -- length, intensity and fundamental frequency -- independently controlled. Subjects would indicate which member of the pair was closest to the key word.
admitted, and the physiological chest pulse theory in particular has been shown to be wrong in its strongest form. The existence of syllabic obstruents is especially problematic for a strictly phonetic theory. As in dealing with all speech phenomena, it is unlikely that a single group of parameters (acoustic or articulatory) will account for all the data.

2. Phonetic Syllable Boundaries

It has been maintained that while it is possible to define phonetic syllabic ity, it is not always possible to determine, on purely phonetic grounds, where syllable boundaries lie. Some, like Jespersen (1904), have generalized that it is always fruitless to locate boundaries; others, like Jones (1949), have said that it is sometimes possible. In an informal experiment using 128 Swedish words and 12 subjects, Gårding (1967) found high agreement on where syllable boundaries fall in Swedish. The literature falls generally into two categories: acoustic evidence and articulatory evidence.

2.1. Acoustic Evidence for Syllable Boundaries

In a pioneering study on the perception of syllable boundaries in VCV sequences, Malmberg (1955) focused on the interaction between the length of the stop gap and the presence or absence of a transition to or from the vowel on
either side of the gap\(^1\). Using synthetic speech he found that when there was a transition from the first vowel to the consonant but no transition to the second vowel, listeners said the syllabication\(^2\) was VC.V; when there was only a transition from the consonant to the second vowel, the syllabication was V.CV. (A period indicates a syllable boundary.) It seems that the presence of the transition put the consonant in the same syllable as the vowel. The stop gap ranged from 20-200 msec and its only effect was that when the gap was very short, the sequence was perceived as V.CV regardless of the transition. Interesting though Malmberg's experiment is, it doesn't represent natural speech production because in natural speech, there are always transitions to and from the consonant regardless of the perceived syllable boundary. His technique suggests a number of variables to take into account in working with a VCV sequence. Acoustic parameters like vowel amplitude and fundamental frequency (considered as stress), presence of aspiration, and vowel duration can all be cues to the location of the boundary. One of these cues, aspiration, in

\(^1\)While Malmberg is using the term transition in the accepted sense of a swiftly changing inflection in the frequency domain, it should be kept in mind that transitions between segments are never missing in real speech, although Malmberg, in his synthetic stimuli, has them either present or absent.

\(^2\)I am using the terms syllabication and syllabification in the following ways: syllabication refers to the division of a phonetic sequence into syllables; syllabification refers to the process of making a segment syllabic.
conjunction with transitions and length of the stop gap was investigated by Christie (1974). He used the nonsense word /asta/ and controlled for the presence or absence of a transition from /a/ to /s/, and a stop gap ranging from 15-135 msec. In contradiction to Malmberg's findings, the presence or absence of the transition did not affect perception as /a.sta/ or /as.ta/. As expected, aspiration was the strongest cue: when present, the word was syllabicated /as.ta/. (The amount of aspiration was not varied; it was an on-off cue.) Increasing the length of the stop gap only worked as a cue in conjunction with aspiration. Both a long gap plus no aspiration and a short gap plus aspiration produced random results. As Christie admits, finer distinctions might have been brought out if different amounts of aspiration had been used instead of just presence or absence.

Lehiste (1961) has described some of the acoustic characteristics of syllable boundaries in Estonian, a language which she argues requires the syllable as the basic unit for describing the distribution of quantity. She says that syllables are phonetically identifiable because the duration of syllables is linguistically significant. Syllable boundaries are either marked by phonetic allophones, or by distributional constraints on consonants, or by stress cues. In vowel-resonant-vowel sequences, the boundary in acoustically signaled by a sudden increase in
intensity within the resonant, apparently indicating that
the resonant is phonetically ambisyllabic\(^1\). In vowel-long
voiceless stop-vowel sequences, the boundary is known by
the release and rearticulation of the stop, also showing
that the stop is phonetically part of both syllables. In
vowel-overlong voiceless stop-vowel sequences, there is
both a decrease in subglottal pressure and an increase in
oral pressure in the stop, and again the syllabication is
VC:.CV. There are also distributional cues and stress cues
to syllable boundaries in Estonian; these will be mentioned
below. In a comparative study of consonant gemination in
English and Estonian (Lehiste, Morton, and Tatham 1973) the
authors found rearticulation in the production of Estonian
medial long and overlong consonants as well as junctural C+C
sequences. But no rearticulation was found in English
junctural C+C sequences.

2.2. Articulatory Evidence for Syllable Boundaries

\(^1\)In recent research on machine recognition of continuous
speech, Lea (1974, 1975) discusses an algorithm for locating
stressed syllables from prosodic features of energy and
fundamental frequency. The procedure is based on local
increases in fundamental frequency and large integrals of
energy within the syllable nucleus. That is, decreases in
fundamental frequency and intensity minima are the criteria
for the placement of acoustic syllable boundaries of
stressed syllables. In similar work related to phoneme
recognition, Todd (1976) argues that "By breaking the
acoustic chain into syllable-sized units prior to phoneme
analysis, it may be possible to use syllable structural
information in the phoneme identification process." (p. 1)
2.2.1. Implosive Versus Explosive Consonants

Grammont (1946) and Saussure (1966) both define phonetic syllable boundaries as occurring at points where articulatory closure is followed by release, or in Saussure's terms, points of implosion are followed by explosion. A VCV sequence is always syllabicated V.CV because the consonant is 'exploded' into a segment with a greater aperture. Where languages allow initial and/or final clusters, syllabication will vary depending on the distributional constraints of the language. This is a phonological consideration though, and while neither Grammont nor Saussure deny the importance of phonological criteria, they are attempting to make universal statements about phonetic syllables, 'unmarked' syllabication, as it were.

Grammont distinguishes eight degrees of aperture among sounds. From greatest to least, the order is low vowels, mid vowels, high vowels, semi-vowels, liquids, nasals, spirants, and stops\(^1\). Presumably, a boundary occurs between a sound of decreasing aperture and a sound of increasing aperture.

---

\(^1\) Grammont's degrees of aperture are only slightly more detailed than Stampe's sonority hierarchy for syllabication. From margin to nucleus the order is from least to most sonorous: obstruent, nasal, liquid, glide and vowel. Any reversal in the order requires placement of a syllable boundary. Hooper (1973) has also discussed degrees of sonority with respect to syllable structure conditions. Kuryłłowicz (1948) points out that the order of elements from margin to nucleus to margin is not always symmetrical.
aperture. In a language like English though, it is still an open question whether a word like master syllabicates before or after the s, particularly if the t is not heavily aspirated. According to Grammont, mast.er would not be a possibility because in intervocalic position, syllables must start with the most closed sound\(^1\). (See below on phonological syllable boundaries for a more thorough discussion of this question.) On the other hand, Saussure notes that a sequence like ar dra can be either ard.ra or ar.dra depending on the "speed of passage from implosion to explosion." (p. 57)

Fliflet (1963) refers to the well-known idea of a loose versus a close contact between the vowel and the consonant in a VCV sequence. A long first vowel results in the perceptual effect of loose contact and the boundary occurs before the consonant. A short vowel makes a close contact, and this results in perceptual ambiguity. In this case

\(^{---}\)

\(^1\)An interesting debate about the universality of CV syllables was started by Sommer (1970) who claimed that a group of Australian dialects has only VC(CCC) syllables. Unfortunately, he gives no phonetic or acoustic evidence that this is the case, but his phonological evidence consists of the fact that all words begin with vowels and end with consonants. He does add, however, that a small subset of words lose their initial vowel and add a word final schwa-like vowel. In a reply to Sommer, Darden (1971) regards that fact as evidence that at least phonetic CV syllables exist. In Olgolo, another Australian language, Dixon (1970) says that all initial consonants were lost, so that word-initial syllables all begin with vowels. But he adds that word medially, syllables are CV(CC). Also certain semantically determined noun classes show a C\(^n\)V free variation in initial position, the C being /w, y/ or /n/. Initial vowels are apparently unstable phenomena.
people tend to say the syllable boundary falls 'within' the consonant, or that the consonant is part of both syllables. Phonetically, close contact leads to gemination of the consonant, or at least a phonetically long consonant. Thus VCV is V.CV and VCCV is VC.CV. Fliflet cites evidence from English, Amharic, Dutch, French, Czech, and Italian which all exhibit a tendency for vowels to be long in open syllables and short in closed syllables. Actually this distinction seems to be circular with respect to placement of boundaries: do open syllables give rise to long vowels, or do long vowels give rise to open syllables? It would seem that both tendencies occur, and so we have to look at higher level phonological constraints in order to determine syllable boundaries. There are also languages like Finnish, Estonian, and Hungarian, which have phonemic contrasts between short and long vowels in open syllables. This question alludes to the problem encountered over and over, that of the necessity to make a distinction between the phonetic and the phonological syllable.

2.2.2 Extrinsic Allophones

The manifestation of phonetic syllable boundaries can sometimes be seen in the extrinsic allophones of syllable-initial versus syllable-final phonemes. Studies relating to juncture phenomena as in a name - an aim are numerous; many are cited in Lehiste (1960) and Gårding
(1967). But in such studies, the terminal allophones are always coincident with word boundaries. Although word boundaries will always be syllable boundaries, the opposite is not the case. For this reason, many researchers have been skeptical of evidence pointing to word medial (or morpheme medial) allophones as signalling syllable boundaries.

Hoard (1966) is interested primarily in the phonological syllable, but he claims that of four essential characteristics of English syllables, two are phonetic. In his view, the syllable (1) consists of one or more segmental phonemes, (2) has an obligatory nucleus and optional margins, (3) has a final phoneme which is 'short' in duration, and (4) has an initial phoneme which is 'long' in duration. (3) and (4) are phonetic criteria. Although he does not elaborate on what he means by 'long' and 'short' he apparently means that there is some inherently medium duration for a given segment type which is lengthened or shortened depending on syllable position. For Hoard, syllabication rules follow from stress placement. He says that because the first vowel in water and the second vowel in potato receive primary stress, the syllabication that results is /wɔt.ər/ and /pə.tət.o/. As expected, the syllable-initial /t/ in potato is long (aspirated would be more likely), and the final /t/ in both words is short
(often called flapped). Hoard explains these different durations as being due to syllable position, but what he ignores is the question of how the consonants got into the syllables they are in. One answer is that there is an articulatory tendency for stressed vowels to 'attract' consonants into a 'close contact,' and thus to create syllables with consonantal margins. This suggests that extrinsic allophones are the result of syllable boundary placement, not the cause of it. A word like plato has two pronunciations, [pléyr.o] and [pléy.tĥ ho]. While the 'stress attracts consonants' rule would predict the first variant, the other criterion for syllabication, that intervocalic consonants syllabicate to the

---

1It is often assumed that the presence of a flap is evidence of syllable-final position, although Lehisite (1961) calls the flap a non-initial, non-final allophone of /t/ and says there are no acoustic cues to indicate which syllable it is in.

---

2The 'stress attracts consonants' principle was pointed out to me by David Stampe. To my knowledge, the question of why this happens has not been answered in the literature. (But cf. Kuryłowicz 1948, for a discussion of the facts.) My conjecture is the following: In Germanic languages an initial stressed vowel is often preceded by a glottal stop. This suggests that whatever physiological mechanism is responsible for stress, its 'unmarked' articulatory correlate is a brief closing of the vocal folds. When there is a supraglottal articulation, that is what coincides with the respiratory mechanism. The segmental allophone -- aspiration, duration, etc. -- is the result of the muscular effort involved. (See Lehisite and Peterson 1959 on the effect of physiological effort on the perception of stress.) (continued on next page)
right, predicts the second. It seems that in some lexical items speakers of English choose one principle over the other. Note also that even in words that don't ordinarily have alternate pronunciations, like water, one can always override the 'stress attracts consonants' rule and in overly articulate, careful speech say [wɔθər]. It is unlikely that one 'decides' on a particular allophone, here an aspirated [tʰ]; rather one uses syllabication rules to divide up the speech continuum for some particular language use, and the segmental allophones follow from that syllabication.

In an experimental study of English syllabication, Davidsen-Nielsen (1974) defines the phonetic syllable as "...a language-specific but non-abstract unit, whose limits are signalled by 'extrinsic allophones.'" (p. 15) He uses aspiration of voiceless stops as the criterion for a syllable-initial allophone, and his data consist of words

(continued from previous page)

This suggests a reason for consonants to be attracted into initial position in stressed syllables, but it does not explain why consonants are attracted into stressed syllable final position. Why do water and potato have final /t/ in the stressed syllable? One possibility is that the consonant is there by default, so to speak. It is not that the stress attracts final consonants, but that the following unstressed syllable does not attract an initial consonant. The decrease in intensity and/or lowering of the fundamental frequency which occurs at the end of a stressed syllable renders the final consonant short, or unaspirated, sometimes unreleased if it is a stop, or generally weak. (See Hooper 1973 for a nice discussion of syllable-final weakening in Spanish.) Thus the /t/ in water is reduced to a short voiced flap because the following unstressed syllable doesn't have the strength to attract it.
containing medial clusters of /s/ followed by /p,t,k/, optionally followed by /r,l,w,y/, followed by a stressed vowel. Subjects were told to pronounce each of about 100 words, and Davidsen-Nielsen used spectrograms and mingograms to measure the presence and amount of aspiration. He found unaspirated stops were produced in the vast majority of cases, indicating a V.sC syllabication. The only exceptions occurred where a transparent morpheme boundary separated the /s/ and the stop, in words like miscalculate, mistime, discourteous, etc., and it was not always the case that a morpheme boundary had this effect. In some cases, Davidsen-Nielsen is unable to explain certain 'irregular' syllabifications, as in transparent, [trans.pʰːrənt] and extrinsic, [ɛ.k.strɪnˈsɪk], since he cites studies on English syllabication which argue that CCC should syllabicate C.CC and CCCC should be CC.CC. But the 'stress attracts consonants' rule, which he does not discuss, would explain why /kstrV/ syllabicates /k.strV/, and the degrees of aperture of Grammont and others, explains why */ks.tr/ is impossible, since /t/ has less aperture than /s/. In transparent, the main stress on the second syllable predicts an /n.sp/ syllabication, although it gets /ns.pʰ/. But trans-, being a morpheme itself, also is not totally unstressed; the vowel never reduces to [ə]. Both of these facts make /ns.pʰ/ the likely syllabication. Thus, we see that higher level considerations like morpheme boundaries
influence syllable boundaries.

Other authors have used information about extrinsic allophones to explain various syllabifications. Although phonetic allophones may follow from syllabication rules, it is also possible to infer where the boundaries were placed once we know how phonetic allophones are constrained. Higginbottom (1964) discusses a dialect of British English that has 'glottal reinforcement'—glottalization that either precedes or takes the place of a syllable-final voiceless stop. Thus, likely is [lai?li] with glottalization, but can also be [lai'kli] with preglottalization and a voiceless /l/. The problem is to determine the syllable boundary in the latter version. According to Higginbottom, it cannot be [lai'kli] because an initial /l/ wouldn't devoice, and it wouldn't be [lai.'kli] because glottalization can't be syllable-initial in this dialect. She concludes that the /k/ is ambisyllabic, and the correct syllabication is [lai'k.kli]. This is a nice example of a purely phonetic argument since there are no phonological reasons to suggest one or the other syllabication.

Another example comes from Bailey (1968). English 'dark' /l/ occurs in syllable-final position (and initially before a back vowel); otherwise the 'light' /l/ occurs. (See Lehiste 1961 and 1964 for an acoustic study of allophones of /l/.) Furthermore Bailey refers to VC.V as
marked syllabication and V.CV as unmarked. The Southern American English pronunciation of *silly* uses unmarked syllabication, [sɪ.li], while the Northern version is marked, [sɪɻ.i]. But Bailey doesn't consider the fact that if stress is taken into account, then the notion of marked-unmarked syllabication is not as clear. With stress on the first syllable, the Northern pronunciation becomes unmarked because the primary stress attracts the /l/ into the first syllable, and as a result, it is velarized. The Southern version is then seen as marked because it overrides the stress principle. Again, different principles produce different results, although both are motivated by purely phonetic criteria.

The only author found to give due credit to the influence of stress is the phonetician Henry Sweet (1877). He says that syllable division can only be known phonetically by the onset of stress. "...syllabification depends not on mere force, but on discontinuity of force." (p. 90) In other words, the juxtaposition of stressed and unstressed vowels determines boundaries. In a word like *ticket*, the /k/ is in the first syllable. In order to get it into the second syllable, one must diminish the stress on /tɪ-/; "...so as to begin a fresh impulse of force on the /k/." (p. 90) This sounds good as far as it goes, but doesn't explain where to place boundaries between adjacent unstressed syllables, as in *classificatory* or *labialization*. 
Presumably we can appeal to the 'consonants to the right' rule in the former word, /klæ.sif.ə.kə.tɔr.i/, and consider the velarization of /l/ as evidence of syllable-final position in the latter word, /ley.bi.əŋ.ə.zəy.ʃən/.

Finally, Hockett (1955) uses the term interlude to refer to a medial consonant cluster which cannot unambiguously be broken up into offset and onset. "An interlude is coda-like and onset-like at the same time, and structurally it belongs both to the syllable which contains the preceding peak and to that which contains the following peak." (p. 52) Furthermore, "When two successive syllables in a language like English are linked by an interlude, there is no 'point of syllable division' between them." (p. 52) He uses the example of /tr/ in nitrate as an interlude, claiming that it cannot be broken into an offset and an onset. Hockett does not mean than an interlude is both a possible onset and a possible offset. Rather it is structurally a single unit which can only occur in medial position. In nitrate, the /r/ is voiceless so it can't be syllable-initial; and the /ay/ of the first syllable is short, a characteristic of vowels in closed syllables. Therefore the /t/ must belong to both syllables and since the voiceless allophone of /r/ is dependent on /t/, the whole cluster takes on its particular structural cohesiveness.
3. Phonological Syllabicity

Most of the arguments for the necessity, or at least the desirability, of incorporating the syllable into phonological theory are based on the fact that phonological syllable boundaries are crucial in the description of a language's phonological processes. But the distinction between phonological syllabicity and syllable boundaries is not always recognized. It is possible to consider the importance of the syllable in stating the domain of tone or stress, without becoming embroiled in the controversy over where the syllable boundaries lie. In this section I review some of the arguments for phonological syllabicity.

Pike (1947) carefully distinguishes between phonetic and phonemic syllables, and defines the latter as "...units of one or more segments in length such that one phonemic syllable constitutes for that language a unit of actual or potential stress placement, or tone placement, or intonation placement, or rhythmic grouping, or of morpheme structure." (p. 60) In discussing the phonological features of tone, Wang (1967) says, "Phonetically, of course, the domain of tone is over the entire voiced portion of the syllable." (p. 95) He says that because of this fact, tone features should be formalized differently from segmental features; namely that they should be regarded as features of individual syllables, although only in tone languages with polysyllabic morphemes. (See Gandour 1974 for arguments
that tones should be represented segmentally in Siamese.) Likewise Stetson (1951) says of the syllable, "...it is the unit for the word- and the sentence stress; it is the unit for the 'tones' of tone languages;" (p. 1) Like others, Fudge (1966) finds it important to distinguish between the phonetic and phonemic syllable. On the phonemic level, its function is to provide a basis for distinctive prosodic features and to account for constraints on possible phoneme sequences. Trubetzkoy (1939) also says, "Prosodic properties do not belong to the vowels as such but to the syllables." (p. 170)

On the other hand, there is at least one language said to be without syllables. Newman (1947) writes, "There are no syllables in Bella Coola, and no phonemically significant phenomena of stress or pitch associated with syllables or words." (p. 132) Recall that Bella Coola words often consist only of strings of voiceless obstruents, which produce phonetic syllables solely in their fortis releases. (See Greenberg 1962 for an argument that even Bella Coola has phonological syllables.)

The prosodic feature that has not been mentioned yet is duration. Several researchers have claimed that the domain of quantity cannot be described without reference to the syllable. For example, Hall (1971) says that one can only account for long versus short consonants in Italian by reference to the syllable: short consonants are
syllable-initial; long consonants are ambisyllabic. Lehiste (1970) claims that the domain of quantity in Icelandic, Norwegian, and Swedish is the syllable. A long consonant must follow a short vowel, and a short consonant must follow a long vowel. Estonian is even more dependent on the syllable -- the duration of the vowel in the second syllable of a word varies with the quantity of the first syllable. She says, "I see no way of accurately predicting the phonetic realization of segmental sounds in Estonian without reference to higher-level phonological units -- syllables and sequences of syllables." (p. 159)

4. Phonological Syllable Boundaries

4.1. Distributional Constraints

By far, most of the literature on the phonological syllable has to do with distributional constraints of vowels and consonants. Any grammar of a language has to include statements of segment distribution and concatenation, and it has been argued by many researchers that these facts are best stated in terms of syllable-sized units.

The most extreme position as to the relationship between the phonetic and phonological syllable is taken by Panconcelli-Calzia (1924), who, according to Malmberg (1954), considers the syllable solely as a psychological or phonemic unit without any acoustic or articulatory counterpart. But Panconcelli-Calzia is actually talking
about syllable boundaries, not syllabicity as such. Also Haugen (1956) argues that the syllable is purely a phonological unit. He claims that it is not identical with the number of vowels because consonants can be syllabic as well; it is not identical to the number of accents in the word because any tone or stress can be given a diacritic and placed on or after the syllable nucleus, and thus one need not refer to the whole syllable (but cf. Wang 1967 who argues that the domain of tone is over the entire voiced portion of the syllable); it is not needed for describing quantity because length can always be treated as a separate phoneme added to the preceding one (but cf. Lehiste 1970 who argues that the syllable is necessary to define duration in some languages); it is not identical with juncture because there are more syllables than junctures in a word. Instead, it is "the smallest unit of recurrent phonemic sequences." (p. 220) The syllable is the best framework for stating the distribution of phonemes.

Pulgram (1967) considers the syllable to be neither a phonetic nor strictly phonemic unit in the traditional sense of 'phonemic' meaning 'distinctive'. "...the syllable is not an emic unit in the sense that it serves a distinctive function." (p. 749) He also says that the only relationship between the phonemic and phonetic syllable is that "...the

1Haugen attempts to solve the problem of undivided interludes by saying that syllable codas should be divided into finals and medials, and syllable onsets into initials and medials. That is, one set of allophones is word-initial and syllable-initial; another set is syllable-initial but not word-initial.
establishment of syllable boundaries on the phonological level by phonotactic means delivers the points where phonetic syllable boundary signals MAY occur; but they NEED NOT occur." (p. 748) In fact, says Pulgram, the only reason a person might use phonetic boundaries would be his desire to divide an utterance into syllables. There is no real communicative function. "Unlike other linguistic units, the syllable is its own and only purpose...this basic meaninglessness... makes its phonetic realization optional..." (p. 749) Elsewhere Pulgram (1970) defines the syllable according to universal syllabication rules based on permissible word initial and word-final clusters. That is, if a cluster can be word-initial, it can also be syllable-initial, and if word-final, then syllable-final. But when medial clusters are ambiguous as to syllable division, two other principles are stated: (1) maximally open syllables, so that Easter is /i.stər/, and (2) minimal codas, maximal onsets, as in the often-cited Spanish problem of how to syllabicate the word transcription. /ns.kr/ is an impossible coda and /n.skɾ/ an impossible onset. Pulgram argues for the minimal coda, /n.skɾ/\(^1\). Essentially the same point concerning initial and final clusters is made by Fudge (1966), Hjelmslev (1936), Jakobson and Halle (1956), and Greenberg (1965), who provides a series of supposed

\(^{1}\)The syllabication trans.scripción is also bolstered by there being a morpheme boundary at the same place as the preferred syllable boundary.
universal implications concerning initial and final consonant sequences.

In a very comprehensive study, Kuryłowicz (1948) outlines his principles of syllabication. They are basically similar to Pulgram's. One difference is that in medial consonants that can be final and initial, the boundary is always within the consonant, unlike Pulgram's maximal onset, minimal coda principle. In fact, regarding medial geminates, Kuryłowicz defines a geminate as a consonant that is in two syllables, rather than delimiting the boundary as lying between the two consonants. That is, instead of saying one places the division between the consonants, he says the phonetic nature of the consonant, being implosive and explosive, determines that there is a phonetic geminate. As a result, there can be no word-initial or word-final geminates. (While an interesting approach, it seems that Kurylowicz is confusing phonetic and phonological boundaries. If a language like Italian has phonemic geminates, then the phonetic facts of syllabication follow from the distinctive function of the doubled consonant. It seems somewhat backwards to say that the syllabication determines whether there is a geminate consonant present.)

For English, the most complete attempt to write that section of the phonology dealing with the distributional distinction between consonants and vowels has been done by
O'Connor and Trim (1953). Their technique is to compare contexts common to pairs of phonemes. For example, in spar and saw, /p/ and /s/ share the context of occurring after initial /s/, and in pray and oral, the same two phonemes share the context, before post-initial /r/. The criterion for assigning phonemes to the same distribution class is whether their shared positions of occurrence are greater than 50% of the positions of occurrence of either member. For example, the total number of positions of occurrence for /η/ is 14, and it shares 10 of them with /p/. Since 10 is more than 50% of 14, /p/ and /η/ belong to the same distribution class. (To the authors, it doesn't matter that 10 is probably far less than the total number of contexts for /p/.)

Of the two classes that emerge, one occurs predominantly initially and finally, the consonants, and the other, post-initially and pre-finally, the vowels. On assigning syllable divisions to ambiguous clusters, the authors say, "The preference for one syllabic division as opposed to another may be explained in terms of the frequency of occurrence of different types of syllable finals and initials." (p. 121) So in VCCV, the relative probability of the three possible divisions, V.CCV, VC.CV, and VCC.V, is the criterion, and for English VC.CV ranks the

---

1Also see Scholes (1966) who discusses phonotactic phenomena of English consonants and the sonority principle.

2According to O'Connor and Trim's method, since /η/ and /h/ share no contexts at all, they must be the same phoneme. This conclusion can only result from using non-phonetic criteria in defining phonemes.
highest.

The same method is used by Arnold (1956) on French although in later research on Polish and Greek (1964) he rejects this approach. Likewise Greenberg (1962) tries to define the distribution of vowels and consonants using non-phonetic criteria. His claim is that the dichotomy of vowel and consonant is universal. The criterion for the dichotomy is the 'maximum recurrence interval,' which is based on counting the number of phonemes that can recur between two members of the same class of phonemes, e.g. the maximum number of consonants that can recur between two vowels. His non-phonetic criteria for the syllable are the following: (1) Every phoneme belongs to some syllable. (If this means a phoneme cannot belong to more than one syllable, then the interlude concept only applies to phonetic syllables.) (2) All phonemes belonging to the same syllable form a continuous sequence. (3) A member of the class of vowels occurs in every syllable. (This can only be the case if, in languages like Bella Coola, sonorants and fricatives can function as members of the vowel class.) (4) There is a finite upper limit to the length of each syllable.

Another example of the distributional approach applied to English is that of Trager and Bloch (1941). In an attempt to limit the number of vowel phonemes to six, /iɛɹʌʊ/, they claim that the so-called long vowels,
/ieuo/, are combined of short vowel plus a glide that agrees in backness. The distributional constraint they presuppose is that short vowels cannot occur in open syllables. In words like *being* and *going*, the phonemic representations are */biyŋ/ and */gʌwŋ/ and the glide is ambisyllabic. Analogously, words like *bedding*, *butter*, and *pudding* have ambisyllabic medial consonants. Thus the constraint on the short vowel distribution allows an economical reduction in the phoneme inventory, because short and long vowels are in complementary distribution: short vowel before a consonant; long vowel before a glide. Their analysis has been criticized by Eliaison (1942) who argued that it simply has never been proved that such medial consonants are in one or the other syllable, and therefore their argument cannot be used to limit the vowel inventory of English. Clearly, Eliaison has phonetic/acoustic 'proof' in mind, and as I have tried to show, even today there are no clear cut articulatory or acoustic cues for phonetic syllable boundaries.

On the other side of the phonological argument is the widely known view of Kohler (1967). He claims that the syllable is not a phonological universal because it is an unnecessary, impossible, and harmful concept. It is unnecessary because syllable boundaries can be deduced from the distributional facts about consonants. It is impossible because sometimes boundaries are arbitrary and thus not
determinable. It is harmful because in cases where boundaries are indeterminate, the underlying phonological structure is obscured. Kohler's arguments have all been answered by Anderson (1969). On the syllable being unnecessary, Anderson says just because segment distribution is predictable without invoking the syllable as a unit does not make it unnecessary. It is better to ask whether phonological rules are best stated with respect to syllable boundaries, and if not, only then would the syllable be unnecessary. On the arbitrariness of syllable division, he proposes underlying ambisyllabic consonants. The underlying structure of **butter** is CVC.CVC, and a later rule deletes the second /t/. (Anderson is apparently not bothered by short stressed vowels in open syllables.) Finally the syllable is not harmful because principles of underlying syllabication (as in **butter**) can be used to illuminate the status of 'ambivalent' medial sequences. Overall, Anderson shows the syllable to be a useful concept in phonology, rather than a necessary one, as he does not address the issue of phonological rules and syllable boundaries. Others though have dealt directly with this question.

4.2 PHONOLOGICAL RULES AND SYLLABLE BOUNDARIES

In recent years evidence has been accumulating that the syllable is a unit which must be part of phonological theory. Simply, the argument is that at least some
phonological rules can only make significant generalizations about a language if they can refer to syllable boundaries. In an excellent article Hooper (1972) says, "...current phonological theory provides no definition of the syllable, and thus no formal means of referring to it as a unit." (p. 525) Using Spanish as an example, she claims that generalizations about nasal assimilation, voicing assimilation, glide obstruentization, and vowel tensing are missed without reference to the syllable in the formal statement of the rule. Specifically, Hooper suggests that there are syllabication rules (some universal, some language specific) that insert syllable boundaries into strings of segments. For example,

(1) $\emptyset \rightarrow \$ / [+syllabic] ____ [+syllabic]$

(2) $\emptyset \rightarrow \$ / [+syllabic] ____ [-syllabic] [+syllabic]$

(3) $\emptyset \rightarrow \$ / [+syll] [-syll] o ____ [-son] [-nas] [+son] [+syll]$

(In this notation, $\$ indicates a syllable boundary.) In (1), $\emptyset VV$ becomes $VV$; in (2) $VCCV$ becomes $VCV$; in (3) clusters are divided such as $Vp.tv$, $Vr.tv$; $Vdrv$, $Vntv$, $Vtv$ etc. Rules (1) and (2) are universal; (3) indicates some of the options a language may choose. Vennemann (1972) also claims that while syllabication rules are universal, their application depends on language-specific consonantal strength hierarchies, plus information about speech rate,
stress, and sandhi phenomenal.

The syllable is conspicuously absent from the standard work on generative phonology, *The Sound Pattern of English* (Chomsky and Halle, 1968). The authors devote considerable space to predicting stress in English yet there is no formal consideration of the obvious relationship between syllables and stress. In his review of SPE, McCawley (1974) says, "There are considerations which lend support to the claim that English stress rules should be stated directly in terms of syllables and moras rather than in terms of segments." (p. 61) Hoard (1971) also argues that not only stress, but issues dealing with aspiration and consonant tensing can be given unified explanation in terms of syllabication rules. Likewise Stampe (1973), in discussing the Middle English rule of tri-syllabic laxing, says that the rule, which accounts for the /ay/Δ/1/ alternation in divine - divinity, must mention a syllable boundary. In research on French, Fudge (1966) says that syllable boundaries are necessary in order to predict masculine and feminine forms of adjectives. Vennemann (1972) also cites open syllables as the conditioning factor for vowel lengthening in Icelandic. Sommerfelt (1931) discussed syllable boundaries and the

---

1Vennemann's strength hierarchies are similar to Grammont's degrees of aperture and Stampe's sonority hierarchy. His theory goes further than the others in that strength is dependent on position in the syllable. Initial position is strongest; final is weakest. See Hooper (1973) for a thorough discussion of Spanish and its strength hierarchies.
historical development of Irish. Finally, Lehiste (1970) refers to the fact that in Finnish and Lappish, the conditioning factor for strong versus weak degrees of stem consonants is the placement of the syllable boundary of the following syllable. "If there is a single consonant between the second and third syllables, the stem consonants appear in the strong degree; a cluster of two consonants or a long consonant between the second and third syllables causes the appearance of the weak degree." (p. 45)

Recently C. Jones (1976) has come up with a very interesting proposal concerning syllable boundary placement in English. Essentially, he says that any intervocalic consonant is ambisyllabic, excluding cases of impermissible onsets and offsets. For example, the syllabication of piper is \([\text{pi}^2\text{p}^\text{er}]\), algebra is \([\text{al}^1\text{g}^2\text{e}^1\text{b}^1\text{ra}]\), although ahead would be \([\text{a}^1\text{head}^2]\) because /h/ cannot be syllable-final. In other words, "...if a certain consonant in a medial cluster could be interpreted as belonging either to the coda of the preceding syllable or to the onset of the following, it is to be interpreted as belonging to both." (p. 122) He further suggests that certain phenomena in the history of English point to "...a directed rule sequence, conspiring to achieve overlap in medial clusters." (p. 126) A brief example will illustrate Jones' point. Several dialects of Old English underwent a medial consonant deletion,
producing alternations like the following:

(5) yntse $\sim$ ynse 'ounce'
    blotsma $\sim$ blosma 'blossom'.

Jones claims that the motivation for this deletion (and similar ones involving /n/ and /d/) was the change from a proper syllable bracketing to the 'unmarked' non-proper bracketing, as in (6):

(6) [ynt] [se] $\rightarrow$ [yn[sje] 'ounce'

    [blots] [ma] $\rightarrow$ [blo[s]ma] 'blossom'.

Thus overlap in medial clusters is achieved. The novel aspect of Jones' thesis is that medial overlap is the unmarked or natural state of affairs in English syllable structure. Although Jones does not generalize to other languages, it may well be interesting to investigate phonological rules in other languages in terms of this

\footnote{It is unclear why /blotsma/ 'blossom' should have started out with a proper bracketing: [blots][ma]. It seems that the /s/ should be both syllable-final and syllable-initial, to give the bracketing [blot[s]ma], since old English did not have a constraint against initial sm-clusters.}
principle and their historical development.

5. Rhythmic Syllabicity

So far, the syllable has been viewed simply as a combination of vowels and consonants. No mention has been made of the syllable as a higher level linguistic unit, a unit that is neurologically programmed and executed. But possibly the most significant aspect of a study of syllables is the investigation into their role as units of performance. A number of researchers have addressed themselves to this issue; some of their theses will be mentioned now.

Rosetti (1959) and deGroot (1926) both speak of the syllable as the smallest unit of 'rhythmic grouping' in a language. The rhythmic group is not actually defined, but deGroot cites three characteristics of it: (1) the tendency toward articulatory grouping of consonant plus vowel; (2) the tendency toward differentiation, to separate the peaks of sonority; and (3) the tendency toward a norm, or standard syllable pattern, such as a preference for open syllables.

The rhythmic group is probably best defined in terms of timing relations. Haugen (1949) sees stress, quantity, juncture, and (to a lesser extent) pitch all as timing phenomena and all dependent on the syllable as a linguistic unit. His definition of the syllable is "...that recurrent sequence of sounds, in terms of which the phenomena of
linguistic timing can be described." (p. 28) In fact the syllable "...is nothing less than the METRONOME of human speech." (p. 281) In the same vein, Fry (1964) says, "The basic idea of the syllable is connected with rhythm, whether of everyday speech or of poetry, and it is therefore fundamentally concerned with the time scheme of speech." (p. 217) Actually, Fry argues that the syllable is necessary almost by default — the timing relations among all the speech organs (lungs, larynx, pharynx, oral articulators, nasal coupling) are so complex that they can only be the result of central organization in the brain. Citing delayed auditory feedback experiments as circumstantial evidence that the syllable is the basic unit of motor control, he says that speech is most disrupted when the feedback delay is 0.1 second. This time interval is slightly less than the mean syllable length in many languages. Specifically, Fry says that most speakers utter between five and seven syllables per second. This means that a 0.1 second delay in auditory feedback comes just when organization of the next syllable is going on, and thus, is the most disruptive. Fry also says that the link between motor organization and

---

Fry also points out that delayed auditory feedback seems to affect stress-timed languages like English and German more than syllable-timed languages like French. This is what one would expect since syllables are of different durations in stress-timed languages. But i.: syllable-timed languages where syllables are all of approximately equal duration, a 0.1 second delay can probably be resolved by a slight overall adjustment of all the syllables in an utterance.
syllabification accounts for the role in perception played by one's native speech habits. That is, one hears the number of syllables in a foreign word according to how such a word would be syllabicated in his own language. (To the extent that this perceptual tendency is valid, it constitutes some evidence for a motor theory of speech perception.)

The research of Kozhevnikov and Chistovich (1965) also points to the syllable as the level of rhythmic organization in speech. Investigating articulatory changes in fast speech, they say that changes in the rate of speech do not affect the relative duration of syllable and word intervals in a phonological phrase (which they call a syntagma). Only the consonants and vowels within the syllable change their relative durations. Specifically, when speech is speeded up, it is at the expense of the vowels. This is evidence that the entire syllable is one programmed rhythmic unit.

Another type of circumstantial evidence for the syllable as a unit of programming comes from slips of the tongue. Fromkin (1971) claims that segmental slips obey a structural law with regard to syllable place: initial consonants slip with initial consonants, e.g. turkey noodle becomes nurkey toodle; and final consonants with final

\[\text{-------------------------}\]

1Kozhevakonov and Chistovich explain the shortening of vowels as opposed to consonants in fast speech as due to the fact that there is a limit on how short a consonant (particularly a noncontinuant) can be, and still be a speech sound, whereas a vowel can become shorter and shorter, or disappear altogether, or be indefinitely stretched out.
consonants, e.g., top shelf becomes /təf/ shelp. Furthermore, some slips involve deletion of entire syllables, e.g., Morton and Broadbent point becomes Morton and Broadpoint. (Evidence from slips was pointed out by Fry also.) Fromkin concludes that "...the linguistic unit corresponding to individual sets of motor commands is not the size of a phoneme or phonetic segment but rather that of a syllable or some combination of phonemic segments." (p. 63)

Finally, the work by many researchers on categorical perception of speech, summarized in Liberman, et al. (1967) and Liberman (1972), has shown that although the phoneme itself has psychological reality, speech is not processed as a series of atomistic acoustic events. Rather, there is parallel processing of speech signals. As Lehiste (1972) puts it, "...the same physical signal (e.g., a frequency change in the second formant) carries more than one kind of information (e.g., the phonetic value of a vowel and the point of articulation of an adjacent consonant)." (p. 193)

The fact of parallel processing points to a unit higher than the phoneme as the minimal perception unit, namely that of a syllable. According to Liberman, et al. (1967), "This parallel delivery of information produces at the acoustic level the merging of influences we have already referred to and yields irreducible acoustic segments of approximately syllabic dimensions." (p. 441) Furthermore, in a recent
article, Studdert-Kennedy (1975), in reviewing literature on speech perception, focuses on the differences between vowels and consonants. The acoustic factors that distinguish stops from vowels are energy and spectral stability. He claims that there are communicative functions for these differences. For example, although vowels are almost always longer than consonants, one can identify a vowel as short as 20 msec. The remaining length is segmentally redundant, but allows suprasegmental information to be displayed, such as fundamental frequency, duration and intensity. Tone languages and quantity languages use these effects linguistically and all languages use some type of accent and intonation. The 'disadvantage' of the typical length of a vowel is that segmental information is transferred more slowly and the vowel is subject to a variety of contextual effects. But adding what Studdert-Kennedy calls 'consonantal attack' to the vowel acts as a source of acoustic contrast between vowels, reduces vowel context effects and increases the phonetic range. Since consonants do not usually carry prosodic information,

"The segmental and suprasegmental loads are therefore divided over consonant and vowel - the first, with its poor auditory store, taking the bulk of the segmental load and the second taking the suprasegmental load. There emerges the syllable, a symbiosis of consonant and vowel, a structure shaped by the articulatory and auditory capacities of its user, fitted to, defining and making possible linguistic and paralinguistic communication." (pp. 119-120)

The preceding survey of literature on syllables and syllabic ity, while surely not complete, points to the scope and depth with which the syllable has been investigated. Perhaps more problems have been raised than questions answered. Certainly the physiological mechanisms corresponding to the syllable have not been uniquely determined. Specifically, the chest pulse theory has been shown not to be correct in detail, although it may be in principle. Most people believe that sonority is the primary acoustic correlate, although the problem of syllabic consonants forces us either to consider sonority not to be the sole property of vowels, or to give up sonority per se as the basic acoustic correlate. As I hope to show in subsequent chapters, mere length of a consonantal segment is perceived as syllabic ity as well. The problem of determining phonetic syllable boundaries is also without consensus. It does seem that in languages with complex medial consonant clusters, like English, there is no unambiguous cue to syllable divisions. With respect to phonological syllabic ity, it appears that prosodic features are best stated on syllable-sized units, although for some languages, there are arguments to the contrary. There is also accruing abundant evidence that phonological theory needs the syllable in order to state generalizations concerning phonological processes. Finally, it has been
suggested that the syllable is the basic unit of neurological organization, the major source of evidence coming from delayed auditory feedback experiments and slips of the tongue. The evidence though is largely inferential and this area is one which awaits further investigation.
CHAPTER II: SYLLABICITY IN ENGLISH: PRODUCTION

0. Introduction

Although the main perceptual experiments to be described in Chapter III involve synthetic speech, it was considered important to discover how English speakers produced the seven word pairs chosen for study. In this chapter I describe the method I used to obtain natural speech recordings, and the criteria for deciding which of the large number of pairs to ultimately use. I also present an analysis of the key words, looking at the durations of the words in various speech styles, and some of their acoustic characteristics. Finally I describe an experiment (Semiloff, 1975) involving the perception of speech styles.

1. Experimental Design

One of the earliest questions about the perception of syllabicity that gave rise to this study was whether a word could be differentially perceived when the listener was told to imagine that he was listening in a particular speech style. Since the variable speech style was to figure prominently in the perception experiments, a technique was devised to elicit the production of the key words in various speech styles. The four styles chosen were Formal Slow
(FS), Formal Fast (FF), Casual Slow (CS), and Casual Fast (CF). The definitions of the four styles are somewhat impressionistic, but as the acoustic analysis given below will illustrate, differences among them can be seen. Formal Slow style is considered carefully and clearly articulated speech, such as that used to an important audience; Formal Fast style is identical to Formal Slow except that it has a quicker tempo. Casual Slow style is the relaxed, often sloppily articulated speech used among friends or family. Casual Fast is identical to Casual Slow except that it has a quicker tempo.

The subjects (described below) were given the following task in order to elicit the key words in all four speech styles. For the Formal Slow and Formal Fast styles, a paragraph task was used. I told each subject to read two unrelated paragraphs, constructed to include all the key words in a coherent story. (Two paragraphs were written because it wasn't possible to get all the key words into one story.) The subjects were told to pretend that they were reading to an important audience and that it was crucial that the audience understand everything being read. For the Formal Slow style I emphasized that there was no time limit on the presentation and told the speakers to read as slowly as possible (within the limits of sounding natural). For the Formal Fast style, I stipulated that there was an imaginary time limit on the presentation, and that the
speakers were to read the paragraphs as formally, yet as quickly, as possible.

In order to elicit the Casual styles, I decided that spontaneous speech, as opposed to reading, would work best. A technique suggested by Garnica (1974) obtained the best results. I presented the subjects with magazine pictures, cut out and pasted on cardboard. A sentence, describing some aspect of each picture, was typed beneath it. The task was to make up a short story about the picture and to use the typed sentence at some point in the story. The purpose was to insure a uniform reading of the sentence with the key word in it, and also to get the subject into a casual, extemporaneous speech mode. To set the scene, I told each subject to pretend that his wife was home sick and that he was telling her stories to pass the time. For the Casual Slow style, subjects were again told there was no time limit on their storytelling; for the Casual Fast style, they were told to rush through the stories because of a class to attend, or some such pretense.

Three subjects participated in the experiment. All were native speakers of General Midwestern English; G.E. was from Cleveland, R.K. was from St. Louis and R.E. was from Des Moines. The recordings were made in an echo-free chamber at the Ohio State University, using an Ampex G-360 tape recorder for G.E. and R.K. and in a sound treated room, using an Ampex G-360 tape recorder at Indiana
University for R.E. The subjects did each of the four tasks four times. By the third recording, each speaker stated that he felt comfortable with the tasks (reading and storytelling), and subsequent analyses were done almost exclusively on the third and fourth recordings.

2. Choice of Key Words

The monosyllabic-disyllabic key words fell into three categories according to the phonetic nature of the disyllabic word. (1) Some words had initial stops, followed by unstressed vowels and a liquid, followed by the stressed vowel, as in BLOW-BELOW, PLIGHT-POLITE, PRAYED-PARADE, DRESS-DURESS, CREST-CARESSED, BROKE-BAROQUE, TRAIN-TERRAIN. (2) Some words had initial unstressed vowels, followed by a resonant consonant and the stressed vowel, as in RACE-ERASE, MUSE-AMUSE, WAKE-AWAKE. (3) Some words had an initial /s/ followed by an unstressed vowel followed by either a stop or another /s/, and then the stressed vowel, as in SPORT-SUPPORT, SCUM-SUCCEDE, SEED-SECEDE, and STAIN-SUSTAIN.

The criterion for inclusion in the perception study was that for any pair of words, the disyllabic member had to have been produced in at least one speech style as a monosyllable. For example, in at least one recording, a speaker had to produce BELOW so that it was impressionistically homophonous with BLOW, POLITE homophonous with PLIGHT, etc. The reason for this
stipulation was to use the production data as a basis for the perception test. That is, it seemed reasonable to require that, in some speech styles, words with syllabic resonants could be produced with a resonant so short that the word was homophonous with its non-syllabic counterpart. In the case of SUPPORT and SUCCUMB, the first vowel would have to be deleted entirely. For SUSTAIN and SECEDE, the first vowel would have to be deleted between the two s's. Although the claim was not made that people can only perceive differences that they produce (which is surely not true) it seemed unlikely that if, for example, BELOW was never produced so that it sounded identical to BLOW, listeners would be less likely to identify any BELOW stimulus as BLOW under any conditions. Under this criterion, RACE-ERASE, MUSE-AMUSE, and WAKE-AWAKE were not used because even in the most casual speech style, speakers did not delete the vowel that formed the first syllable of the disyllabic word. Among the s-words, SECEDE was produced with only a long /s/ in the CS style by R.K. A spectrogram of this production showed that there were two peaks of intensity in the /s/. Preliminary synthesis of SECEDE with only a lengthened /s/ was not very successful, and indicated that the intensity curve of the /s/ probably was a significant cue. Consequently, that pair of words was discarded. For the same reason, STAIN-SUSTAIN was not used since STAIN with a long /s/ never really sounded like
SUSTAIN. In SCUM-SUCUM, the disyllabic word was never produced with the first vowel deleted so that pair was discarded. In SPORT-SUPPORT, SUPPORT was produced without the first vowel in the CF style by R.K., so that pair was included in the perception experiment. Among the pairs of words with stop-vowel-liquid sequences, all pairs were used except TRAIN-TERRAIN. In some dialects, initial tr is pronounced [c] so that if any of my identification test subjects turned out to be speakers of such a dialect, they may have been unable to accept any TRAIN-TERRAIN stimuli as tokens of the word train. The seven pairs chosen were BLOW-BELOW, PLIGHT-POLITE, DRESS-DURESS, CREST-CARESSED, BROKE-BAROQUE, PRAYED-PARADE AND SPORT-SUPPORT.

3. Acoustic Analysis of Speech Style Differences

In order to examine differences among the four speech styles, spectrograms were made of the 14 key words, some on a Voiceprint 7000 series spectrograph at The Ohio State University and others on a Kay Sonograph at Indiana University. A conversion scale that had previously been worked out for the measurement of duration gave 76 msec for 10 millimeters on the spectrogram. Table 2.1 gives the durations of the key words in the four speech styles. For purposes of the following generalizations, Style will refer to the Formal versus Casual dimension, and Rate will refer to the Slow versus Fast dimension. From Table 2.1, the
<table>
<thead>
<tr>
<th></th>
<th>FS</th>
<th>FF</th>
<th>CS</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLOW</td>
<td>305</td>
<td>185</td>
<td>200</td>
<td>180</td>
</tr>
<tr>
<td>BELOW</td>
<td>340</td>
<td>210</td>
<td>295</td>
<td>190</td>
</tr>
<tr>
<td>PLIGHT</td>
<td>325</td>
<td>210</td>
<td>350</td>
<td>245</td>
</tr>
<tr>
<td>POLITE</td>
<td>480</td>
<td>340</td>
<td>370</td>
<td>275</td>
</tr>
<tr>
<td>PRAYED</td>
<td>380</td>
<td>230</td>
<td>290</td>
<td>230</td>
</tr>
<tr>
<td>PARADE</td>
<td>510</td>
<td>270</td>
<td>470</td>
<td>235</td>
</tr>
<tr>
<td>CREST</td>
<td>365</td>
<td>325</td>
<td>335</td>
<td>300</td>
</tr>
<tr>
<td>CARESSSED</td>
<td>470</td>
<td>395</td>
<td>400</td>
<td>335</td>
</tr>
<tr>
<td>BROKE</td>
<td>400</td>
<td>310</td>
<td>260</td>
<td>235</td>
</tr>
<tr>
<td>BAROQUE</td>
<td>420</td>
<td>325</td>
<td>365</td>
<td>280</td>
</tr>
<tr>
<td>DRESS</td>
<td>455</td>
<td>290</td>
<td>310</td>
<td>265</td>
</tr>
<tr>
<td>DURESS</td>
<td>435</td>
<td>340</td>
<td>400</td>
<td>310</td>
</tr>
<tr>
<td>SPORT</td>
<td>425</td>
<td>290</td>
<td>435</td>
<td>310</td>
</tr>
<tr>
<td>SPORTS</td>
<td>380</td>
<td>355</td>
<td>365</td>
<td>280</td>
</tr>
</tbody>
</table>

Table 2.1. Duration in msec of the key words in the four speech styles. (Because sports was used in the casual style and sport in the formal style, they are listed separately.)
following statistics emerge:

1. Words in the Casual style are an average of 16% shorter than words in the Formal style.
2. Words spoken at the Fast rate are an average of 25% shorter than words at the Slow rate.
3. Words in the Formal Fast combination are an average of 28% shorter than the Formal Slow.
4. Words in the Casual Fast combination are an average of 23% shorter than the Casual Slow.
5. Words in the Formal Fast combination are an average of 16% shorter than the Casual Slow.
6. Words in the Casual Fast combination are an average of 36% shorter than the Formal Slow.
7. Words in the Casual Slow combination are an average of 18% shorter than the Formal Slow.
8. Words in the Casual Fast combination are an average of 12% shorter than the Formal Fast.

As expected, the greatest durational difference occurs between the Formal Slow and Casual Fast styles. Looking at the effects of Rate versus Style, it appears that Rate has a

\[\text{It must be remembered that the sentence contexts for the Formal style were different from the Casual style. The original purpose of the recording was simply to elicit the key words in four speech styles, and the necessity for a uniform sentence context was overlooked. This should be kept in mind when comparing the word durations across styles, since considerations like sentence length, position in paragraph (see Lehiste, 1975), and the possibility for changing styles in mid sentence (Shockey, 1973) may have had varying effects on the words in different contexts.}\]
greater effect on duration than Style. That is, Casual words are shorter than formal words by 16%, but Fast words are shorter than Slow words by 25%. When the two dimensions are combined, Rate provides the greatest effect, as evidenced by the fact that Formal Fast words are 16% shorter than Casual Slow words. What these differences could reflect is that (1) Rate differences may be easier to produce than Style differences, and (2) Style differences may be produced by different mechanisms than Rate differences. Regarding the first, subjects informally said it was more difficult to sit and talk casually in an echo-free chamber than to sit there and talk formally. Regarding the second, the acoustic analysis below provides some evidence which indicates that casual speech, while usually of shorter duration than formal speech, has additional articulatory characteristics.

Table 2.2 presents the durations of the liquids /l/ and /r/ in the liquid pairs, and the presence of /ə/ and

---

1When measuring the duration of linguistic units, the problem arises as to how to measure words (phrases, sentences, etc.) in which elements have been deleted. One might use a slightly abstract standard duration for a word, based on a moderate rate and fairly careful pronunciation. Then, regardless of deletions, like cluster simplification, vowel deletion, etc., the word is measured from that baseline. But another method would be to make a standard duration based on the addition of each segment length. Then if a segment is deleted, that segment is not counted in the total. This method has problems too because segmentation of the speech continuum is often difficult (if not impossible) and as a result, very arbitrary decisions on segment duration would have to be made. This problem has been addressed in Goldman-Eisler (1961).
<table>
<thead>
<tr>
<th></th>
<th>BLOW</th>
<th>BELOW</th>
<th></th>
<th>PLIGHT</th>
<th>POLITE</th>
<th></th>
<th>PRAYED</th>
<th>PARADE</th>
<th></th>
<th>SPORT</th>
<th>SUPPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FS</td>
<td>FF</td>
<td>CS</td>
<td>CF</td>
<td>FS</td>
<td>FF</td>
<td>CS</td>
<td>CF</td>
<td>FS</td>
<td>FF</td>
<td>CS</td>
</tr>
<tr>
<td>[a]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>55</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[l]</td>
<td>70</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>70</td>
<td>40</td>
<td>60</td>
<td>35</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[p]</td>
<td>ASPIRATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[l]</td>
<td>-</td>
<td>15</td>
<td>-</td>
<td>55</td>
<td>55</td>
<td>10</td>
</tr>
<tr>
<td>[r]</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>20</td>
<td>40</td>
<td>90</td>
<td>85</td>
<td>40</td>
<td>90</td>
<td>40</td>
<td>85</td>
</tr>
<tr>
<td>[a]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>90</td>
<td>40</td>
<td>85</td>
<td>40</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[p]</td>
<td>ASPIRATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[l]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>120</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>[r]</td>
<td>40</td>
<td>15</td>
<td>15</td>
<td>20</td>
<td>120</td>
<td>40</td>
<td>70</td>
<td>30</td>
<td>90</td>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>[k]</td>
<td></td>
<td></td>
<td>ASPIRATION</td>
<td></td>
<td>[a]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>55</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[p]</td>
<td>CLOSURE</td>
<td>105</td>
<td>105</td>
<td>90</td>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>[r]</td>
<td>55</td>
<td>30</td>
<td>40</td>
<td>40</td>
<td>105</td>
<td>90</td>
<td>115</td>
<td>90</td>
<td>105</td>
<td>105</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[a]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[l]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.2 Average Duration (in msec) of Segments in the Key Words
aspiration of /p/ in SPORT-SUPPORT. The criteria for segmentation were based on Peterson and Lehiste (1960). For example, in a pair like PLIGHT-POLITE, aspiration of the voiceless stop is measured, but it is considered as an anticipatory voiceless version of the following vowel or liquid. It is perhaps arbitrary, but the liquids have been measured from the preceding stop through their steady states; that is, until the beginning of the transition to the following vowel.

A comparison of Tables 2.1 and 2.2 shows that overall word duration in the four speech styles does not always correlate with segment duration in the four styles. Only PARADE, DURESS and BAROQUE show that duration proceeds from longest to shortest for both the whole word and the /r/ in the order FS>PF>CS>CF. In all other cases, while the FS is usually the longest for the word length and /r/ and /l/ segment length, and CF the shortest, there is a fair amount of variability. For example, in PRAYED word duration shows FS>CS>PF=CF but the /r/ duration shows FS>CF>PF>CS; and in CREST word duration gives FS>CS>FF>CF and /r/ duration gives FS>PF=CS=CF. It appears that segment shortening (or lengthening) will not show a regular correlation with word duration. This is particularly the case where acoustic features, like aspiration, are deleted altogether. It seems premature to generalize from such a small group of speakers and from such a small corpus, but the trend here does offer
some support for the claims of Kozhevnikov and Chistovich (1965), whose research showed that changes in the rate of speech do not affect the relative durations of syllables and word intervals within a phonological phrase (their syntagma), but that consonants and vowels do change their relative durations.

Turning to some other acoustic characteristics of the key words in the different speech styles, the following observations can be made:¹

(1) BAROQUE. The CS version shows an /u/-offglide on the stressed vowel (typical of both speakers' dialects) but the CF version does not have an offglide.

(2) CARESSED. In CS, F₂ and F₃ reach a steady state for the /e/ and remain there; in CF, F₂ and F₃ only point toward that region.

(3) CREST. The CF version shows that the final /t/ has been deleted. (It is also interesting that the CS version, with /t/, has a shorter /s/ than the CF with no /t/. One interpretation is to speak of compensation. That is, in order to preserve some predetermined word length, the /s/ is lengthened to compensate for the loss of the /t/. Another interpretation is to say that in order to affect 'ease of articulation' in casual speech, the /t/ undergoes

¹Spectrograms of the key words illustrating the four speech styles can be obtained from the author on request.
spirantization since it already shares place of articulation and voicing with the preceding /s/. Again, overall word duration is preserved. Without knowing more about higher level language processing, it doesn't seem possible to choose between these alternatives.)

(4) SUPPORT. The first unstressed vowel is completely deleted in the CF style.

(5) PARADE. The CF does not show a steady state for the offglide of the stressed vowel, although the CS version does. Also the FS shows a clear transition from the stressed vowel offglide to the final /d/, but the FF does not.

(6) BELOW. There is no offglide on the stressed vowel in CF, but FS, FF, and CS do have an offglide.

(7) PLIGHT. The CF version has a lower F2 for the steady state of the vowel offglide than CS, indicating some centralization of the vowel, typical of a casual speech style.

(8) POLITE. In the CF style, F2 does not reach a steady state for the offglide of the stressed vowel.

(9) SPORT. In the FF style, the final /t/ is a short voiced flap, but the FS /t/ is not. (The following word begins with a vowel so the context for flapping is available.)

To summarize these characteristics, speeding up speech rate is undoubtedly responsible for the steady states of
spirantization since it already shares place of articulation and voicing with the preceding /s/. Again, overall word duration is preserved. Without knowing more about higher level language processing, it doesn't seem possible to choose between these alternatives.)

(4) SUPPORT. The first unstressed vowel is completely deleted in the CF style.

(5) PARADE. The CF does not show a steady state for the offglide of the stressed vowel, although the CS version does. Also the FS shows a clear transition from the stressed vowel offglide to the final /d/, but the FF does not.

(6) BELOW. There is no offglide on the stressed vowel in CF, but FS, FF, and CS do have an offglide.

(7) PLIGHT. The CF version has a lower F₂ for the steady state of the vowel offglide than CS, indicating some centralization of the vowel, typical of a casual speech style.

(8) POLITE. In the CF style, F₂ does not reach a steady state for the offglide of the stressed vowel.

(9) SPORT. In the FF style, the final /t/ is a short voiced flap, but the FS /t/ is not. (The following word begins with a vowel so the context for flapping is available.)

To summarize these characteristics, speeding up speech rate is undoubtedly responsible for the steady states of
some vowels not being reached and clear transitions to the
following consonants not being present. Lindblom (1963), in
discussing vowel target undershoot, says:

"As a vowel becomes shorter, there is less and less
time for the articulators to complete their 'on-' and
'offglide' movements within the CVC syllable. Provided
that the neural events corresponding to the
phonemes actually stay invariant, the speech organs
fail, as a result of the physiological limitations, to
reach the positions that they assume when the vowel is
pronounced under ideal steady-state conditions. In
the acoustic domain, this is paralleled by undershoot
in the formant frequencies relative to the bulls-eye
formant pattern." (p. 1779)

In other words, the timing of linguistic units is a
neurological activity and the articulators are simply
following orders. (Cf. Hammarberg 1976 for an argument
against an undershoot theory.) It is also interesting that
the greatest changes occur between the Casual Fast style and
all the others, instead of between Casual and Formal.
Apparently, although overall word durations vary among all
four combinations, it takes a maximally casual, sloppy style
for durational differences to have any substantive effect on
segments.

4. The Perception of Speech Style and Rate

As a corollary to the study of the production of the
key words in the four speech styles, I ran an experiment to
determine whether listeners could perceive differences among
the styles (Semiloff, 1975). From the original recordings
(using data from speakers G.E. and R.K. only), I chose 108
full sentences: 24 each of Formal Slow and Formal Fast and 30 each of Casual Slow and Casual Fast. The sentences were copied and randomized, and presented as a perceptual listening test to two groups of 30 subjects each, students at the Ohio State University. Both groups were given the same instructions but in reverse order. Group I was told to listen to each sentence and make two decisions about it -- (1) to determine whether it was fast or slow and (2) to determine whether it was formal or casual. Group II was told to make the style decision first (formal-casual) and the rate decision second (fast-slow). The different orders were used to determine whether style decisions affected rate decisions or vice versa. Answer sheets were provided. The styles and rates were defined for the listeners exactly as they had been for the speakers.

Figure 2.1 gives the percent of correct identification in terms of the style and rate intended by the speakers. The overall percent correct for Group I (rate decision first) was 80% and for Group II (style decision first) was 79%. This difference was nonsignificant, and the hypothesis that the order of judgments makes no difference was accepted. (Test for Significance of Difference between two Proportions, z=1.42, p=.05). But although overall order was nonsignificant, one difference did occur when style and rate were separated. Subjects did better on rate when they made the rate decision first. Group I had 84% correct and Group
Figure 2.1 Percent Correct for Style and Rate
II 81% on rate and this difference was significant (z=3.33, p=.05).

Figure 2.2 shows the overall percent correct for each of the four style-rate combinations. For both groups, the Formal Slow and Casual Fast sentences were more often correctly identified than the 'mixed' combinations, Formal Fast and Casual Slow. This fact related directly to the question of the independence of rate and style. A Pearson Product-Moment Correlation test was used to establish the correlation, if any, between rate and style decisions. Results showed that rate decisions were in fact made independently of style decisions. Only 7 of the 60 subjects had Pearson r's that reached significance at the .05 level. The hypothesis that rate can be perceived independently of style was accepted.

Of some interest is the effect of pauses on rate identification. Only the Formal sentences were long enough to contain pauses, and of the 10 Formal Fast productions spoken with pauses of at least 300 msec, 8 received less than 50% identification as fast. These results support the research of Goldman-Eisler (1961), who found that when the flow of speech is broken up by pauses, it is experienced as slow speech. The question though is whether the pause itself was responsible for the wrong rate identification, or whether those 10 sentences were longer despite the pauses. Because there were two speakers, many sentences occurred
Figure 2.2 Percent Correct for Style-Rate Combinations
twice on the test tape, and of the 10 pause sentences six also occurred in the Formal Fast combination, but with no pause present. In all six cases, when the pauses were subtracted, the sentences were still longer by an average of 850 msec than their counterparts produced with no pauses. That is, if a sentence was long enough to require a pause, it was spoken more slowly, even when the pause was not included. Apparently both the pause and the overall greater duration were responsible for the identification of these Formal Fast sentences as Slow.

Finally it is interesting to look at the patterns of errors that listeners made. Table 2.3 is a confusion matrix showing the distribution of responses. Among the Formal Slow, Casual Slow, and Casual Fast sentences, the majority of errors were errors of style. For example, instead of identifying a sentence correctly as Formal Slow, subjects tended to call it Casual Slow; Casual Slow sentences were thought to be Formal Slow; and Casual Fast sentences were taken to be Formal Fast. Among the Formal Fast sentences, due to the inclusion of pauses as discussed above, errors were mostly rate.

To summarize, as expected, the greatest differences in production and perception occurred in the maximally differentiated styles: Formal Slow and Casual Fast. In production, there is more shortening in the Formal Fast than Casual Slow combination; thus rate has more effect on
<table>
<thead>
<tr>
<th>ACTUAL SPEECH STYLE</th>
<th>FORMAL-SLOW</th>
<th>FORMAL-FAST</th>
<th>CASUAL-SLOW</th>
<th>CASUAL-FAST</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORMAL-SLOW</td>
<td>86</td>
<td>--</td>
<td>14</td>
<td>--</td>
<td>100</td>
</tr>
<tr>
<td>FORMAL-FAST</td>
<td>32</td>
<td>43</td>
<td>9</td>
<td>16</td>
<td>100</td>
</tr>
<tr>
<td>CASUAL-SLOW</td>
<td>31</td>
<td>5</td>
<td>51</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td>CASUAL-FAST</td>
<td>2</td>
<td>16</td>
<td>7</td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2.3. Confusion Matrix, All Responses, All Subjects
duration than style. In comparing word and segment length, it was found that their durations do not always follow the same order. In perception, listeners can identify rate and style separately fairly well; but in combination, the Formal Fast and Casual Slow styles cause confusion. Finally the purpose of gathering the production and perception data was to provide a data base for the perception experiments to be described in Chapter III.
Chapter III. THE CATEGORICAL PERCEPTION EXPERIMENT

0. Introduction

The primary question to which this chapter addresses itself is whether the duration of the steady state portion of a segment in a word can serve as an acoustic cue in the perception of syllabic complexity. In other words, in a pair of words like DRESS-DURESS, which differ primarily in that DURESS has a syllabic /r/ and DRESS has a non-syllabic /r/, can that difference in the number of syllables be represented by simply increasing the duration of the steady state portion of the /r/? The second question is whether listeners can judge speech stimuli as belonging to a particular speech style and rate. That is, can listeners make a judgment that a stimulus belongs to one linguistic category when told to judge that stimulus from the reference of a particular speech style, for example, Formal Fast, but to judge the same stimulus as belonging to another linguistic category when told to judge it from the reference of a different speech style, for example, Casual Slow? The perception experiment described in Chapter II did show that listeners can perceive speech style differences when complete natural speech sentences are used. But whether a similar task can be done when the stimuli are synthetically
produced single words is another question.

In this Chapter, I describe two experiments which were designed to explore the categorical nature of the perception of segment duration as a cue to the perception of syllabic nature. Briefly, categorical perception refers to the fact that listeners perceive speech stimuli which are varied in a step-wise fashion, not in equal steps along a continuum, but as quantal jumps from one linguistic category to another. (Liberman 1972) Categorical perception is not a characteristic of non-speech auditory events, because they have been shown to be perceived in a more nearly continuous fashion (Mattingly, Liberman, Syrdal and Halwes, 1971). It is only when acoustic stimuli are presumed to be speech, that is, when listeners are in the speech mode, that they predispose themselves to categorize such stimuli as belonging to some linguistic category¹.

In the following sections, I first describe an identification (or labeling) test, in which subjects listen to various speech stimuli and make a forced choice decision as to which linguistic category each stimulus belongs; and second, a discrimination test in which subjects are asked to

discriminate between stimuli which are either one, two, or three steps apart along a particular acoustic continuum. According to the traditional categorical perception model, if listeners are able to categorize stimuli in the labeling test into one or another linguistic category, they should be unable to discriminate between stimuli which fall within the same category. However, they should be able to discriminate stimuli which fall on either side of the linguistic boundary.

1. Identification Experiment

1.1. Design and Method

Based on the production experiment described in Chapter II, I chose the following seven word pairs for the identification test: BLOW-BELOW, PLOIGHT-POLITE, BROKE-BAROQUE, DRESS-DURESS, CREST-CARESSED, PRAYED-PARADE, and SPORT-SUPPORT. In the first six pairs, the segment of interest was the liquid following the initial stop, and in the seventh pair, the feature was the aspiration of /p/.

The monosyllabic members of each pair were synthesized on an OVE III C serial synthesizer at Haskins Laboratories. A DDP 224 computer facility provided the analysis of the natural speech words, which were recorded in an anechoic chamber by one male speaker. The input waveform was analyzed by a Ubiquitous Spectrum Analyzer, model U64A, and was inputted to an analog to digital converter. The output
of the Ubiquitous Spectrum Analyzer consisted of 128 frequency channels, each channel 40 cycles tall and the series of 128 channels was repeated every 12.8 msec. The sampling rate was 10,000 spectral samples per second.

The OVE III C synthesizer parameters used were (1) Hisp Amplitude, used for the simulation of aspiration of the initial voiceless stops in PLIGHT, CREST, and PRAYED, the release of /p/ in SPORT and the release of the final stops in SPORT, BROKE, PLIGHT and CREST; (2) Vowel Amplitude; (3) Fundamental Frequency set at 120 Hz throughout; (4) First Formant; (5) Second Formant; (6) Third Formant; (7) Bandwidth of \( F_1 \); (8) Bandwidth of \( F_2 \); (9) Bandwidth of \( F_3 \); (10) Fricative Amplitude used for the /s/ in DRESS, CREST and SPORT; (11) Fricative Formant 1; and (12) Fricative Formant 2, both also used to synthesize /s/. The Nasal Formant was set at 200 Hz throughout.

The preparation of the stimuli proceeded in the following way: The seven key words were recorded in list fashion by one male speaker who used a moderate conversational speech style and rate. The words were digitized and both natural and stylized spectrograms were examined in order to locate some portion of the /l/ in BLOW and PLIGHT and the /r/ in BROKE, DRESS, PRAYED and CREST which could be said to constitute the 'steady state'. In the case of /l/ in BLOW and PLIGHT, the steady state was considered to be the time period immediately before the
transition to the following /o/ and /ay/, respectively. It was the \( F_1 \) transition which was crucial, since /l/ is characterized by a rapid \( F_1 \) transition which precedes the \( F_2 \) transition. Also \( F_3 \) was maximally high at the point before it lowered to the next vowel. For BLOW, /l/ had the following formants during its steady state: \( F_1=479 \) Hz, \( F_2=951 \) Hz, \( F_3=3063 \) Hz. For PLIGHT, \( F_1=549 \) Hz, \( F_2=1029 \) Hz, \( F_3=2998 \) Hz. In the case of /r/, the steady state was considered to be the lowest value of \( F_3 \) before its transition to the following vowel. The formant frequencies for BROKE were \( F_1=511 \) Hz, \( F_2=1172 \) Hz, \( F_3=1782 \) Hz; for DRESS, \( F_1=469 \) Hz, \( F_2=1415 \) Hz, \( F_3=1795 \) Hz; for CREST, \( F_1=489 \) Hz, \( F_2=1233 \) Hz, \( F_3=1542 \) Hz; for PRAYED, \( F_1=518 \) Hz, \( F_2=1189 \) Hz, \( F_3=1670 \) Hz. The portion of SPORT which was of interest was the period of aspiration after release of the /p/.

Since the variable I wanted to manipulate was duration, the frame that constituted the middle of the steady state of the /l/, the /r/, and the aspiration of /p/ was incremented in steps of 10 msec, until each key word sounded impressionistically like its disyllabic counterpart. A 10 msec increment was chosen because, according to Lehiste (1970), the difference limen for duration is between 10 and 40 msec, so that 10 msec gave the most stringent test for categorical perception. DRESS, for example, had a steady state /r/ of 30 msec. The middle frame of /r/ was increased 13 times in 10 msec steps, until there was no doubt that the
word DURESS would be perceived. At that point the /r/ steady state was 150 msec long. The same procedure was used for the other words, the only difference being that for PLIGHT the steady state of /l/ was only 20 msec long, so the stimuli ranged from PLIGHT with a 20 msec /l/ steady state to POLITE with a 140 msec steady state, again 13 stimuli in total. BLOW, CREST, PRAYED and BROKE all originally had steady state liquids of 30 msec, so in their longest versions, the steady states were 150 msec. For SPORT, the aspiration on the /p/ was 20 msec long so 13 increments produced a word with 140 msec of aspiration. (As later results will show, this word was never perceived as SUPPORT.)

The 13 versions of each of the seven words gave 91 separate stimuli. The stimuli were randomized and repeated twice for the test tape. Ten dummy stimuli were inserted at the beginning to familiarize subjects with synthetic speech and their particular task (described below). The final tape then consisted of 192 items, with four seconds of silence between items in which subjects were to mark their responses on the answer sheet. After every ten words, a natural speech recording announcing the upcoming number was spliced into the tape in order to help subjects keep their place and to break the monotony of hearing the synthetic stimuli.

The identification task was for the subjects to make a forced choice decision as to which linguistic category each
stimulus belonged: the monosyllabic word or the disyllabic word. For each stimulus, the subject had to check off which of two words he thought he heard: BROKE or BAROQUE, CREST or CARESSED, PRAYED or PARADE, DRESS or DURESS, BLOW or BELOW, PLIGHT or POLITE, SPORT or SUPPORT.

One additional aspect of the final test tape must be mentioned. I wanted to know whether it would be possible to shift the listener's perception from the monosyllabic to the disyllabic word (or vice versa) by using a precursor frame intended to suggest a particular speech style. To put subjects into the appropriate 'speech style mode', there were four different versions of the final tape. Each tape had a natural speech precursor phrase before each stimulus. The phrase was "The word you will hear next is ______. The four speech styles were those described in Chapter II.

Each of four groups of subjects was told to try to judge the words on the tape as though they were being spoken in one of the four style-rate combinations. For example, the Formal Slow group received the following instructions:

-------

1Due to a slight communication error the precursor sentence which was recorded was not exactly the one I had specified. I had wanted the sentence to be "The word you are going to hear next is ______", instead of "The word you will hear next is ______". The former sentence, being longer, would have offered more cues to the listener as to the speech style and would have allowed more phonetic reduction in the casual speech forms. Whether this error was responsible for any part of the results of the identification is hard to judge.
"The following experiment is an identification test of synthetically produced English words. Each word will be preceded by the following careful, slowly articulated natural frame sentence: 'The word you will hear next is _ _ _ _ _'. You are to put a check mark next to the word on the answer sheet that comes closest to being identical to the synthetic word. For purposes of this experiment, pretend you are listening to slow, formal, carefully articulated speech and judge each synthetic word against this formal, slow criterion. For example, if the word in number 1 sounds like blow if you were judging it according to sloppy, conversational speech but below if you were judging it according to a slow formal speech style, then check below..."

A fifth group of subjects, used as a control, received no speech style instructions at all. They were simply told to identify the stimuli, and on their tape, there was no precursor sentence before each stimulus.

The subjects for the identification test consisted of five groups of 15 listeners in each group. Subjects were students in an introductory psychology class at Indiana University who received credit towards an experiment-participation requirement. In order to make an attempt at dialect homogeneity, in case that turned out to be significant, all subjects came from Illinois, Indiana or Ohio excluding the large metropolitan areas of Chicago, Cleveland and Cincinnati. That is not to say that speakers from northern Indiana and southern Indiana speak the same dialect (they do not), but within the limits of subject availability, I imposed the strictest constraints possible. I do not think that dialect differences were responsible for any part of the results to be described below.
Subjects took the 25 minute identification test in groups of 15. They heard the tape in a sound treated room over headphones. Some discussion of the instructions preceded the test since subjects usually had questions about the 'speech style mode'. The experiment did not begin until all the listeners were sure that they understood the instructions.

1.2. Results

The results of the identification test can be shown in the form of graphs. Since the synthesized words at the short end of the scale (those with the shortest steady state liquid or /p/ aspiration) are presumed to be perceived as monosyllabic, and those at the long end of the scale, to be perceived as disyllabic, the purpose of the graph is to show where the crossover lies. That is, at what point do the identifications cross over from the monosyllabic word to the disyllabic word?

1.2.1. SPORT-SUPPORT

The first issue to be disposed of is the question of the SPORT-SUPPORT pair. Recall that SUPPORT was meant to be simulated by increasing the duration of the aspiration on the /p/ in SPORT. Figure 3.1 shows the pattern of identification for the five groups combined. (as I explain in the next section, there is no statistical reason to keep
Figure 3.1 Crossover, SPORT-SUPPORT, All Groups
the five groups separate, so Figure 3.1 is an accurate representation of the SPORT-SUPPORT responses.) It is obvious that there was no crossover. Most subjects perceived the stimuli as SPORT no matter how long the /p/ aspiration became. It is interesting though, that as the aspiration increases, the two curves do approach each other, indicating more guessing. My interpretation of the increased guessing is that at long durations of /p/ aspiration (more than 70 msec) the stimuli began sounding unnatural. Since the task of the subjects was a forced binary choice, the long aspiration stimuli did sound more like SUPPORT than SPORT, so more responses went in that direction. Although confidence ratings for subjects' responses were not elicited, I suspect that the SUPPORT responses to the long /p/ aspiration would have been given very low ratings. With respect to the question of categorical perception of syllabic structure, one has to conclude that lengthening the aspiration of a voiceless stop in a context like SPORT does not result in perception which crosses over to SUPPORT. Rather, the long aspiration stimuli sound like someone is saying SPORT with an unnaturally exploded /p/. In this example, duration of aspiration is not an acoustic cue to the perception of syllabic structure. The SPORT-SUPPORT stimuli will not be considered any further.
1.2.2. Speech Style Differences.

Table 3.1 presents the crossover durations for each word pair and each speech style group. For example, for the BLOW stimuli, the Formal-Slow group showed a crossover at 86 msec, Formal-Fast at 78 msec, Casual-Slow at 86 msec, Casual-Fast at 83 msec, and the Control group at 92 msec. Crossovers were obtained for all of the six key words, although none of them showed particularly steep slopes, as one would expect in truly categorical perception. Statistically, differences due to segment duration were highly significant. Table 3.2 gives the results of an analysis of variance test for independence (two-factor mixed design: repeated measures on one factor). Treatment refers to variation due to changes in the stimuli. In Table 3.2, all five speech style groups are combined.

The most important question is whether there were any statistically significant differences among the five speech style groups. In fact, there were not. Table 3.3 gives the F-ratios for group differences. Treatment here refers to group variation. Only PRAYED showed an F-ratio that reached significance at the .04 level. What Table 3.3 means is that whatever variability there was among the five speech style groups was due to chance, and not to any variable under control in the experiment. Figures 3.2-3.7 give the identification percentages for each word pair, all five speech style groups combined. In subsequent analyses, the
<table>
<thead>
<tr>
<th>WORD PAIR</th>
<th>SPEECH STYLE</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FORMAL-SLOW</td>
<td>FORMAL-FAST</td>
<td>CASUAL-SLOW</td>
<td>CASUAL-FAST</td>
<td>CONTROL</td>
<td>AVERAGE</td>
<td></td>
</tr>
<tr>
<td>BLOW-BELOW</td>
<td>86</td>
<td>78</td>
<td>85</td>
<td>83</td>
<td>92</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>PLIGHT-POLITE</td>
<td>55</td>
<td>45</td>
<td>49</td>
<td>44</td>
<td>54</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>BROKE-BAROQUE</td>
<td>64</td>
<td>64</td>
<td>55</td>
<td>55</td>
<td>66</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>DRESS-DURESS</td>
<td>70</td>
<td>60</td>
<td>75</td>
<td>66</td>
<td>68</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>CREST-CARESSED</td>
<td>63</td>
<td>74</td>
<td>70</td>
<td>78</td>
<td>62</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>PRAYED-PARADE</td>
<td>70</td>
<td>67</td>
<td>59</td>
<td>65</td>
<td>52</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
<td>69</td>
<td>65</td>
<td>66</td>
<td>65</td>
<td>66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1. Crossover Durations (in msec) for the Six Word Pairs
<table>
<thead>
<tr>
<th>Key Word</th>
<th>SS (df: 12)</th>
<th>df</th>
<th>MS (840)</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLOW: STIMULUS DIFFERENCES</td>
<td>188.695</td>
<td>12</td>
<td>15.725</td>
<td>106.480*</td>
</tr>
<tr>
<td>Treatment</td>
<td>124.048</td>
<td>840</td>
<td>.148</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>312.743</td>
<td>852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLIGHT: STIMULUS DIFFERENCES</td>
<td>125.392</td>
<td>12</td>
<td>10.449</td>
<td>70.163*</td>
</tr>
<tr>
<td>Treatment</td>
<td>125.101</td>
<td>840</td>
<td>.149</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>250.493</td>
<td>852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BROKE: STIMULUS DIFFERENCES</td>
<td>170.339</td>
<td>12</td>
<td>14.195</td>
<td>114.046*</td>
</tr>
<tr>
<td>Treatment</td>
<td>104.552</td>
<td>840</td>
<td>.124</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>274.891</td>
<td>852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRESS: STIMULUS DIFFERENCES</td>
<td>182.424</td>
<td>12</td>
<td>15.202</td>
<td>117.486*</td>
</tr>
<tr>
<td>Treatment</td>
<td>108.691</td>
<td>840</td>
<td>.129</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>291.115</td>
<td>852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CREST: STIMULUS DIFFERENCES</td>
<td>144.811</td>
<td>12</td>
<td>12.068</td>
<td>83.289*</td>
</tr>
<tr>
<td>Treatment</td>
<td>121.705</td>
<td>840</td>
<td>.145</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>266.516</td>
<td>852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRAYED: STIMULUS DIFFERENCES</td>
<td>153.666</td>
<td>12</td>
<td>12.806</td>
<td>93.434*</td>
</tr>
<tr>
<td>Treatment</td>
<td>115.126</td>
<td>840</td>
<td>.137</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>268.792</td>
<td>852</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at \( p = .001 \)

Table 3.2. Stimulus Differences for the Key Words
### BLOW: SPEECH STYLE DIFFERENCES

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2.856</td>
<td>4</td>
<td>.714</td>
<td>.971</td>
</tr>
<tr>
<td>Error</td>
<td>51.477</td>
<td>70</td>
<td>.735</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>54.333</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PLIGHT: SPEECH STYLE DIFFERENCES

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>3.041</td>
<td>4</td>
<td>.760</td>
<td>.962</td>
</tr>
<tr>
<td>Error</td>
<td>55.293</td>
<td>70</td>
<td>.790</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>58.334</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### BROKE: SPEECH STYLE DIFFERENCES

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>4.279</td>
<td>4</td>
<td>1.070</td>
<td>1.951</td>
</tr>
<tr>
<td>Error</td>
<td>38.375</td>
<td>70</td>
<td>.548</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>42.654</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### DRESS: SPEECH STYLE DIFFERENCES

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2.513</td>
<td>4</td>
<td>.628</td>
<td>.583</td>
</tr>
<tr>
<td>Error</td>
<td>75.436</td>
<td>70</td>
<td>1.078</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>77.949</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### CREST: SPEECH STYLE DIFFERENCES

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>7.024</td>
<td>4</td>
<td>1.756</td>
<td>2.104</td>
</tr>
<tr>
<td>Error</td>
<td>58.421</td>
<td>70</td>
<td>.835</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>65.445</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PRAYED: SPEECH STYLE DIFFERENCES

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>4.110</td>
<td>4</td>
<td>1.027</td>
<td>2.556***</td>
</tr>
<tr>
<td>Error</td>
<td>28.134</td>
<td>70</td>
<td>.402</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>32.244</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***Significant at $a = .04$

Table 3.3. Speech Style Differences for the Key Words
Figure 3.2 Crossover, BLOW-BELOW, All Groups
Figure 3.3 Crossover, PLIGHT-POLITE, All Groups
Figure 3.4 Crossover, BROKE-BAROQUE, All Groups
Figure 3.5 Crossover, DRESS-DURESS, All Groups
Figure 3.6 Crossover, CREST-CARESSED, All Groups
Figure 3.7 Crossover, PRAYED-PARADE, All Groups
five groups will be collapsed.

1.2.3. Trial Differences

As explained above, each stimulus word occurred twice on the tape. This provided 30 responses to each word, although there were only 15 subjects in a group. In doing the statistical analysis, it turned out that this repeated measures design caused some significant differences. Table 3.4 shows the F-ratios for the trial differences. Here, Treatment refers to the two trials for each subject, each stimulus. What Table 3.4 means is that except for CREST and PLIGHT, part of the variation in the subjects' responses can be explained as due to the first versus the second occurrence of the word. Figures 3.8-3.13 present the trial differences graphically. The plots show the percent identification for the disyllabic member of each pair.

The figures indicate that at the longer durations of /l/ and /r/, subjects were somewhat more disposed to identify the word as disyllabic on the second trial than on the first trial. This tendency though, is not overwhelming in any sense. That the test results came out significant is due in part to the large numbers of degrees of freedom associated with the large numbers of subjects and stimuli. The results also possibly point to a slight learning effect. That is, by the time subjects listened to the stimuli the second time, they were used to listening to synthetic speech.
<table>
<thead>
<tr>
<th>Key Word</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLOW: TRIAL DIFFERENCES</td>
<td>Treatment</td>
<td>4.767</td>
<td>12</td>
<td>.397</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>93.323</td>
<td>840</td>
<td>.123</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>98.090</td>
<td>852</td>
<td></td>
</tr>
<tr>
<td>PLIGHT: TRIAL DIFFERENCES</td>
<td>Treatment</td>
<td>1.435</td>
<td>12</td>
<td>.120</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>108.528</td>
<td>840</td>
<td>.129</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>109.963</td>
<td>852</td>
<td></td>
</tr>
<tr>
<td>BROKE: TRIAL DIFFERENCES</td>
<td>Treatment</td>
<td>3.181</td>
<td>12</td>
<td>.265</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>85.017</td>
<td>840</td>
<td>.101</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>88.198</td>
<td>852</td>
<td></td>
</tr>
<tr>
<td>DRESS: TRIAL DIFFERENCES</td>
<td>Treatment</td>
<td>5.594</td>
<td>12</td>
<td>.466</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>78.992</td>
<td>840</td>
<td>.094</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>84.586</td>
<td>852</td>
<td></td>
</tr>
<tr>
<td>CREST: TRIAL DIFFERENCES</td>
<td>Treatment</td>
<td>2.553</td>
<td>12</td>
<td>.213</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>111.070</td>
<td>840</td>
<td>.132</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>113.623</td>
<td>852</td>
<td></td>
</tr>
<tr>
<td>PRAYED: TRIAL DIFFERENCES</td>
<td>Treatment</td>
<td>4.114</td>
<td>12</td>
<td>.343</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>96.790</td>
<td>840</td>
<td>.115</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>100.904</td>
<td>852</td>
<td></td>
</tr>
</tbody>
</table>

* Significant at α = .001
** Significant at α = .002

Table 3.4. Trial Differences for the Key Words
Figure 3.8 Trial Differences, BLOW-BELOW
Figure 3.9 Trial Differences, PLIGHT-POLITE
Figure 3.10 Trial Differences, BROKE-BAROQUE
Figure 3.11 Trial Differences, DRESS-DRESSS
Figure 3.12 Trial Differences, CREST-Caressed
Figure 3.13 Trial Differences, PRAYED-PARADE
and possibly were paying more attention to the task they were performing.

2. Discrimination Experiment

The traditional categorical perception paradigm includes discrimination testing as well as identification testing. In a discrimination experiment, subjects are asked to discriminate between stimuli that lie next to or near each other along some acoustic continuum. As Liberman, et al. (1957) state, "[Discrimination testing] is important...because it provides a basis for the question...whether or not, with similar acoustic differences, a listener can better discriminate between sounds that lie on opposite sides of a phoneme boundary than he can between sounds that fall within the same phoneme category." (p.358) In the present case, the 'phoneme boundary' refers to the 'syllabic boundary' between the monosyllabic and the disyllabic word. The strong version of the categorical perception model claims that listeners should be able to discriminate between stimuli only to the extent that they can label them. So, if 'a' is labeled as belonging to category X, and 'b' and 'c' to category Y, then 'a' should be able to be discriminated from 'b' and 'c', but 'b' and 'c' should not be discriminable from each other.

---

I would like to acknowledge the help of Hideki Kasuya of the Speech Communications Research Laboratory for his help in setting up the discrimination experiment and his advice and comments.
Researchers usually calculate the predicted discrimination function from the labeling data. (There are several mathematical methods for doing this, cf. Liberman, et al. 1957, Pollack and Pisoni 1971.) In most cases the obtained discrimination functions are in fact higher than the predicted. According to Studdert-Kennedy et al. (1970), "...the obtained function almost always lies somewhat above the predicted, indicating that there remains some basis for discrimination, however marginal, between stimuli that are placed in the same category." (p.236)

Discrimination tests can be done in various formats. The most widely used is the ABX procedure (Liberman, et al. 1957). In this case the subject listens to a triad of stimuli, usually from half a second to a second apart, and has to decide whether the third stimulus was more like the first or more like the second. The first two stimuli are always different, and the third is always identical to one of them. Another format has been called the 4IAX procedure (Pisoni 1971). In this test, two sequences of two stimuli each are presented to listeners: AA, AB or AB, AA. The listeners must decide whether the first pair was more alike than the second. The disadvantage of the ABX procedure is explained by Pollack and Pisoni (1971) as due to the fact that a heavy load must be placed on memory since the third stimulus must be compared both to the second and to the fairly remote first stimulus. But the 4IAX procedure avoids
this pitfall, since the listener makes only a pair-by-pair comparison. That is, it is never necessary to compare stimuli which are separated by an intervening stimulus. A variation of the 4IAX procedure is the 2IAX (Pollack and Pisoni, 1971) or simply the AX format (Fujisaki and Kawashima, 1970). Listeners hear one sequence of stimuli, either AA or AB and they indicate whether the two members sound the same or different. Like the 4IAX test, only a pairwise comparison is necessary. In all three types of discrimination test, the dimension along which the stimuli differ need never be specified by the experimenter, since only 'same-different' responses are indicated.

2.1. One-Step Discrimination

The stimuli to be compared are usually one step, two steps, or three steps apart along the acoustic continuum. The most stringent discrimination test is the one step.

Because of the large number of stimuli used in the labeling test (6 words x 13 steps) discrimination tests were only run on one of the key word pairs, PRAYED-PARADE. The ABX procedure was used, so that the test tape consisted of triads of the synthetic PRAYED-PARADE stimuli. The four possible combinations were ABA, ABB, BAA and BAB, each stimulus one step (=10 msec) away from the other. Three subjects (employees at the Speech Communications Research Laboratory) participated in the experiment. They were all
right-handed, with no history of hearing disorders and were native English speakers. The total number of different triads was 48 and they were randomized on four tapes, each triad appearing four times per tape. Each subject listened to each tape twice, during eight 25 minute listening sessions, taken on eight weekdays. Each subject responded to each triad a total of 32 times. There were four dummy triads at the beginning and end of each tape to avoid any 'primacy' or 'recency' learning effects.

Figures 3.14-3.17 show the results of the one step discrimination tests. The three subjects' scores are combined in each of the four stimulus-order treatments: ABA, ABB, BAA, and BAB. 'Stimulus number' refers to each ABX triad along the acoustic continuum. For example, stimulus 1 is the triad composed of the PRAYED-PARADE stimuli where the /r/ steady state is 30 msec long for A, 40 msec for B, and 30 msec for X. Table 3.5 gives the average percent correct and average standard deviations for each subject, and the statistical analysis of these results is presented in Table 3.6. It is clear from the figures that none of the subjects was able to discriminate among the stimuli when they were only one step apart. In other words, the 10 msec difference between the stimuli was simply too small to detect, even at the 'syllabic boundary', which for PRAYED-PARADE, fell at 63 msec, or at the fourth discrimination stimulus.
Figure 3.14 ABA One-Step Discrimination, All Subjects
Figure 3.15 ABB One-Step Discrimination, All Subjects
Figure 3.16  BAA One-Step Discrimination, All Subjects
Figure 3.17 BAB One-Step Discrimination, All Subjects
<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>TREATMENT</th>
<th>ABA</th>
<th>ABB</th>
<th>BAA</th>
<th>BAB</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average % Correct</td>
<td>47.66</td>
<td>59.12</td>
<td>55.21</td>
<td>47.92</td>
<td>52.47</td>
</tr>
<tr>
<td>2</td>
<td>Average % Correct</td>
<td>46.09</td>
<td>51.04</td>
<td>52.86</td>
<td>48.70</td>
<td>49.67</td>
</tr>
<tr>
<td></td>
<td>Av. Stnd. Deviation</td>
<td>7.75</td>
<td>10.58</td>
<td>9.51</td>
<td>10.11</td>
<td>9.89</td>
</tr>
<tr>
<td>3</td>
<td>Average % Correct</td>
<td>52.60</td>
<td>56.77</td>
<td>58.85</td>
<td>55.73</td>
<td>55.90</td>
</tr>
</tbody>
</table>

| Average | Average % Correct | 47.78 | 55.64 | 55.64 | 50.78 | 52.71 |
|         | Av. Stnd. Deviation | 7.16 | 5.93 | 5.27 | 2.98 | 6.33 |

Table 3.5. Average Percent Correct and Standard Deviations for the Three Subjects, One-Step Discrimination.
### Differences Among the Four ABX Orders (Treatments)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>85.46</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>148.61</td>
<td>3</td>
<td>49.54</td>
<td>5.90</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Error</td>
<td>50.35</td>
<td>6</td>
<td>8.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>284.42</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Differences Among the Three Subjects (Treatments)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>148.61</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>85.46</td>
<td>2</td>
<td>42.73</td>
<td>5.09</td>
<td>&lt;.10</td>
</tr>
<tr>
<td>Error</td>
<td>50.35</td>
<td>6</td>
<td>8.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>284.42</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.6. Analysis of Variance One-Step Discrimination
The top part of Table 3.6 shows that there were differences among the four stimulus-order treatments, which were statistically significant at the .05 level. In order to determine which of the groups differed from each other both a T-test for differences among several means and Duncan's Multiple Range Test were performed. The results were identical and showed that the ABA order was significantly different from the ABB and BAA orders (t=2.31, 8 df, p=.05). The bottom part of Table 3.6 shows that there were no significant differences among the three subjects. (The F-ratio does not reach significance until p=.10.)

The treatment effect is probably due to the inherent bias in the ABX procedure. In the ABB and BAA orders, the X-stimulus is identical to the immediately preceding stimulus. Apparently, when X is identical to stimulus two instead of stimulus one, similarity is easier to perceive, at least for the one step discrimination. This tendency offers some evidence for the well known short term memory effects. That is, stimulus one has been lost (or partially lost) to short term memory by the time stimulus three reaches it, although stimulus two is still available for comparison. When stimuli two and three are the same, the auditory image left by stimulus two facilitates making an accurate comparison, although when they are different, it does not.
2.2. Two-Step and Three-Step Discrimination

The same three subjects from the one-step test took the two- and three-step tests. The format was the same except for the following changes: For the two-step test, the number of different triads was 44 and each subject responded to each triad 20 times (instead of 32 times) over five sessions. There were two dummy triads (instead of four) at the beginning and end of each tape. For the three-step test, there were 40 different triads, subjects heard each triad 20 times over five sessions and no dummy triads were included on any of the tapes.

Figures 3.18-3.21 show the results of the two-step test. Although the individual subject's scores are not plotted, the data showed that there was a large amount of inter-speaker variation. In addition, the pattern of responses is clearly not categorical. There is no peak of discriminability at any stimulus along the continuum. (Recall that the syllabic boundary for PRAYED-PARADE lies at stimulus 4.) Finally, there is a large amount of intra-speaker variation. The average standard deviations for each speaker are large and only subject three shows an

---

1 Because of certain technical problems, the two- and three-step discriminations were run several weeks after the one-step test.
Figure 3.18 ABA Two-Step Discrimination, All Subjects
Figure 3.19 ABB Two-Step Discrimination, All Subjects
Figure 3.20  BAA Two-Step Discrimination, All Subjects
Figure 3.21 BAB Two-Step Discrimination, All Subjects
average discrimination curve slightly above chance\(^1\). These numbers are given in Table 3.7.

Table 3.8 shows the results of analysis of variance on the two-step data. The top table gives no significant effect due to experimental treatment. The bottom table does show a significant effect due to subjects. This is not surprising, since it was mentioned above that subject three did considerably better than the other two. A t-test for differences among several means indicated that subjects one and two did differ significantly from subject three, but not from each other. \((t=2.26, 9\text{ df}, p=.05)\)

Figures 3.22-3.25 present the findings of the three-step discrimination tests. Like the two-step results, there is both inter- and intra-speaker variation and it is clear that perception is not categorical. Table 3.9 presents the average percent correct and average standard deviations for each subject and Table 3.10 gives the analysis of variance for treatments and subjects. The top part of Table 3.10 shows that there was a significant difference among the four treatments \((p=.005)\). A subsequent t-test for differences among means indicated that except for

\(^1\) Subject 3 did considerably better than the other subjects on both the two- and three-step tests. I think there are several possible reasons for this. First, she seemed to take the task more seriously than subject 2, who felt considerably inconvenienced and bored by it. Second, she had a strong musical background and I suspect she was simply better at perceiving durational differences among speech or non-speech stimuli.
<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>TREATMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABA</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Average &amp; Correct</td>
<td>53.64</td>
</tr>
<tr>
<td>Av. Stnd. Deviation</td>
<td>10.58</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Average &amp; Correct</td>
<td>44.54</td>
</tr>
<tr>
<td>Av. Stnd. Deviation</td>
<td>5.89</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Average &amp; Correct</td>
<td>65.45</td>
</tr>
<tr>
<td>Av. Stnd. Deviation</td>
<td>16.91</td>
</tr>
<tr>
<td>Average</td>
<td></td>
</tr>
<tr>
<td>Average &amp; Correct</td>
<td>54.54</td>
</tr>
<tr>
<td>Av. Stnd. Deviation</td>
<td>6.10</td>
</tr>
</tbody>
</table>

Table 3.7. Average Percent Correct and Standard Deviations for the Three Subjects, Two-Step Discrimination
### Differences Among the Four ABX Orders (Treatments)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>712.15</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>44.95</td>
<td>3</td>
<td>14.98</td>
<td>.89</td>
<td>---</td>
</tr>
<tr>
<td>Error</td>
<td>757.10</td>
<td>6</td>
<td>16.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>857.56</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Differences Among the Three Subjects (Treatments)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>44.95</td>
<td>3</td>
<td></td>
<td></td>
<td>&lt;.005</td>
</tr>
<tr>
<td>Treatments</td>
<td>712.15</td>
<td>2</td>
<td>256.07</td>
<td>21.27</td>
<td>&lt;.005</td>
</tr>
<tr>
<td>Error</td>
<td>757.10</td>
<td>6</td>
<td>16.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>857.56</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.8. Analysis of Variance Two-Step Discrimination
Figure 3.22 ABA Three-Step Discrimination, All Subjects
Figure 3.23 ABB Three-Step Discrimination, All Subjects
Figure 3.24 BAA Three-Step Discrimination, All Subjects
Figure 3.25  BAB Three-Step Discrimination, All Subjects
<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>TREATMENT</th>
<th>ABA</th>
<th>ABB</th>
<th>BAA</th>
<th>BAB</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average % Correct</td>
<td>53.50</td>
<td>67.00</td>
<td>65.50</td>
<td>62.00</td>
<td>62.00</td>
</tr>
<tr>
<td></td>
<td>Av. Stnd. Deviation</td>
<td>16.95</td>
<td>10.18</td>
<td>19.80</td>
<td>10.67</td>
<td>15.88</td>
</tr>
<tr>
<td>2</td>
<td>Average % Correct</td>
<td>46.00</td>
<td>58.00</td>
<td>53.50</td>
<td>49.50</td>
<td>51.75</td>
</tr>
<tr>
<td>3</td>
<td>Average % Correct</td>
<td>74.50</td>
<td>97.00</td>
<td>83.00</td>
<td>84.50</td>
<td>84.75</td>
</tr>
<tr>
<td></td>
<td>Av. Stnd. Deviation</td>
<td>26.45</td>
<td>4.64</td>
<td>15.39</td>
<td>5.75</td>
<td>17.72</td>
</tr>
<tr>
<td>Average</td>
<td>Average % Correct</td>
<td>58.00</td>
<td>74.00</td>
<td>67.33</td>
<td>65.33</td>
<td>66.17</td>
</tr>
<tr>
<td></td>
<td>Av. Stnd. Deviation</td>
<td>12.74</td>
<td>4.08</td>
<td>9.76</td>
<td>6.31</td>
<td>10.58</td>
</tr>
</tbody>
</table>

Table 3.9. Average Percent Correct and Standard Deviations for the Three Subjects, Three-Step Discrimination.
### Differences Among the Four ABX Orders (Treatments)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>2282.17</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>390.34</td>
<td>3</td>
<td>130.11</td>
<td>13.54</td>
<td>&lt;.005</td>
</tr>
<tr>
<td>Error</td>
<td>57.66</td>
<td>6</td>
<td>9.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>2730.17</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Differences Among the Three Subjects (Treatments)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>390.34</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>2282.17</td>
<td>2</td>
<td>1141.08</td>
<td>118.74</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Error</td>
<td>57.66</td>
<td>6</td>
<td>9.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>2730.17</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.10. Analysis of Variance, Three-Step Discrimination
order BAA and BAB, all other orders did differ significantly from each other. (t=2.31, 8 df, p=.05) The bottom part of Table 3.10 shows that there were differences among the three subjects (p=.001), and the t-test found differences among all three subjects (t=2.26, 9 df, p=.05).

In summary, the one-step and three-step discrimination tests showed that the stimulus order within the triad was in fact responsible for much of the variation that resulted among the four groups. In both cases, the ABB and BAA treatments showed a higher percentage of correct responses than the ABA and BAB treatments. As suggested earlier, I think this difference is attributable to the inherent bias in the ABX procedure.

Regarding differences among the three subjects, the two-step and three-step tests showed that significant differences existed. Considering the overall results of the one-step test, where none of the subjects was able to discriminate among the stimuli above chance level, it is no surprise that there were no differences among them. But once there was better than chance level discrimination, subject three was clearly better than the others.

Finally, the results of these tests leave no doubt that syllabicity was not perceived in a categorical fashion. There were no consistent peaks of discrimination for any of the stimuli.

The fact that there was so much variation within a
single speaker remains mysterious. For example, in the three-step ABA test, subject one had a range from 70% correct to 10% correct and subject three for the same test ranged from 100% correct to 40% correct. (The low ends of these subjects' ranges did not occur on the same stimuli.) One explanation might be that the subjects' attention spans waxed and waned so to speak, and inattention resulted in lower scores. If this had been the case, then one could hypothesize that the subjects were able to discriminate, at least the two- and three-step stimuli, better than their responses indicated. One way to test this hypothesis would be to separately analyze each half of each listening session, on the assumption that attention levels were higher at the beginning of a session than at the end. The data from the three-step test were analyzed this way and precisely the opposite results were obtained: all three subjects averaged slightly better results on the second half of each session than on the first half. Subject one averaged 15 more correct discriminations on the second half; subject two averaged six more correct; subject three averaged eight more correct. (The re-analysis was not done on the two-step data because the randomization of the stimuli did not evenly distribute the triads such that exactly half occurred in both halves of each experimental session.) Thus, the large degree of intra-speaker variation remains unexplained.
Chapter IV. CONCLUSIONS AND IMPLICATIONS

0. Introduction

The results of the labeling and discrimination experiments described in the preceding chapter constitute little more than facts and figures at this point. Graphs and statistics are useful only insofar as they can be interpreted within the context of a theory, or as the basis for testing hypotheses. The perception experiments were meant to test three hypotheses: 1) The suprasegmental feature, duration, is a cue to the perception of syllability. 2) The perception of syllability is categorical. 3) The use of a precursor phrase meant to suggest a particular speech style can shift listeners' perception from a monosyllabic to a disyllabic word (and vice versa). In this last chapter, I discuss to what extent these hypotheses are supported.

A certain amount of information about the production and perception of English liquids is necessary in order to understand the differences between syllabic and non-syllabic segments. This is undertaken in Section 1. In Section 2, I discuss the question of speech style differences and offer some suggestions as to why there were none. Section 3
focuses on the question of the categorical perception of syllabic ity. Section 4 suggests reasons for the differences in average crossover durations for the six key words. Section 5 attempts to relate the findings of this research to current theories of speech perception and to phonological theory. Section 6 consists of suggestions and implications for further research on the perception of syllabic ity, and the last Section is a summary of the research undertaken in this dissertation.

1. Syllabic and Non-Syllabic Liquids in Production

A very thorough study of the acoustical characteristics of /r/ and /l/ has been done by Lehiste (1964). She measured the formant frequencies and durations of numerous tokens of English liquids, not only in initial, intervocalic and final positions, but also in morphological contexts. Although initial stop-liquid clusters were not dealt with, some of her results can be generalized to that context.

Regarding the six key word pairs used in the present study, one question is whether the monosyllabic member constituted an example of an initial liquid and whether the disyllabic member constituted a syllable-final syllabic liquid. This question is central to the whole problem of segmentation of the speech continuum, an issue of great concern in computer assisted speech recognition studies. The problem, for example in POLITE, is whether the /l/ is
syllable-initial: [pʰə.laйт]; syllable-final: [pʰəʔ.ейт] or [pʰə.lейт]; or both: [pʰə.lейт] or [pʰəʔ.лейт]. Without employing very stringent and often arbitrary segmentation criteria, it is impossible to decide this question from the information seen in spectrograms. With respect to /l/ at least, researchers usually say that a velarized /l/, with its lower F₂, is characteristic of syllable-final position, so that simply measuring the F₂ frequency and having a non-velarized /l/ as a reference should provide information about position in the syllable. Likewise, if the /l/ is velarized at the onset and shows an increase in F₂, it may be possible to say that it is ambisyllabic. The latter case is especially hard to tell from a spectrogram and often other parameters, like overall duration or intensity are measured in order to attempt to locate syllable boundaries. /r/ is even more difficult to segment because it has no clear positional allophones. Compounding the problem is the high incidence of not only inter-speaker variation but intra-speaker variation. In both cases, speech style and rate of articulation can play confounding roles.

The Lehiste study measured real speech productions. In my study, the synthetic words were based on digitized representations. The question is whether the 'simulated' syllabic liquid stimuli, as in POLITE, PARADE, etc., produced by lengthening the steady state of the liquid, show acoustic similarities to Lehiste's 'real' productions of
syllabic liquids, like ample, able, other, over, etc. To the extent that there are similarities, then perception can be said to mirror production. That is, a motor theory of speech perception is viable. A strict motor theory (Studdert-Kennedy, et al. 1970) says that we perceive speech by reference to the way we produce it. So if my synthetic POLITE and PARADE also have acoustic characteristics like real syllabic liquids, there is weak evidence for a motor theory. I say 'weak' evidence because it is well known that similar acoustic representations can be obtained by different articulations, so that even if the (so-called) syllabic /l/ in POLITE is similar to the syllabic /l/ in able, it does not automatically follow that identical articulatory movements are responsible. To the extent that the synthetic syllabic liquids are not acoustically similar to the real speech productions, we have evidence against a motor theory, namely that it is possible to simulate syllabic liquids by lengthening their steady states, instead of whatever means are employed in production.

Syllabic /l/ in Lehiste's study has characteristics such as the following: (1) Syllabic /l/ has a lower F2 than final /l/. (2) F3 is somewhat lower than in final /l/. (3) The highest F2 values occur after velars, and the lowest F2 values after labials. (4) Syllabic /l/ has a relatively great intensity, and in many cases "... the unstressed
syllabic /l/ actually had a greater amplitude than a stressed high vowel occurring in the first syllable"). (p. 32)

Comparing the real speech syllabic /l/’s to the synthetic, it is obvious that they do not share most acoustic characteristics. As the results of the labeling tests showed, PLIGHT with a long /l/ is perceived as POLITE, so one can assume a syllabic /l/ has been perceived. But no parameter except duration of the steady state has been altered - F₂ and F₃ were not lowered nor was there any change in the amplitude curve. For /l/ then, perception does not mirror production.

Lehiste also compares syllabic /r/ in words like catered and Hungary with non-syllabic /r/ in hatred and hungry respectively. The syllabic /r/ typically has a lower F₂ and F₃ than the non-syllabic /r/ but the most striking acoustic difference lies in the approximately 50% greater duration of the syllabic allophone.

Looking at the key words with /r/ in this study, BROKE, DRESS, CREST and PRAYED, which were perceived as BAROQUE, DURESS, CARESSSED and PARADE when the steady state portion of the /r/ was lengthened, it appears that perception comes closer to mirroring production than with /l/. That is, the synthetic stimuli have a non-syllabic /r/ which simply gets longer, and likewise in Lehiste's corpus, the primary factor is a difference in the duration of the /r/ phoneme. Overall
then, it appears that the simulation of syllabicity by increased duration supports a motor theory of speech perception for /r/ but does not support it in its strongest version for /l/.

2. Speech Style Differences

The third hypothesis stated at the beginning of this chapter was that a precursor speech style phrase could shift listeners' perception from the monosyllabic to the disyllabic word (and vice versa). This hypothesis was not supported; there were no significant differences among the five groups. There were two expectations regarding the speech styles. First, I hypothesized that the control groups' responses would be most similar to those of the Formal-Slow group, based on the assumption that in the absence of instructions to the contrary, listeners would judge the stimuli as though they were in a maximally differentiated speech style. But the control groups' average crossovers were the closest to the Formal-Slows for only three of the six words: BLOW-BELOW, PLIGHT-POLITE and CREST-CARESSED. Furthermore, averaging over all the words, the control group was actually the farthest from the Formal-Slow, as Table 4.1 shows.
The second hypothesis was that I expected either the Formal-Slow or the Control crossovers to be the latest. Taking PLIGHT-POLITE as an example, I predicted that the Formal-Slow group would require that the /l/ in PLIGHT be considerably lengthened before the word would be acceptable as POLITE. On the other side, I expected the Casual-Fast group to show the earliest crossovers, since in casual, sloppy speech, PLIGHT with a fairly small /l/ lengthening should be acceptable as POLITE. Regarding the first, the Formal-Slow crossovers were the latest for only two words, PLIGHT-POLITE and PRAYED-PARADE. For the second, the Casual-Fast crossovers were the earliest for only one word, PLIGHT-POLITE. Regarding the mixed styles, Formal-Fast and Casual-Slow, I predicted the Formal-Fast to show the earlier crossovers since my previous research had shown that Formal-Fast words were of shorter duration than Casual-Slow, and that the Slow-Fast dimension was easier to both produce and perceive than the Formal-Casual. But the labeling results showed that only three words had earlier average crossovers for Formal-Fast versus Casual-Slow: BLOW-BELOW, PLIGHT-POLITE and DRESS-DURESS. Referring to Table 4.1 again, averaging over all words, the Formal-Slow crossovers

<table>
<thead>
<tr>
<th>F-S</th>
<th>F-F</th>
<th>C-S</th>
<th>C-F</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>69</td>
<td>65</td>
<td>66</td>
<td>65</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 4.1. Average Crossovers (in msec), 5 Speech Styles
average only 4-5 msec later than the other groups; both Formal-Fast and Casual-Fast show the same crossovers; and the Control and Casual-Slow groups average the same crossovers. Indeed, speech style cannot be said to have influenced the 75 subjects at all.

I think there are several reasons for the failure of the speech style hypothesis. First, it is possible that the precursor phrases were not good examples of the four speech styles. The speaker who recorded the phrase as much as admitted that he was not very happy with the Formal-Fast and Casual-Slow versions. Specifically, he said that the 'mixed' styles were harder to produce, considering that the phrase consisted of only one short utterance, instead of a long stretch of speech where one could more or less work up to the desired style and rate combination. Second, it is possible that the subjects simply stopped paying attention to the precursor at all after a while\(^1\). Considering that the exact same phrase preceded each stimulus, subjects could have tuned out the phrase very early in the session, once they realized that it was going to be repeated almost 200 times. One solution to this problem would have been to vary the precursor while keeping the same speech style. That is, several different phrases could have been used, randomly

\(^{1}\text{Peter Ladefoged suggested this reason to me, and made the suggestion that the precursor might have been varied, rotating among different phrases to retain the subject's attention.}\)
assigning each to a different stimulus. This technique should have at least retained the subjects' attention to the precursor. Third, the listeners in each group had no style and rate to compare the stimuli against. That is, each group listened to only one version of the precursor phrase, so that it was probably very hard to decide whether a word was Casual-Slow, for example, when the listener did not have a Casual-Fast precursor with which to compare it. The fourth reason has to do with an asymmetry inherent in the experimental task, presented graphically in Figure 4.1. Called a Speech Style Acceptance Table, the table is actually a hypothesis about how the labeling could have been done. Taking BLOW-BELLOW as an example, at the short durations of /l/ the Formal-Slow group would be likely to accept the stimuli as monosyllabic but reject them as disyllabic. But at the same short durations the other extreme group, Casual-Fast, would be likely to accept the stimuli as either monosyllabic or disyllabic. The reason I suggest is that in casual fast speech phonetic [biʌ] could be lexically either BLOW or BELOW. The short /l/ allows the word to be perceived as BELOW but it does not disallow it from being perceived as BLOW. Similarly, at long durations of /l/, the Casual-Fast group would be likely to reject the stimuli as BLOW and accept them as BELOW. But the Formal-Slow group could accept them as either BLOW or BELOW. Again for the latter group, the long /l/ allows the stimuli
<table>
<thead>
<tr>
<th></th>
<th>Formal-Slow Speech Mode</th>
<th>Casual-Fast Speech Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short (121)</td>
<td>Blow</td>
<td>Blow or Below</td>
</tr>
<tr>
<td>Long (191)</td>
<td>Blow or Below</td>
<td>Below</td>
</tr>
</tbody>
</table>

Figure 4.1 Speech Style Acceptance Table
to be perceived as BELOW but still does not disallow them from being perceived as the monosyllabic BLOW. Evidence in support of this 'Acceptance Hypothesis' would be found in the labeling results of the Formal Slow and Casual Fast groups. Percentages close to the 50% line would indicate more guessing on the subjects' part, so that tabulating a 'guessing average' for the curves on either side of the crossover should indicate whether the Acceptance Hypothesis is borne out. I made the assumption that the crossover duration perceptually divides the short liquid from the long one. Using BLOW-BELOW again, I added up the percentages for BELOW on the left side of the crossover point, and divided by the number of stimuli left of the crossover. The BELOW percentages before the crossover total 80. Divided by 6 (the number of stimuli preceding the crossover) the guessing average is 13.3. The same procedure is followed for the BLOW responses to the right of the crossover point. The total is 158 divided by 7 stimuli, giving 22.6 as the guessing average. The higher the number, the closer to the 50% chance level, and the more guessing was assumed to take place. Table 4.2 shows the results of this procedure.

For the Formal Slow group a higher guessing average for the long liquid duration constitutes support for the Acceptance Hypothesis, and for the Casual Fast group, a higher guessing average for the short liquid duration constitutes support for the Acceptance Hypothesis. From the
<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Style</th>
<th>Guessing</th>
<th>Average</th>
<th>Support for Acceptance Hypothesis?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Short Liquid</td>
<td>Long Liquid</td>
<td></td>
</tr>
<tr>
<td>BLOW-BELOW</td>
<td>FS</td>
<td>13.3</td>
<td>22.6</td>
<td>YES</td>
</tr>
<tr>
<td>BLOW-BELOW</td>
<td>CF</td>
<td>20.0</td>
<td>31.4</td>
<td>NO</td>
</tr>
<tr>
<td>PLIGHT-POLITE</td>
<td>FS</td>
<td>21.8</td>
<td>22.9</td>
<td>YES</td>
</tr>
<tr>
<td>PLIGHT-POLITE</td>
<td>CF</td>
<td>37.7</td>
<td>24.4</td>
<td>YES</td>
</tr>
<tr>
<td>BROKE-BAROQUE</td>
<td>FS</td>
<td>23.5</td>
<td>23.2</td>
<td>NO</td>
</tr>
<tr>
<td>BROKE-BAROQUE</td>
<td>CF</td>
<td>23.0</td>
<td>12.9</td>
<td>YES</td>
</tr>
<tr>
<td>DRESS-DURESS</td>
<td>FS</td>
<td>10.8</td>
<td>21.0</td>
<td>YES</td>
</tr>
<tr>
<td>DRESS-DURESS</td>
<td>CF</td>
<td>15.0</td>
<td>27.9</td>
<td>NO</td>
</tr>
<tr>
<td>CREST-CARESSED</td>
<td>FS</td>
<td>21.8</td>
<td>25.2</td>
<td>YES</td>
</tr>
<tr>
<td>CREST-CARESSED</td>
<td>CF</td>
<td>24.6</td>
<td>25.8</td>
<td>NO</td>
</tr>
<tr>
<td>PRAYED-PARADE</td>
<td>FS</td>
<td>15.8</td>
<td>15.4</td>
<td>NO</td>
</tr>
<tr>
<td>PRAYED-PARADE</td>
<td>CF</td>
<td>28.2</td>
<td>12.9</td>
<td>YES</td>
</tr>
</tbody>
</table>

Table 4.2. Speech Style Guessing Averages (in msec)
table, we see that the hypothesis is supported in seven cases, and not supported in five cases. In two cases, the guessing average for the short versus the long liquid differs by less than 1% — BROKE-BAROQUE, FS and PRAYED-PARADE, FS — not very strong evidence against the hypothesis. Overall, the Acceptance Hypothesis receives moderate support. I conclude that the potentially ambiguous responses for the Casual Fast group at short liquid durations and the Formal Slow group at long durations were factors in the failure of speech style differences to emerge, but not overwhelmingly.

It should also be kept in mind that the subjects might have been unable to do the task simply because it was too difficult to do, especially the mixed styles, Formal Fast and Casual Slow. One way of simplifying the task would be to have subjects listen to the tape twice and ask them to make a Fast-Slow decision about the stimuli in one session, and a Formal-Casual decision in another session.

3. The Categorical Perception of Syllabicity

In this Section, I focus on the results of the labeling and discrimination tests described in Chapter III. They are interpreted within the context of the theory of the categorical perception of speech. First I review some recent literature on the perception of liquids. Second, I interpret the results of the experiments mentioned above.
3.1. The Categorical Perception of Liquids

Early work on the perception of the liquids /r/ and /l/ was done by O'Connor et al. (1957). Using handpainted spectrograms to synthesize the segments, they found that /w, y, r, l/ could be distinguished by manipulating $F_2$ and $F_3$. A high $F_2$ was needed for /y/, a low $F_2$ for /w/ and intermediate values for /l/ and /r/. The starting frequencies of /r/ and /l/ overlapped and were dependent on the following vowel, as Table 4.3 shows.

<table>
<thead>
<tr>
<th>Context</th>
<th>/r/</th>
<th>/l/</th>
</tr>
</thead>
<tbody>
<tr>
<td>___i, e, ε</td>
<td>940-1560</td>
<td>960-1800</td>
</tr>
<tr>
<td>___o</td>
<td>840-1200</td>
<td>840-1680</td>
</tr>
<tr>
<td>___o, u</td>
<td>600-1200</td>
<td>840-1680</td>
</tr>
<tr>
<td>___a</td>
<td>not given</td>
<td>840-1800</td>
</tr>
</tbody>
</table>

Table 4.3. $F_2$ Onset Frequencies for Liquids (in Hz)
(after O'Connor, et al. 1957)

$F_2$ alone did not distinguish /r/ and /l/ but $F_3$ did. The authors say, "The third formant starting frequency for /l/ is close to that of the vowel third formant, whereas the third formant onset of /r/ needs to be lower in frequency, fairly close to the second formant onset. This being the case, it is possible to pass from /l/ to /r/ by no other change in the pattern than a gradual lowering of the
starting frequency of the third formant". (p. 34) It is the 
interrelation between $F_2$ and $F_3$ that is interesting. Before 
front vowels, only two formants were necessary to perceive 
/r/ and the $F_3$ range could extend from the /r/’s $F_2$, around 
840 Hz to as high as 1920 Hz; but before front vowels, /l/ 
could be no lower than $F_3$ of the vowel. Before back vowels, 
where $F_2$ is falling, the reverse holds: two formants are 
sufficient to perceive /l/, and the $F_3$ onset can be as low 
as 1920 Hz, but $F_3$ is needed for /r/ and can’t be any higher 
than 1680 Hz.

O’Connor et al. also found that the duration of the 
transition from the liquid to the following vowel affected 
identification. A 100 msec transition was suitable but 
60-70 msec produced better /l/’s and transitions longer than 
100 msec gave better /r/’s.

$F_1$ did not distinguish /w,y,l,r/ but it did distinguish 
/l/ from the nasals. In order to perceive /l/, the $F_1$ 
starting transition had to be at 360 Hz or higher; lower 
frequencies were identified as nasals. It was also the case 
that the $F_1$ transition of /l/ was important. "For /r/, 
there is more information in the steady state onset to the 
extent that the onset approximates an /r/-colored vowel." 
(p. 31)

More recently, an interesting experiment on the 
categorical perception of /r/ and /l/ by speakers of English 
and Japanese was done by Miyawaki, et al. (1975). While
English has two phonemes, /r/ and /l/, Japanese has only one, a more or less flapped variety of /r/. Varying only $F_3$ to produce the liquids and performing the traditional labeling and discrimination tests on groups of American and Japanese listeners, the authors found that the Americans perceived the stimuli in a completely categorical fashion: low $F_3$ values made the stimuli sound like /r/; high $F_3$'s gave /l/. But the Japanese listeners perceived them more continuously, demonstrating the conclusion of the authors that "...the effect of linguistic experience is specific to perception in the 'speech mode'." (p. 331)

The relationship between the perception of liquids and hemispheric specialization has been investigated by Cutting (1974). Although not under all conditions, it is generally found that consonants, with their considerable encodedness, are processed in the left hemisphere - the so-called 'language hemisphere' - and vowels are processed in the right hemisphere. (See Studdert-Kennedy 1976 for a thorough discussion of this topic.) A good question is: where are liquids processed? Distinctive feature systems emphasize the vocalic and consonantal properties of /r/ and /l/, for example, Jakobson, Fant and Halle (1956) and Chomsky and Halle (1968). Liquids have steady state formants like vowels but swiftly changing transitions like consonants. Cutting used the technique of dichotic listening to try to answer that question. In dichotic listening experiments,
different stimuli are presented at the same time to the right and left ears. Because of the contralateral pathways from the auditory nerves to the hemispheres, signals to the right ear go directly to the left hemisphere, and signals to the left ear go to the right hemisphere. If a speech signal is typically processed in the left hemisphere and if it is fed to the left ear, it will take longer to reach the left hemisphere than if it were fed to the right ear. This is because the signal must be transferred by means of inter-hemispheric pathways from the right hemisphere to the left. A signal in the right ear, going directly to the left hemisphere, results in a 'right ear advantage' because the listener's response to that signal will be faster than to a signal to the left ear.

In Cutting's dichotic listening experiments using stop-liquid-vowel syllables, he found a right ear advantage for stops, a left ear advantage for vowels and a moderate right ear advantage for liquids (less than for stops). He concluded that liquids are perceptually (and acoustically) more like consonants than vowels.

But further dichotic listening tests for stops and liquids in final as well as initial position showed the following: a right ear advantage for stops in both positions and /l/ in initial position but no ear advantage for final /l/. (Subjects had 63% correct identification of /l/ for the right ear and 65% correct for the left ear, a net 2%
left ear advantage.) Cutting concludes that "...initial liquids yield results which are typical of consonants, while final liquids yield results which are typical of vowels." (p. 606) He offers further discussion of the perceptual differences among the liquids (although this particular experiment did not include final /r/). Cutting suggests that perceptual differences in liquids stem from acoustic cues. Even though initial and final /r, l/ are mirror-images of each other, the cues for initial /r, l/ are found in the transitions (like stops) while the cues for final /r, l/ are found in the steady state of glides (like vowels). As a result,

"If the perceptual system takes advantage of where these cues are located in the sound pattern of speech, the differential ear advantages for initial and final liquids are explicable in acoustic terms not phonetic terms. That is, some contribution to the right ear advantage may stem from the acoustic form of the signal not from the phonetic labeling of it." (p. 606)

Cutting does not consider the 2% left ear advantage for final /l/ enough to be considered a significant left ear advantage. It would be interesting to know whether final /r/ produces a real left ear advantage. Referring back to the O'Connor et al. (1957) study of the steady state of liquids, recall that steady state /l/ was not as good a stimulus as steady state /r/, so that one might expect what Cutting says about final /l/ to be even more true for final /r/.
Research directly related to the question of /r/ and /l/ perception and syllable position has been done by McGovern and Strange (1976). Using the liquids and the vowel /i/, they constructed mirror-image stimuli, /ri/, /li/, /ir/ and /il/. They manipulated only $F_3$ and ran both identification and discrimination tests. The lack of right ear advantage found by Cutting (1974) for final /l/ would suggest that final liquids are not perceived categorically. But McGovern and Strange found that both liquids were perceived categorically in both syllable positions. The authors' only comment about their findings versus Cutting's is, "These results leave us with one difficulty: the dichotic listening and categorical perception methodologies lead to different conclusions regarding the perceptual status of the final liquids." (p. 5) What might also be suggested is that the right hemisphere might also work somewhat categorically in speech perception.

There is one aspect of the McGovern and Strange study that raises a question about the results. The authors used the vowel /i/ in their test syllables. If they are referring to the high front vowel as in *beet*, then they have produced syllables that don't occur in English in the case of the final liquids, /ir/ and /il/. /ri/ and /li/ (the front or 'light' /l/ I assume) are English sounding syllables but the liquid-final syllables, said to be mirror-images, are not, since the maximally high front vowel
does not occur before /r/ or before the front /l/. Syllable-final /l/ is almost always velarized in English and when it does occur after /i/, it usually becomes syllabic as in feel [fiɛl] or [fiyɛl]. Also, when /i/ occurs before /r/, the /r/ becomes syllabic as in seer [sɪɛr] or [sɪyɛr]. Mirror-images of syllable-initial liquid+/i/ would simply not produce stimuli of the same 'believability'. (I should not discount the possibility that the issue is 'typographical' - it is possible that in the McGovern and Strange paper, /i/ was typed but /r/ was intended. If so, these remarks are only partially valid. /r/ occurs in English, but /r1/, where /l/ is not velarized, still does not.) If the question about /i/ is legitimate, then it is possible that the liquids in final position were perceived more like true consonants (and more like syllable-initial liquids) and thus more categorically. If the final liquids had been more like real English syllables, the perception test results may have been more continuous than categorical and not have shown the disparity with Cutting's results.

3.2. Interpretation of Labeling and Discrimination Tests

At first glance, the results of the labeling tests and the discrimination tests appear very much at odds. Specifically, the perception of syllabicity appears to be categorical when the labeling tests are considered since fairly clear crossovers are seen for all of the /r/-/l/
words. But perception is definitely not categorical when the discrimination results are considered because there is no peak in discrimination that coincides with the 'syllabicity boundary' for the PRAYED-PARADE stimuli. (Since the PRAYED-PARADE stimuli seemed typical in the sense that they were good synthetic tokens and their labeling crossover did not show very wide fluctuation among the subjects, there is no reason to think the discrimination results would have been different using one of the other word pairs.) To recapitulate the discrimination results, subjects were unable to discriminate the one-step stimuli at all, the curves all hovering around the 50% mark; subjects were still unable to discriminate the two-step stimuli, subject 3 being the only one to perform above the 50% level; but in the three-step test, only subject 2 remained at chance level, the others (particularly subject 3) being able to discriminate among the stimuli.

One problem with interpreting the results of the labeling and discrimination tests is that, as mentioned in Chapter III, different subjects were used in the two experiments. This means that when the labeling results are compared with the discrimination results, considerable latitude is being taken in making generalizations. To compensate for this, I had the three discrimination subjects take the labeling test as well. (Since I was convinced that no speech style differences would emerge, they listened to
the Control tape, with no precursor phrase.) The results are
not very different from those of the original 75 subjects.
Crossovers were obtained for each of the 6 /r/-/l/ words,
and no crossover was obtained for SPORT-SUPPORT. The
average crossover durations are presented in Table 4.4.

<table>
<thead>
<tr>
<th>WORD PAIR</th>
<th>AVERAGE CROSSOVER DURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>BROKE-BAROQUE</td>
<td>65</td>
</tr>
<tr>
<td>DRESS-DURESS</td>
<td>65</td>
</tr>
<tr>
<td>PLIGHT-POLITE</td>
<td>70</td>
</tr>
<tr>
<td>PRAYED-PARADE</td>
<td>70</td>
</tr>
<tr>
<td>CREST-CARESSED</td>
<td>75</td>
</tr>
<tr>
<td>BLOW-BELOW</td>
<td>93</td>
</tr>
</tbody>
</table>

Table 4.4. Average Crossover Durations (in msec 3 Subjects)

Like the original results, BLOW-BELOW has the latest
crossover and CREST-CARESSED the second latest. But
PLIGHT-POLITE now has an intermediate crossover, 70 msec, as
opposed to its original very early crossover, 49 msec. The
others are all within 10 msec of their original crossover
durations. Except for PLIGHT-POLITE, the differences are
not striking.

Since data were now available from the same subjects
for both experiments, it was possible to use the labeling
results to predict the discrimination functions. Figures
4.2-4.4 show the original ABA discrimination curves, but
with their predicted functions added. (The formula for
predicting discrimination curves from identification scores
is taken from Liberman, et al. 1957.) It is usually found
Figure 4.2 ABA One-Step Discrimination, Predicted and Obtained Values, All Subjects
Figure 4.3 ABA Two-Step Discrimination, Predicted and Obtained Values, All Subjects
Figure 4.4  ABA Three-Step Discrimination, Predicted and Obtained Values, All Subjects
that the obtained scores are higher than the predicted, and this was the case here for the two- and three-step tests, but not the one-step where the obtained scores were lower than the predicted.

The questions to be answered are: (1) Although both sets of labeling results show crossovers (original subjects plus discrimination subjects), are they sharp enough to be considered categorical? (2) If they are categorical, how can we reconcile the disparity between those results and the clearly non-categorical results of the discrimination tests?

The first question is actually very hard to answer. There are no 'guidelines' that I am aware of for determining how steep identification curves have to be before the responses can be unambiguously called categorical. Simply the presence of a crossover point does not mean that subjects were secure in categorizing the stimuli. In the case of the /r/-/l/ stimuli, the steady state range was quite large, 120 msec between the shortest and longest liquid durations and there was no doubt that the short steady states would produce monosyllabic judgments and the long steady states, disyllabic judgments. That crossovers occurred then, is no surprise. The question is, is the crossover sharp or is it fuzzy? The data leave little doubt that the crossover is fuzzy. In almost every case, there seems to be roughly a 40 msec time period around the crossover where subjects were mostly guessing. In the
BLOW-BELOW data, the curves even cross over each other several times, a result which happened only once, although in several instances other curves come close to re-crossing. In general, a 40 msec syllabicidity boundary does not indicate very categorical perception. One way to get the results to look more categorical would be to increase the durational increments between the steps of the stimuli. As mentioned in Chapter III, the difference limen for duration is between 10-40 msec (Lehiste, 1970), so that my stimuli, varying in 10 msec steps, did lie at the most stringent end of the scale. Changing the increment to 20 or 30 msec probably would have helped slightly, but a 40 msec step interval would have been too large. This is because we know from the results of the three-step discrimination test, that subjects can discriminate among stimuli that differ by 30 msec, indicating that this much of a difference is clearly perceptible. I must conclude that the labeling results are ambiguous: perception appears to be moderately categorical, if such a phrase is not a contradiction.

The next question to be answered is how to reconcile the clearly non-categorical discrimination results with the partially categorical identification results. First, I will state that I do not think that this experiment, taken as a whole, showed that the perception of syllabicidity is categorical. This is based on the fuzzy results of the identification test coupled with the continuous-perception
results of the discrimination test. Essentially, the question refers to the difference between segment perception and suprasegmental perception. The acoustic variable hypothesized to be responsible for the perception of syllabic \(/r/\) and \(/l/\) was simply the duration of the steady state of the liquid, and duration is traditionally referred to as a suprasegmental phonetic feature. Along with pitch and intensity, duration is an acoustic event that can span several phonemic segments (although it need not); sometimes the suprasegmentals are considered as overlaid features. (See the Introduction to Lehiste 1970 for a lucid discussion of what suprasegmentals are.) Suprasegmental features in general have more in common with vowels than with consonants, particularly stops. Pitch and intensity are almost always defined over vowels, although duration can be defined over both vowels and consonants. Regarding categorical perception, the basic difference between vowels and stop consonants is that the former are not perceived categorically while the latter are. (Vowels can be forced into categorical perception if they are short enough. Cf. Studdert-Kennedy 1976 for a review of the literature on vowel perception.) What the discrimination tests in this study showed is that the perception of syllabicity, using steady state duration as the acoustic variable, is similar to vowel perception. Both are perceived in a more continuous fashion than consonants.
It appears that suprasegmentals are perceived like vowels. This statement should be regarded as precautionary though because of at least two factors. First, only duration was tested. The other prosodic features, pitch and intensity, were not, so that reaching a conclusion about all suprasegmental features is not warranted. Second, the results of this study are completely counter to another study on the perception of duration in Japanese. (Fujisaki, Nakamura and Imoto 1973.) Vowel and consonant duration is phonemic in Japanese. Fujisaki, et al., performed labeling and AX discrimination tests for vowels and consonants ([oi] 'nephew' versus [o:i] 'cover' and [ise] 'a place name' versus [isse] 'a unit of area') and came to the following conclusion: "The peaks of the two discrimination curves roughly correspond to locations of phoneme boundaries...and clearly indicate that the discrimination performance is influenced by categorical judgments to almost the same extent both in vowels and in fricative consonants, when the stimuli are varied on the continuum of duration." (p. 290) In a language that has phonemic duration, duration is perceived categorically.

The only way to redeem the claim, based on my study, that duration is perceived continuously, is to argue that the facts of Japanese are different in kind from English. Two factors bear on this: (1) It is the very essence of the categorical perception model that the listener's linguistic
experience determines his responses to the identification and discrimination tests. The fact that duration is phonemic in Japanese but not in English is the reason that the discrimination tests produced such different results in the two languages. I have said that duration is not phonemic in English but my labeling tests did show that listeners use duration in assigning stimuli to one or another linguistic category. Perhaps we should re-address the question: Is duration phonemic in English? The issue is discussed further in Section 5. (2) The second way in which the English and Japanese conclusions might not be comparable is that the perception of syllabic ity might not be the same as the perception of phonemic segmental duration. This also is developed further in Section 5; suffice it to say here that if suprasegmentals are secondary, overlaid functions of inherent features (according to Lehiste 1970), then syllabic ity might be considered an overlaid function of suprasegmentals. Lehiste says, "Every segment must have a certain duration in the time domain to be identifiable as a segment; the feature of quantity involves manipulation of inherent duration." (p. 2) I would add another level to this hierarchy: that syllabic ity involves the manipulation of quantity for linguistic purposes in languages that do not have phonemic quantity.

Leaving the issue of whether suprasegmentals are perceived more continuously than non-vowel segments, the
discrepancy between the labeling and discrimination tests can be looked at another way. The labeling tests seemed to show that syllabic ity is categorized but there may be an inherent bias in the test procedure which obscures the real trends. The labeling test was a binary forced choice: there was no provision for a listener to indicate, a) that a stimulus did not fit either category, or b) whether a stimulus was a relatively good or relatively poor token of the category to which it was assigned. For example, in the PRAYED-PARADE set, a steady state /r/ of 70 msec was perceived as PARADE but sc was a steady state /r/ of 150 msec. No confidence ratings were elicited from subjects; there was no way to indicate that the 70 msec /r/ was slightly more like PARADE than PRAYED but that the 150 msec /r/ was clearly and unambiguously PARADE. In fact, it is entirely possible that none of the stimuli was a totally believable token of its respective word. There is some support for this notion in the fact that out of the total of 91 different stimuli, 100% agreement on identification by any one group of subjects was reached for only 10 of the stimuli. When the five speech style groups are combined and averaged, not a single stimulus shows 100% agreement among the 75 listeners. What this suggests is that the labeling test is a somewhat false barometer of categorical perception. This is not to totally discredit the labeling paradigm, but in all the research of which I am aware, it is
the discrimination test, not the labeling test, which provides the most compelling evidence for categorical perception. I know of no examples of speech perception research where the results of the discrimination test were categorical while the labeling responses were non-categorical. On the other hand it is not uncommon to find labeling results looking categorical and discrimination results continuous (the present research notwithstanding).

Given the mildly categorical-looking labeling results in this study, another way of doing the labeling test might have produced either non-categorical responses or at least shifted the syllabic boundary considerably to the right. Instead of asking listeners to choose between two lexical items, I might have asked them to decide whether each stimulus was monosyllabic or disyllabic. This technique would partially cancel out the lexical effect of the PRAYED-PARADE forced choice. Instead, the answer sheet would not have any 'words' for subjects to check off, only a column for 'monosyllabic' and 'disyllabic'. My hypothesis is that at the least, PRAYED would have to have a fairly long steady state /r/ before it would be perceived as disyllabic and at the most, it never would be perceived as disyllabic, resulting in continuous perception. In the latter case, then the labeling test would be in agreement with the present discrimination results. The hypothesis remains to be tested though; it should make an interesting
4. Crossover Differences Among the Key Words.

The original labeling test resulted in average crossover durations for the steady state liquids as shown in Table 4.5.

<table>
<thead>
<tr>
<th>WORD PAIR</th>
<th>AVERAGE CROSSOVER DURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLIGHT-POLITE</td>
<td>49</td>
</tr>
<tr>
<td>BROKE-BAROQUE</td>
<td>61</td>
</tr>
<tr>
<td>PRAYED-PARADE</td>
<td>63</td>
</tr>
<tr>
<td>DRESS-DURESS</td>
<td>68</td>
</tr>
<tr>
<td>CREST-CAressed</td>
<td>70</td>
</tr>
<tr>
<td>BLOW-BELOW</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 4.5. Average Crossover Durations (in msec) for the 6 Key Words, All Groups Combined

The average crossover durations for the three discrimination subjects are not included since there were only six responses per stimulus, as opposed to 150 responses for the original speech style groups. Two questions need to be answered: (1) Does this variability represent statistically significant differences among the six words?, (2) If so, how can the variability be accounted for? In this Section, I show that the answer to the first question is yes, and then I suggest a number of approaches to explaining the variability. Six possible explanations are discussed and then rejected. Finally I propose an explanation for the differences, which combines inherent differences among segment types plus some characteristic differences between
That the variability shown in Table 4.5 was statistically significant can be seen in Table 4.6.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>ms</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>1.83</td>
<td>5</td>
<td>.37</td>
<td>16.79</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Error</td>
<td>8.08</td>
<td>370</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>9.91</td>
<td>375</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6. Analysis of Variance Among the 6 Key Words

An analysis of variance (treatments by subjects or repeated measures design) among the six key words was computed. All five speech style groups are combined, and treatment refers to the six different key words. The F-ratio is clearly significant. What Table 4.6 does not tell us is which words differed from which other words. To determine this, another statistic, Duncan's Multiple Range Test for nearly equal n's, was used. (Bruning and Kintz, 1968). As expected, BLOW-BELOW with its late average crossover was significantly different from all the other words. In addition, CREST-CARESSED with the second latest crossover was significantly different from PLIGHT-POLITE, BROKE-BAROQUE and PRAYED-PARADE. Finally, DRESS-DURESS was significantly different from PLIGHT-POLITE. These differences are shown
graphically in Figure 4.5.

Figure 4.5. Differences Among the Key Words After Duncan's Multiple Range Test

It is clear that those words with the latest crossovers differed statistically from those with earlier crossovers.

Now we must ask how to account for these differences.

4.1. Voicing of Preceding Stop.

One approach to explaining the differences in crossover durations for the steady state portion of the liquids would be to consider the voicing of the preceding stop.

<table>
<thead>
<tr>
<th>KEY WORD</th>
<th>CROSSOVER DURATION IN MSEC</th>
<th>INITIAL VOICED STOP</th>
<th>INITIAL VOICELESS STOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLIGHT-POLITE</td>
<td>49</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>BROKE-BAROQUE</td>
<td>61</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>PRAYED-PARADE</td>
<td>63</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DRESS-DURESS</td>
<td>68</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>CREST-CARESSED</td>
<td>70</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>BLOW-BELOW</td>
<td>86</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 4.6. Voicing of Preceding Stops and Crossover Durations

But Figure 4.6 shows that there is no obvious pattern. If one were tempted to argue that a non-syllabic liquid following a voiced stop must be long in duration before it is perceived as syllabic, there would be no support for that
hypothesis.

4.2. Place of Articulation of Preceding Stop

Another place to look for reasons for the different average crossovers might be the place of articulation of the preceding stop.

<table>
<thead>
<tr>
<th>KEY WORD</th>
<th>CROSSOVER DURATION</th>
<th>PLACE OF ARTICULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLIGHT-POLITE</td>
<td>49</td>
<td>X</td>
</tr>
<tr>
<td>BROKE-BAROQUE</td>
<td>61</td>
<td>X</td>
</tr>
<tr>
<td>PRAYED-PARADE</td>
<td>63</td>
<td>X</td>
</tr>
<tr>
<td>DRESS-DURESS</td>
<td>68</td>
<td>X</td>
</tr>
<tr>
<td>CREST-CARESSED</td>
<td>70</td>
<td>X</td>
</tr>
<tr>
<td>BLOW-BELOW</td>
<td>86</td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 4.7. Place of Articulation of Preceding Stop and Crossover Duration

Figure 4.7 shows that the words with the earliest average crossovers start with bilabial stops. But the word with the very latest crossover also begins with a bilabial. Some research (Lehiste and Peterson, 1960) has shown a relationship between place of articulation and the amount of aspiration after voiceless stops (as place of articulation moves back in the mouth, aspiration increases). But voiced stops would not be included.

4.3. Inherent Differences Between /r/ and /l/

It is also possible to consider the intrinsic characteristics of /r/ and /l/ in explaining the different crossover durations. The production and perception of the
liquids have been discussed above but it suffices to say that the differences in /r/ and /l/ are not very revealing at this point since the word with the earliest crossover has an /l/ and so does the one with the latest crossover, PLIGHT-POLITE and BLOW-BELOW, respectively.

4.4. Front Versus Back /l/

While one could not say that /r/ must be longer than /l/ (or vice versa) in order to be perceived as syllabic, it is possible to look at the difference between the front or clear /l/ and the back or velarized /l/. Velarization is a coarticulatory effect associated with back consonants and vowels and lip rounding. Its acoustic manifestation is a lowered $F_2$ due to the narrowed vocal tract. Looking at BLOW and PLIGHT, both /l/’s precede back vowels but in BLOW, the following vowel is rounded as well. In my synthetic stimuli, the $F_2$ of /l/ is somewhat lower before /oU/ in BLOW, 951 Hz, than before /ay/ in PLIGHT, 1029 Hz. The difference is only 80 Hz though. In addition to coarticulation due to vowel context, in English, /l/’s are velarized in syllable-final position. Again we are faced with the question of where the syllable boundaries lie in BELOW and POLITE. If the boundary lies within or after the /l/ and if listeners expect syllable-final /l/’s to be velarized, then any /l/ which is not velarized (i.e., does not have a lowered $F_2$) might have to be considerably longer
than a non-syllabic /l/ before it is perceived as syllabic. In other words, perception might be based on a 'compensatory lengthening' principle - to compensate for the non-velarized nature of the non-syllabic /l/, the /l/ would have to be particularly long before it is perceived as syllabic. The trouble is, this would explain why the crossover in BLOW-BELOW is quite late, but it does not account at all for the crossover in PLIGHT-POLITE, being the earliest of all the words.

4.5. Amplitude Differences

Another possible factor in explaining the crossover differences is amplitude. In preparing the synthetic stimuli, as described in Chapter III, a single frame in each key word was increased in duration and no other acoustic parameter was manipulated. But it is possible that as the overall waveform changed, that small - though perceptible - changes in amplitude occurred also. If amplitude differences were present in one or another of the key words, one might account for crossover differences as related to them. For example, an amplitude difference in BLOW at a long but not a short /l/ duration may have influenced the syllabicity boundary. To determine whether there were any amplitude changes as word duration increased, amplitude

\[\text{---}\]

\(^1\)Robert J. Hanson suggested this possibility to me.
displays of the stimuli were made\(^1\). Figures 4.8-4.11 show some of these displays. Figure 4.8 gives the amplitude display of the shortest BLOW stimulus and Figure 4.9 shows the longest BLOW. Figures 4.10 and 4.11 are the shortest and longest CREST stimuli, respectively. The amplitude corresponding to the liquid portion lies within the rectangle. The difference in amplitude of the transitions from the stop to the liquid and to the following vowel were so small, on the order of 2-3 dB, that amplitude was ruled out as affecting the perception of syllabic ity in these stimuli.

4.6. Semantic Differences\(^2\)

Finally, semantic differences among the key words may have accounted for their crossover variability, at least in part. I doubt that this was the case at all, however, because after each group of subjects took the labeling test, I asked if any of the words had been unfamiliar or in any way problematic and not one of the 75 subjects mentioned a single word. But if their perception was affected, it does not necessarily mean they were consciously aware of it, so that some of the word differences should be mentioned anyway. First, BELOW is the only preposition, all the other

\(^1\)Amplitude displays were made at the Speech Communications Research Laboratory, Inc.

\(^2\)David J. Broad suggested this possibility to me.
Figure 4.8 Amplitude of BLOW-BELOW, Short /1/
Figure 4.9 Amplitude of BLOW-BELLOW, Long /1/
Figure 4.10 Amplitude of CREST-CARNSLD, Short /r/
Figure 4.11 Amplitude of CREST-CARESSED, Long /r/
words being nouns, verbs and adjectives. It seems intuitively unlikely that perceptual judgments were influenced by this fact. Second, of all the words, DURESS, BAROQUE and PLIGHT may be said to be more 'learned' than the others. If unfamiliarity did influence responses, I predict it would be in the direction of a later crossover. That is, if the disyllabic member of a pair is a less common word, the liquid should have to be considerably long before the stimulus is perceived as disyllabic. Unfortunately, this theory gains little support from the data. Both DURESS and BAROQUE showed intermediate crossovers, 68 msec and 61 msec, respectively. If the words were so uncommon, later crossovers should have occurred. Only PLIGHT is favored by the theory, since if PLIGHT is considered a fairly uncommon word, then even small /l/ increases would be sufficient to shift the listeners' perception to the more common disyllabic word, POLITE. This is exactly what happened, the crossover occurring at 49 msec.

4.7. The Interaction of Aspiration, Place of Articulation, and Differences Between /r/ and /l/ Perception

While none of the above six arguments fully explains the different crossovers among the key words, I think certain aspects of several of them are significant.

Concerning the relation between voicing and the perception of syllabic identity, consider the following
explanation\(^1\). After a voiceless stop, one expects a voiceless resonant. In fact, the spectrograms of PLIGHT, PRAYED and CREST from the production experiment described in Chapter II all show that except for a short period of voicing toward the end of the liquid, most of it is coterminous with the voiceless aspiration period following the release of the stop. Since listeners expect the liquids in PLIGHT, PRAYED and CREST to be voiceless and associate them with being non-syllabic, a fairly small amount of voicing will shift the listeners' perception to the disyllabic word. This would explain why the /l/ in PLIGHT-POLITE shows such an early crossover, and perhaps accounts for the intermediate crossover of PRAYED-PARADE. But CREST-CARESSED, having the second latest crossover of 70 msec, would not fit the pattern.

Although place of articulation alone did not appear to elucidate the problem, taken in conjunction with aspiration on voiceless stops, some further insights may be gained. Lehiste and Peterson (1960) found that on the average, English voiceless bilabial stops were followed by 58 msec of aspiration, alveolars by 69 msec and velars by 75 msec\(^2\). If we can make the assumption that a longer period of

\(^1\)I owe the substance of this explanation to Ilse Lehiste.

\(^2\)This finding is qualified somewhat. The authors also found that of the two main allophones of /k/, the front allophone, [c] and the back allophone, [k], the former averaged 78 msec of aspiration and the latter 72 msec. Alveolars always had a shorter aspiration period than velars though.
aspiration will also require a longer period of voicing of the liquid in order to be perceived as syllabic, then CREST-CARESSED, requiring a later crossover than the other voiceless stop pairs, is explained as well. On the other hand, having just argued that after voiceless stops, any amount of voicing of the liquid should signal a syllabic liquid, it is not clear why a liquid after a voiceless velar stop should act differently from one after a labial or alveolar stop. I have no solution to this paradox, except to hypothesize that it is the ratio of voiceless to voiced liquid that governs this perceptual phenomenon, not the absolute amount of voicing. That is, short aspiration needs only a short amount of voicing to make the liquid syllabic, the durational ratios, whatever they are, remaining constant.

We are now left with the voiced stop pairs, BROKE-BAROQUE, DRESS-DURESS and BLOW-BELOW. I am not familiar with any experimental evidence to the effect that place of articulation of voiced stops influences the duration of a following resonant (or vowel). Likewise, Lehiste (1970) says, "The influence of the manner of articulation of a preceding consonant on a following vowel seems to be largely unexplored." (p. 27) If such evidence existed, one might expect the alveolar pair, DRESS-DURESS, to show a later crossover than the bilabial pairs, BROKE-BAROQUE and BLOW-BELOW. DRESS-DURESS has an
intermediate crossover, 68 msec, and BROKE-BAROQUE, somewhat earlier, 61 msec. But BLOW-BELOW does not fit the pattern (if it is a pattern) at all, having the latest crossover, 86 msec.

It appears that BLOW-BELOW is the most anomalous pair. I think one has to look at not only the presence or absence of aspiration and the place of articulation of the preceding stop, but at the acoustic-perceptual characteristics of /r/ and /l/.

Earlier in this chapter some of the acoustic characteristics of liquids were discussed. In perceptual terms it was found that F₃ alone was enough of an acoustic cue to distinguish /r/ and /l/ (keeping F₁ and F₂ constant). The measurement of these sounds in real production showed that there are other acoustic differences, especially involving syllabic /l/. To briefly recapitulate the findings of Lehiste (1964), syllabic /r/ can be considered

---

¹Victoria Fromkin has suggested another reason for BLOW-BELOW having the latest crossover: it is the only stimulus that ends with a long, open vowel. All the other word pairs end in single consonants or clusters, and so the stressed vowel is always short. This fact in itself does not account for the late crossover duration of BLOW-BELOW; all it really does is point to one phonetic fact that separates this stimulus from all the others. But if, in production, BELOW differs from BLOW other than in the syllabic versus non-syllabic /l/, one would expect it to be in the duration of the stressed vowel. My production data show this to be the case, although the comparison is not strictly fair since the phonetic contexts for the words were not identical. This suggests that in order that BLOW plus a long /l/ be perceived as BELOW, the /l/ has to be considerably long to compensate for the fact that the stressed vowel is still at a duration characteristic of BLOW.
acoustically 'simpler' than syllabic /l/: "The only measurable difference between a syllabic /r/ and a nonsyllabic allophone of /r/ occurring in an analogous position may be the greater length of the syllabic allophone." (p. 190) But syllabic /l/ has both a lower F₂ and greater intensity than its nonsyllabic allophone.

With these differences in mind I would like to suggest that /r/-perception is facilitated by its relatively simple acoustic structure and that conversely, /l/-perception is hampered by its relatively complex structure. This claim receives additional support from the O'Connor et al. (1957) study. Speaking of the perception of their steady states, "...[/l/] contains some identifying information but naive listeners have considerable difficulty in perceiving the steady state as /l/... For /r/ there is more information in the steady-state onset, to the extent that the onset approximates an r-colored vowel." (p. 31)

What I am suggesting is that a word with a voiced stop + /l/ will require the /l/ to be quite long before it is acceptable as disyllabic. That this is so is not due solely to the /l/, but is due to the voiced stop also. Since there is no period of aspiration in which /l/’s spectral information would be present, there is no /l/ information at all until full voicing starts. But if BLOW-BELOW is now explained, then PLIGHT-POLITE is not. Actually this is not so, because the BLOW-BELOW 'theory' depends on there being
no aspiration present, which is of course not the case in PLIGHT-POLITE. There appears to be a hierarchy of cues for /l/-perception. At the top is presence or absence of aspiration. If aspiration is present, then only a small amount of voicing is necessary to shift the listeners' perceptions to the disyllabic word. If aspiration is not present, then a much longer voicing period is needed because of the complex acoustic cues associated with syllabic /l/. On the second rung of the hierarchy is place of articulation for voiceless stops — since the amount of aspiration increases as articulation moves back, so does the amount of liquid voicing needed to shift the listeners' perceptions. That this seems to hold for the two voiced pairs, BROKE-BAROQUE and DRESS-DURESS may be accidental.

5. Do We Perceive Segments or Syllables?

In the section on the labeling and discrimination tests in this chapter I argued that the perception of syllabicity was not categorical based on the results of the discrimination tests. The labeling tests, on the other hand, did produce the crossovers typical of the categorical perception model and I attempted to account for this discrepancy by claiming that, (a) the discrimination paradigm is a more potent test for categorical perception, and (b) the forced choice technique of identification tests has the inherent bias of disallowing 'neither category' as a
choice. Furthermore, listeners were not asked to give confidence ratings on their identifications.

In this section, I plan to retreat slightly from the strong conclusion summarized above. I am not retreating from the position that syllabicity is not perceived categorically; only from the implication that duration is never an acoustic cue to syllabicity. I think that extending the steady state duration of liquids in the contexts of the key words in this study does function as a cue to the number of syllables the word contains. In this section, I apply this finding to a general theory of speech perception and to phonological theory.

As a first approximation toward answering the question which is the title of this section, it might be rephrased as the following: Do listeners respond to the duration of a segment itself or to the overall waveform of the stimulus? Put this way, given the generally encoded nature of speech perception (Liberman, et al. 1967), and the fact that listeners did categorize the stimuli in the labeling experiment, it is reasonable to conclude that they do respond to the overall waveform of a stimulus and not to individual segments. The reason is that if subjects were simply processing /r/’s and /l/’s with unusually long steady states, there would have been no reason to think that the stimuli sounded like anything other than PLIGHT, BROKE, etc. with a rather long second segment. If that had been the
case, then at the very least, the crossover points would have been considerably later; at the most, there would have been no crossover at all. Translating back to the original question, it looks like we can say that the level of syllable perception is more basic than the level of phoneme (or segment) perception.

The question of which level is basic in speech perception is not answered though. Not only is there a wealth of (often conflicting) literature on the feature - allophone - phoneme - syllable level as the most basic in speech perception, but I contend that the title of this section itself could be misleading.

Before framing the question in a more appropriate way, I would like to summarize Studdert-Kennedy's (1976, p. 252) remarks about the segment versus syllable controversy. He says that the arguments against segment-level perception and for syllable-level perception are the following: (a) No phoneme-size segment can be isolated in the acoustic signal. (b) Some phonemes (stops) cannot be spoken in isolation. (c) We speak in syllables and syllables carry stress and overall rhythm. Furthermore, he cites research by Savin and Bever (1970) and Warren (1971), which showed that reaction time to syllables in a list was faster than reaction time to the initial phonemes of the syllables. Finally, he says that some researchers believe that the invariance and segmentation problems would be solved if the syllable were
an unanalyzed unit of perception. On the other side, there is also evidence that the segment is more basic than the syllable. Studdert-Kennedy cites work by Day and Wood (1972), Kozhevnikov and Chistovich (1965) and Savin and Bever (1970) which all showed that there are reaction time differences in the identification of vowels and consonants within a syllable. Such differences shouldn't occur if the syllable were a singular perception unit.

The unanalyzable-level controversy can continue indefinitely particularly as machinery for measurement becomes more sensitive and experimental design and methodology become more creative. Perhaps the only way out of this quagmire is to put the question in a different way. Again, based on the remarks of Studdert-Kennedy (1976), we should ask, what is the basic acoustic unit of speech perception and what is the basic linguistic unit of speech perception? Studdert-Kennedy phrases it this way: "There is wide agreement among writers whose views may otherwise diverge, that the basic acoustic unit of speech perception (and production) is of roughly syllabic length... This does not mean... that the syllable is the basic linguistic and perceptual unit." (p. 253) What Studdert-Kennedy implies, and what I would like to propose is that the syllable is basic from an articulatory and acoustic standpoint, but a unit like the segment (to the extent that segment and phoneme coincide) is basic from the standpoint of linguistic
function. Another way of putting the same idea is to say that phonemes, which function to distinguish meaning in language, are the smallest units that we normally hear under ordinary communicative language use, but syllables are what we normally produce, due to physiological (perhaps neurological) constraints on articulation.

The next question to ask is, considering that the suprasegmental feature, duration, can function to distinguish meaning among minimal pairs, what implications for phonological theory does this finding suggest? Since duration was the acoustic cue responsible for the crossovers in the identification experiment, one might ask whether duration is phonemic in English. If segmental duration makes a difference in meaning, then in most traditional phonological theories, duration should be phonemic. But first, I do not think duration in English can be considered phonemic even on traditional grounds; and second I do not think the right question is being asked.

To say that English has phonemic length would be a novel statement. In the case of vowels, English has predictable, and therefore not phonemic, vowel length. It is well known that English vowels are longer before voiced consonants than before voiceless consonants; they are also longer in open syllables than in closed syllables. Evidence for the entirely predictable nature of English vowel length comes from backwards speech (saying words backwards) and
slips of the tongue\textsuperscript{1}. In the case of backwards speech, a word like \textit{dock} /dak/ with a short vowel will become /k\textipa{\textexclam}d/ with a lengthened vowel in backwards speech, the vowel lengthening because it is now preceding a voiced stop. Some slips of the tongue came from Fromkin (1973): \textit{it's the red book} becomes \textit{red [bud]} and \textit{I'll wring her neck} becomes \textit{frink} [neg]. In both examples the vowel before the voiced segment, [bud] and [neg], is phonetically lengthened\textsuperscript{2}. Regarding consonantal length, if there were such a phoneme it would be very limited in its positions of occurrence, not only in the segments over which it occurs, but in its phonetic context\textsuperscript{3}. The distinction would be limited to the context stop-liquid-vowel precisely in the words that were used in this study. (English has syllabic nasals also but I don't know of any experimental evidence to

\textsuperscript{1}David Stamps has brought these types of evidence to my attention.

\textsuperscript{2}An example from Pig Latin shows that the lengthening rule is not always phonetically motivated because a word boundary can suppress it. In most dialects of Pig Latin, initial consonants or clusters are moved to the end of the word and a phonetic [ey] is added and in some dialects, words with initial vowels simply add [ey]. For those people there is a distinction between \textit{ice} and \textit{sigh}. \textit{Ice} is [\textipa{\textexclam}ysey] and \textit{sigh} is [\textipa{\textexclam}ysey]. Since the original word boundary follows the /s/ in \textit{ice}, the vowel is short. But in \textit{sigh}, the original boundary follows the vowel, so the vowel remains long even though phonetically it is followed by /s/ in the Pig Latinized version.

\textsuperscript{3}One is reminded of the classical arguments for and against /\textipa{\textexclam}l/ as a phoneme, as in Sapir (1925) who argued against /\textipa{\textexclam}l/ and Swadesh (1934) who called it a 'defective phoneme'.

the effect that simply extending the duration of the nasal's steady state can be responsible for its perceived syllabicity.) There are few, if any, English phonemes that have such a restricted position of occurrence. (Even the contexts for /h/, /ŋ/ and /ʒ/ can be stated more generally than this 'length phoneme'.)

In point of fact, I believe we are asking the wrong question when we ask whether length is phonemic in English. It is not the duration of the liquid itself which is distinctive, but rather its function, which is that the liquid is syllabic. The real question is - is syllabicity phonemic in English? In a strictly functional sense, it seems that syllabicity is phonemic since it distinguishes between otherwise identical words, fulfilling the minimal pair requirement for phonemes. If syllabicity were in fact phonemic, then using PLIGHT-POLITE as an example, the underlying representation of these words would be the same, except that PLIGHT would have /l/ and POLITE would have /L/.

But there are several reasons why [l] (and [ɾ], [ŋ] and [m]) should not be considered English phonemes. Two fairly

---

That underlying /l/ would violate the biuniqueness criterion for underlying representations (Chomsky, 1964) would be a problem for those who accept biuniqueness. Presumably, in words like feel [fiː], the [l] does not come from underlying /l/. It comes from /l/ plus a rule that /l/ becomes syllabic in final position after a front tense vowel. That means it would not be clear, given a phonetic [l] whether that segment came from an underlying /l/ or arose by phonological rule, thus violating biuniqueness.
weak reasons are that, (a) the number of phonemes would be increased by four, and (b) we would not be able to explain the spelling of words like POLITE, DURESS, CARESS, BELOW, BAROQUE and PARADE. If the liquid were syllabic in the underlying representation, why are there so many different vowels which can precede it, e.g. o, u, e, or a? (This argument could be countered by saying that there are historical reasons for the spelling differences and that the liquids were not originally syllabic. Because the original vowel + liquid was phonetically realized as a syllabic liquid and because there were no morphophonemic alternations, the phonological system was restructured. That is, the vowel + liquid in this context was restructured to syllabic liquid, yet the spelling remained unchanged.) The third, most potent, reason for denying phonemic status to syllabic liquids and nasals is that they are all derivable from either a final liquid which is preceded by a tense vowel, as in feel, fire1, or from [ə] + resonant. (Resonant /R/ here refers to the class of liquids and nasals. I am assuming that [ə] is not a phoneme in English but derives from a full vowel by vowel reduction in unstressed syllables, although that assumption is not strictly relevant.) In Traeger and Bloch's (1941) study of

---

1Davis Stampe has pointed out that the syllabification of liquids does not always occur after a tense vowel, but rather it occurs in final position after a nonsyllabic. That is, in some dialects feel is [fi:k] but file is [faɪ].
the syllabic phonemes of English, the authors claim that the syllabic resonants all come from /əR/. (It is of no consequence for my argument that for Trager and Bloch, /ə/ is a phoneme.) They argue,

"The phonetic similarity of the nucleus of such syllables [as in apple, button, rhythm, H.S.] is greatest to some allophones of the already established phoneme /ɔ/ and this lateral- or nasal-colored syllability is in complementary distribution with the members of that phoneme. We conclude that [ʌ,ʊ,œ] are /əl,ən,əm/. The phoneme /ɔ/ has then a wide range of allophones: [ʌ] in undone, [œ] in sofa, syllability in apple, button rhythm." (p. 232)

In the next paragraph the authors claim that the /ɔ/ in certain, father is also from /ər/ and thus acts like the resonants mentioned above. The interesting part of Trager and Bloch's argument for my purposes is that syllability is considered as an allophone, not a phoneme. Another reason for deriving syllabic resonants from /VR/ is that morphophonemic alternations, like able - ability and polar - polarity, can be derived if we say that the phonetic vowels in the second syllables in ability and polarity are the underlying vowels which reduce to [ə] and then to syllabic resonants when they are in unstressed syllables. Furthermore, citing evidence again from backwards speech, if words like butter and bottle had underlying syllabic /r/ and /l/, then we could not explain why, when these words are produced backwards, a vowel is pronounced after the resonant
(albeit reduced, depending on how one stresses backwards speech). Butter backwards is /rɛt.əb/, not */ɛt.əb/ and bottle is /lətab/, not */ɛl.əb/.

To summarize, the results of the labeling tests in this study, indicating that duration can be a perceptual cue to syllabic itinerary, raise the possibility that syllabic itinerary (although not duration per se) is phonemic in English. If that were the case, pairs like PLIGHT-POLITE would have underlying representations like /pλaɪt/-/pλaɪt/. Yet there are good arguments against positing underlying syllabic resonants for English. This leads us to say that the underlying representations for PLIGHT-POLITE are /pλaɪt/-/poləɪt/. (possibly /pɛləɪt/). Now there is no reason to say that syllabic itinerary is phonemic because there is a real vowel in the underlying representation of POLITE. Another way of capsulizing the issue is to say that looking solely at the acoustic-phonetic data provides one interpretation of the facts, but considering these data in the context of phonological theory provides another interpretation.

To treat phonetic facts as isolated phenomena tends to ignore the fact that language is a system of hierarchically organized levels. The phonetic surface structure is the

---

1Getting people to pronounce words backwards is not easy, particularly when they go by the spelling instead of the pronunciation. One person said /ɛl.əb/ for bottle because of the final orthographic e.
tip-of-the-iceberg, so to speak. Although we have instruments to record and measure our phonetic output and no instruments to record the underlying internalized linguistic structure which is responsible for that output, this should not mean that theories about the underlying structure are unsupportable conjectures. (Some are of course, but some phonetic measurements can turn out to be spurious too.) The importance of a happy relationship between phonetic and phonological research is vital to linguistic research in general. It seems to me to be self-defeating on both sides to do research in a vacuum. The goals of both fields would suffer as a result.

6. Further Research

Part of the value of any scientific research lies in the questions it raises (as well as answers) and in the additional research it generates. Several further paradigms for looking at the perception of syllabicity in English have emerged from both analyzing the present data and talking with people who are interested in similar questions. In this section, I briefly mention several, hoping to interest the reader in the investigation of syllabicity, as well as elicit comments and additional suggestions.
6.1. Fundamental Frequency and the Perception of Syllabicity

The acoustic cue in the present study was duration. As a suprasegmental, it was not found to be perceived categorically as syllabicity. It is possible to try another prosodic feature, like fundamental frequency, to see whether different results obtain. Preliminarily, it must be determined what the F0 contour is like in natural speech productions of the key word pairs. Lehiste (personal communication) has pointed out that one might find that the changing F0 contour begins with the onset of voicing in a word like BLOW but does not begin to change until the second syllable in a word like BELOW. (If that happened one would also gain information about where the syllable boundary is.) If this were the case, then varying the timing of the onset of the pitch contour would be the variable to manipulate. Based on the results of the labeling tests in the present study, a stimulus that received fairly ambiguous responses as mono- or disyllabic would be used, and the hypothesis would be that manipulating the timing of the onset of the pitch inflection is a cue to the number of syllables in the word.

6.2. Intensity and the Perception of Syllabicity

I am grateful to Ilse Lehiste for bringing to my attention the importance of analyzing natural speech F0 contours.
The format for the examination of intensity and syllabic perception is similar to that of $F_0$ and syllable. First one would have to look at natural speech amplitude curve differences over the liquid in the mono- and disyllabic word pairs. One finding might be that in the disyllabic member, there is a changing amplitude curve on the syllabic liquid, but a level curve on the monosyllabic member. Then, again using an ambiguous stimulus from the duration experiment, one could manipulate the amount of intensity that is distributed over the liquid portion of the word. The hypothesis would be that increasing the amplitude of the liquid is a cue to the perception of the word as disyllabic.

I mentioned in Chapter II that some disyllabic words with initial /s/, like SEED-SECEDE, STAINT-SUSTAIN and SPORT-SUPFORT were produced in the Casual Fast style with what impressionistically sounded like a long initial /s/ and no following vowel. But in synthesizing the monosyllabic counterparts, lengthening the /s/ alone did not successfully simulate the disyllabic words. Since the data showed that the Casual Slow version of SECEDE seemed to show two intensity peaks over the long initial /s/, this acoustic cue might be used to investigate the perception of syllabic of words with initial fricatives. Specifically, both the duration of the /s/ and the extent of amplitude decrease in the middle of the /s/ would be manipulated for pairs like
SEED-SECEDE and STAIN-SUSTAIN. For SPORT-SUPPORT, it was suggested by Ilse Lehiste to decrease the intensity on the /s/ in SPORT near the /p/ boundary. The variable to be manipulated would be the timing of the onset of the amplitude decrease of the /s/. What these experiments suggest about the production of syllabic in fricatives is related to Stetson's chest pulse theory. I referred in Chapter I to Ladefoged's research on the correlation between the syllable and the chest pulse. (Ladefoged, 1967) He claimed that a word like sport may be produced with two bursts of activity by the intercostal muscles, the first burst associated with the /s/. What might be going on when SUPPORT is produced with a long /s/ is that the decrease in muscular activity starts before the /p/ articulation. Similarly SECEDE with a long /s/ and two intensity peaks might indicate that there are two peaks of muscular activity in the /s/ itself.

6.3. Syllabic in Word-Final Position

If increasing the steady state duration of a liquid in a stop-liquid-vowel environment can cause the liquid to be perceived as syllabic, would the same results occur in the environment vowel-liquid/nasal-consonant? For example, would increasing the steady state of the /r/ in horde

\[------
1David Wong has stimulated my thinking about this experiment.
produce the word horrid?\(^1\) I do not envision particularly successful simulation of the disyllabic words with this technique. For one thing, syllabic resonants are typically described as arising from V+R sequences, not R+V. The reason, I think, is that the transition to the resonant contains more of the resonant's characteristics (nasalization, retroflexion, etc.) than the transition from the resonant to the following segment. It is probably unlikely that a Casual Fast natural speech production of horrid would be homophonous with horde, as opposed to the present study where casual fast POLITE did impressionistically sound like PLIGHT. If this final position syllable did work out, then duration would have to be considered a more potent cue to syllabicity than the present study showed, considering that syllabicity did not turn out to be perceived categorically.

6.4. Duration of Transitions

In the present experiment, only the steady states of the liquids were manipulated. Transitions were not altered, but synthesized to directly match the digitized version of the natural speech recordings. Another way of using duration as an acoustic cue would be to extend the

\(^1\)The horde-horrid example is not the best since there are some Eastern American dialects where the vowel in horde is close to /ɔ/ and in horrid close to /a/. But there are other pairs like pelt-pellet.
transition to the liquid, instead of just its steady state. In the case of pairs with initial voiced stops, like BLOW-BELOW, the transition from the release of the /b/ to the steady state of the /l/ would be 'stretched' so to speak. The beginning and ending formant values would be specified for each overall transition duration and the appropriate interpolation would be performed. This procedure holds less promise for words with initial voiceless stops because the transitions have aspiration associated with them. The SPORT-SUPPORT data reported in Chapter III already showed that increasing the aspiration alone could not force identification of the stimuli as disyllabic. There is a difference between what was done to the SPORT-SUPPORT stimuli and expanding the transitions of voiceless stops. In SPORT-SUPPORT, a single frame from the center of the aspiration period was incremented instead of interpolating between beginning and ending frequencies. I could not say whether stretching the transitions as a whole would result in better (i.e. more categorical) identification.

6.5. Syllabicity Identification

The final suggestion for further research is one that has been mentioned in an earlier Section of this Chapter, 'Interpretation of Labeling and Discrimination Tests'. One of my comments about the categorical nature of the labeling
results was that responses might have been less categorical if the instructions had been to say whether each stimulus had one syllable or two. I would like to further suggest that instead of meaningful English words, nonsense words be used. For example, /grob/ - /gərəb/, /plɪd/-/pəlɪd/, etc. This technique would avoid any lexical bias and of course the 'words' would all be plausible English morphemes. It would even be possible to make the identification more open-ended by telling listeners to write down the number of syllables each stimulus had, instead of making the task a forced choice: one or two syllables. The basic purpose in using this experimental design would be to try to reconcile the results of the labeling and the discrimination tests which produced such disparate results. If, as I propose, the fault lies with the lexical forced choice labeling task, it would be worthwhile to change the task while still aiming for the same type of information.

7. Summary

The purpose of this study was to investigate the question of whether the perception of syllabic ity in English is categorical. Using synthetically produced stimuli, I ran identification and ABX discrimination tests. I synthesized the words BLOW, PLIGHT, BROKE, DRESS, CREST, PRAYED and SPORT and then lengthened the steady state of the /l/ or the /r/ (or aspiration of /p/ in SPORT) in 10 msec increments.
The identification task was to judge whether each stimulus was a token of one of the monosyllabic words listed above, or a token of its disyllabic counterpart: BELOW, POLITE, BAROQUE, DURESS, CAressed, PARADE or SUPPORT. In addition, subjects were divided into five speech style groups. Four groups heard a tape where each stimulus was preceded by a precursor phrase intended to suggest a particular speech style. The styles were Formal-Slow, Formal-Fast, Casual-Slow and Casual-Fast. The fifth group served as a control and heard the stimuli with no precursor phrase. The purpose of this aspect of the task was to determine whether it would be possible to shift the listeners' perception from monosyllabic to disyllabic (or vice versa) by using one of the precursor phrases.

The results of the identification tests showed that there were no differences among the five speech style groups. I concluded that subjects were unable to differentially judge the syllabic ity of the words using the reference of a particular speech style. But the results also showed crossovers in responses such that when the /l/ or the /r/ (although not the /p/ aspiration of SPORT) was short, listeners perceived the word as monosyllabic, and when the liquid was long it was perceived as disyllabic. At the medial durations though, the stimuli were often ambiguous, and the crossovers that were obtained were not steep. I concluded that the labeling test pointed to
moderately categorical perception.

There were differences among the six word pairs in terms of where the crossover occurred. For all words, the steady state of the liquid ranged from 20 or 30 msec to 140 or 150 msec. The earliest crossover occurred in PLIGHT-POLITE at 49 msec, the latest occurred in BLOW-BELOW at 86 msec and the others fell between 60-70 msec. In explaining these differences, I suggested that early crossover durations could be associated with preceding voiceless stops. The reason is that liquids following voiceless stops are usually voiceless, so that when even a small amount of voicing is added, perception will shift from the monosyllabic to the disyllabic word. I also suggested that BLOW-BELOW showed a particularly late crossover because of the complicated acoustic cues associated with /l/ as opposed to /r/. I considered these explanations tentative because of the small number of words used in the experiment.

The ABX discrimination experiment consisted of one-, two- and three- step tests, using the PRAYED-PARADE word pair. The three listeners were unable to discriminate the one-step differences above chance level, although the two-step differences were slightly above chance and the three-step well above. But the discrimination curves showed no peaks at the 'syllabicity boundary', and I had to conclude that the perception of syllabicity was continuous, not categorical.
In attempting to reconcile the difference between the categorical-looking results of the labeling test and the non-categorical results of the discrimination test, I felt that the discrimination paradigm was a more powerful barometer of categorical perception than the labeling paradigm. In particular, the forced choice labeling task of lexical identification should be replaced by a task where the listeners actually note how many syllables a stimulus has. This might make the identification results look less categorical and more congruent with the results of the discrimination tests.

The implications of this type of research for theories of speech perception and phonological theory often reduce to two questions. (1) Is the syllable a more basic unit of speech perception than the phoneme? (2) If duration serves as an acoustic cue to syllabic identity in English, is syllabic phonemic in English? I argued that the first question is not really framed properly since there seems to be evidence for both the phoneme and the syllable as the minimal unit of speech perception. Instead, based on an article by Studdert-Kennedy (1976), I drew a distinction between the segment/phoneme, as a linguistic unit and the syllable, as an articulatory-perceptual unit. In response to the second question, I said that syllabic identity could only be considered phonemic if one posited syllabic resonants in underlying representations. But there are good reasons for claiming
that English does not have underlying syllabic resonants, namely that syllabic resonants can always be derived from an underlying vowel + resonant. I concluded that syllability is not phonemic in English.
BIBLIOGRAPHY


27. Fujisaki, H. and T. Kawashima, Same Experiments on Speech Perception and a Model for the Perceptual


1956.


73. Lehiste, I., The Units of Speech Perception, *Speech and Cortical Functioning*, edited by J.H. Gilbert, New York:


