

THE ROLE OF CONSONANT CONSTRUCTIVES AND  
TRANSITIONS IN THE PERCEPTION OF ONGOING SPEECH

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

ALLAN KARL BIRD

1972

LIBRARY  
Michigan State  
University


This is to certify that the  
thesis entitled  
THE ROLE OF CONSONANT CONSTRUCTIVES AND  
TRANSITIONS IN THE PERCEPTION  
OF ONGOING SPEECH

presented by

Allan Karl Bird

has been accepted towards fulfillment  
of the requirements for

PhD degree in Audiology and  
Speech Sciences



Major professor

Date 11-10-72

ABSTRACT

THE ROLE OF CONSONANT CONSTRICTIVES AND  
TRANSITIONS IN THE PERCEPTION  
OF ONGOING SPEECH

by  
Allan Karl Bird

Considerable research has been undertaken in an effort to determine the critical acoustical elements involved in the perception of speech. Almost all of the studies to date have used either nonsense syllables or single words as the speech sample. The major research has proposed two major sources for identification of consonants. One of these is the consonant constriction and the second is the transition or the shift from the constriction to the vowel. Recently the question has been as to whether the perception of ongoing speech is similar to the perception of syllables or words.

This study was designed to answer the following questions:

1. Does the use of context provide cues for the perception of words which have been acoustically altered by the removal of various phonetic segments?
2. How much intelligibility is maintained because of co-articulation when only the vowel sound is heard both in context and out of context?

3. Is there a significant difference in the perception of contextual speech when transitions are removed and when constrictive elements are removed?
4. If significant differences occur in ongoing contextual speech, do the same differences also appear in non-contextual sentences?

Eight experimental tapes were made. Four of the tapes used sentences of five words each. The other four tapes had the same 30 words in random order. The tapes within an order were altered as follows: one had all of the transitions missing; one had all of the constrictive portions missing; one had the constriction and transition portions missing; and one had the transitions and vowels missing.

Different groups of 10 listeners each heard each condition and every person in the group was asked to write down the words which they heard.

A statistical analysis of the obtained data was done to determine whether differences in results were significant. The analysis of variance was used and an F-ratio score indicated a significant difference between orders of presentation, differences in segmentation effect and interaction at the 0.01 level. Graphs were made to demonstrate what the interaction effect was.

A post hoc comparison was performed showing the significant seven crosswise comparisons to be contributing to the overall significance. Of special importance was the statistically significant comparison between the



meaningful sentence with transitions omitted and the meaningful sentence with constrictions omitted.

A confusion matrix was constructed to examine the type of perceptual errors which were made under each of the orders of presentation. It was found that, in classifying errors as errors of place, voicing and affrication, differences in these three perceptual categories did exist.

The vowels alone were found to contain little information although the subjects did appear to apply some linguistic constraints to the material as shown by the high number of the's that the subjects guessed. In terms of five word sentences, one would expect to find a the in each of the sentences.

All of the findings of the study suggest that contextual perception is a different process than when one hears the words only as single entities. Further, it appeared that the segments of the speech signal removed seemed to make a difference.

Accepted by the faculty of the Department of  
Audiology and Speech Sciences, College of Communication  
Arts, Michigan State University, in partial fulfillment  
of the requirements for the Doctor of Philosophy degree.

Thesis Committee: \_\_\_\_\_ Co-Director  
Leo V. Deal, Ph.D.

\_\_\_\_\_ Co-Director  
Oscar Tosi, Ph.D.

\_\_\_\_\_  
Herbert J. Oyer, Ph.D.

\_\_\_\_\_  
Gordon Aldridge, Ph.D.

THE ROLE OF CONSONANT CONSTRUCTIVES AND  
TRANSITIONS IN THE PERCEPTION  
OF ONGOING SPEECH

By  
Allan Karl Bird

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Department of Audiology and  
Speech Sciences

1972

G78471

## TABLE OF CONTENTS

LIST OF TABLES . . . . .	iv
LIST OF FIGURES . . . . .	v
<b>Chapter</b>	<b>Page</b>
I. INTRODUCTION . . . . .	1
Statement of the Problem and Purpose of the Study . . . . .	6
Importance of the Study . . . . .	6
Definitions of Terms Used . . . . .	7
II. REVIEW OF THE LITERATURE . . . . .	9
Theories of Speech Perception . . . . .	9
Distinctive Features Theory . . . . .	10
Motor Theory of Speech Perception . . . . .	10
Coarticulation Theory of Sound Production . . . . .	11
Phrase Theory . . . . .	13
Contextual Probability Theory . . . . .	14
Studies Related to the Perception of Consonants . . . . .	20
Studies Related to the Importance of the Constriction in Speech Perception . . . . .	22
Studies Related to the Importance of the Transition in Speech Perception . . . . .	35
Studies Relating to the Role of Context in Perception . . . . .	43
Studies Relating to Time Compressed and Interrupted Speech . . . . .	48
III. EQUIPMENT, MATERIALS, SUBJECTS, AND PROCEDURES . . . . .	54
Equipment . . . . .	54
Materials . . . . .	54
Preparation of Materials . . . . .	55
Subjects . . . . .	62

Procedures . . . . .	65
Analysis of the Data . . . . .	66
IV. RESULTS AND DISCUSSION . . . . .	68
Results of Comparing Conditions . . . . .	69
Effects of Context . . . . .	74
Effects of Segmentation . . . . .	80
Amount of Information Contained in Vowels . . . . .	83
Perceptual Differences in Ongoing Contextual and Non-contextual Speech . . . . .	84
V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS .	91
Summary . . . . .	91
Conclusions . . . . .	94
Recommendations for Further Research . . . . .	96
BIBLIOGRAPHY . . . . .	98
APPENDIX . . . . .	106

## LIST OF TABLES

Table	Page
1. Contents of the Experimental Tapes . . . . .	64
2. Group Means and Standard Deviations of Ten Subjects for Each of Eight Test Conditions Ranked in Descending Order . . . . .	70
3. Analysis of Variance . . . . .	71
4. Differences in Averages Contrasting Each of the Conditions with all other Conditions . . .	75
5. Number of Correct Identifications for Each of the 30 Words under Each of the Experimental Conditions . . . . .	76
6. Soderberg's Information Values of Words in Clauses Classified by Length . . . . .	78
7. Total Number of Correct Responses for Each of the Five Sentence Positions . . . . .	79
8. Classification of the Consonants Used to Construct the Confusion Matrices . . . . .	87
9. A Confusion Matrix for Meaningful Sentences with Transitions Omitted or Constrictions Omitted and First Order Approximations with Transitions Omitted or Constrictions Omitted . . . . .	88

## LIST OF FIGURES

Figure	Page
1. Schematic Diagram Showing the Place of an Incoming Perceptual Pattern in a Phoneme Class with the Aid of Statistical Information Based on the Immediate Past of the Message (Fry, 1964, p. 16) . . . . .	17
2. An Expanded Model Showing the Place of an Incoming Perceptual Pattern in Speech Perception . . . . .	19
3. Schematic Diagram of Identification Procedure . . . . .	27
4. Diagram showing Preparation of Tapes to Form the Eight Experimental Tapes used in this Study . . . . .	63
5. Standard Deviation and Mean Number of Words Correctly Identified for each Condition Based on 10 Judges in Each Group . . . . .	73
6. Confusion Errors for Contextual and Non-contextual Speech . . . . .	89

## CHAPTER I

### INTRODUCTION

Of all the abilities which man possesses, the one which gives him his most distinctive nature is the ability to use language as a tool of communication. It is little wonder that an active interest in speech and language has been present in man's search for self understanding. The question of language and speech has been of great interest to the audiologist, the linguist, the psychologist, and the speech pathologist. From the communication standpoint, there has been an active interest in the information-bearing elements of speech. Each person's interest in communication is concerned with knowing and understanding what the elements of the oral language code are which allow man to communicate with man. How one knows what another person has said and what allows a person to say it are questions of extreme interest.

Traditionally, the study of speech perception has examined the phoneme as the basic unit of speech; and the inquiries on speech perception have been directed



at discovering what the elements were which allowed the listener to recognize the differences in phonemes. A myriad of studies have been conducted looking at the phoneme as the unit of speech perception.

Recently, however, the use of the phoneme as the unit of perception has been questioned; and it has been suggested that perception of speech actually occurs along more meaningful lines with more meaningful units. Ladefoged and Broadbent (1957) suggest that the study of the isolated phoneme is probably of only limited value in understanding the perception of speech. Brady (1960) and Brady, House and Stevens (1961) found an increasing volume of evidence becoming available that the perception of ongoing speech and language entails operations that differ markedly from the perception of other acoustical events. Mattingly and Liberman (1970) have stated that analysis has shown that the same acoustical cues carry information for different phonemes and, furthermore, that the same phoneme is often signaled by differing acoustical information.

Another problem in the study of the isolated phoneme has been pointed out by Hendrik (1965). He states that it is only in the study of the isolated signal that the listener has all possible cues available to him. He says:

phonemic system only pertains to the exceptional and rather abnormal situation of isolated words enunciated

by one particular speaker. Application of the phoneme system to other situations, including the normal, is an as yet unproven extrapolation. (Hendrik, p. 426)

Harris (1953) conducted a study designed to look at the effects of using the phoneme as building blocks for ongoing speech. Consonant phonemes and vowel phonemes were joined together to form words. The resulting "ongoing speech" was found to be very unnatural sounding and highly unintelligible.

Lisker and Abramson (1967) found that certain cues, which on the basis of acoustical analysis allowed for the distinction between sounds in single words, were much less useful in ongoing speech. They suggested that in ongoing speech the cues which are available to the listener in the isolated word either are blurred or not present at all. They further hypothesized that correct perception was still possible because of the linguistic and contextual constraints which were available in ongoing speech. Therefore, they argue that even though the acoustical cue was not clear, the slack created was compensated for by context. Fry (1964) has stated that the most important perceptual information that a listener can have about an utterance is that it makes sense. The role of context appears to be one of considerable importance. Context can perhaps best be defined using the concept of information theory:

If we define a message as a sequence of code symbols, in the present case, phones, the accuracy of decoding a message of a given length will be greater than the accuracy of encoding the individual symbols if and only if there are conditional probabilities that are not all equal (Strauss, 1950, p. 709).

This idea can perhaps best be demonstrated by looking at the graphic symbols of our language. If, for example, a person is asked to spell a word and is given no information as to what the first letter of the word is, he must randomly guess what the first letter is.

Obviously even in this random guessing, the probability of certain letters occurring is greater. Assume now that the person has guessed that the first letter is "q." If the person has a knowledge of the English language, the chances of the person correctly guessing the next letter are excellent. Most people knowledgeable of the language would guess the next letter to be "u." Even when the case is not as obvious as the previous example, the speller is able to limit his alternatives by prior information. There are, as a result of certain learned or innate rules, constraints placed upon our selection.

The suggestion has then been made that the idea of context might also apply to the perception of the spoken code. Three types of context have been suggested which may aid the listener in the comprehension of ongoing speech (Ladefoged, 1967). The first type of context which he suggests might be of assistance to the listener is the "context of the situation." These are

events which occur at the time of the utterance but are not an intrinsic part of them. For example, a person watching a sporting event might assume that someone addressing him may be discussing the sporting event. The second type of cues are the linguistic cues of the language which the listener might apply. For example, the person hearing "he say" might well be able to supply the "s" to the utterance even though he had not heard it since he knows some linguistic rules regarding singular and plural. The third type of contextual cues are the acoustical cues.

Lieberman (1967) and Fry (1964) have both suggested that the listener relies heavily on the linguistic cues for speech perception.

Black (1964) has pointed out that since English words are not nonsense syllables, they cannot be randomly assembled. There are limitations which are placed on the alternatives from which a listener might select. Reed and Wang (1961) were able to demonstrate that contextual cues override acoustical cues for perception.

While several people have alluded to the fact that perception of ongoing speech varies from the perception of the isolated word, only a slight beginning has been made to investigate this question. Warren (1970) and Warren and Obsek (1971) found that a listener was able to supply a phoneme even when it was not present by

simply removing one sound from a sentence. They called this ability phonetic restoration.

#### STATEMENT OF THE PROBLEM AND PURPOSE OF THE STUDY

This study seeks to answer some basic questions regarding the perceptual ability of the speaker. Specifically, the study is designed to determine whether those elements of speech perception which have been suggested as being important in the perception of the isolated phoneme carry significance in ongoing speech.

The following questions were formulated to define the research:

1. Does the use of context provide cues for the perception of words which have been acoustically altered by the removal of various phonetic segments?
2. How much intelligibility is maintained because of co-articulation when only the vowel sound is heard both in context and out of context?
3. Is there a significant difference in the perception of contextual speech when transitions are removed and when constrictive elements are removed?
4. If significant differences occur in ongoing contextual speech, do the same differences also appear in non-contextual sentences?

#### IMPORTANCE OF THE STUDY

It has been hypothesized that context is important in the perception of speech. Further it has been suggested that various portions of the speech signal

carry varying degrees of information (Dukelskiy, 1967). This question has not been examined in ongoing speech. It is the purpose of this study to attempt to provide some basic information relative to phonemic restoration ability of the listener and to the present knowledge about the information-carrying units of speech.

The information obtained from this study is viewed as having practical application in the area of sound discrimination and information processing as well as having possible significant impact upon future research in speech perception.

#### DEFINITION OF TERMS USED

Constriction.--The term constriction or constrictive element as used in this paper refers to the portion of the consonant which represents the quasi "steady" part of that consonant. It is that portion of the consonants which Potter, Kopp and Green (1947) refer to as spikes and fills. The more generic term "constriction" was selected for use in this paper.

Context.--For the purposes of this paper, context has reference to the effect of linguistic restraints upon the perception of speech. It refers to the meaning the listener might gain as a result of syntactical, semantic, and morphological rules which the listener applies.

First order approximation.--A first order approximation sentence for the purposes of this study refers to a sentence which is constructed out of randomly selected words. The 30 words from the meaningful sentences were randomly assigned to the six first order approximation sentences by use of random number tables. These sentences have no contextual information.

Phonetic restoration.--Phonetic restoration is a term coined by Warren and Obusek (1971). It refers to the ability of the listener to supply or restore perception of the missing phonetic elements of a speech signal.

Segmentation.--Segmentation has reference to the process of temporally removing portions of the phoneme.

Steady state vowel.--The steady state vowel is the portion in which the amplitude spectrum of the vowel is quasi constant.

Transition.--In ongoing speech, there is a shift which occurs as the flow of speech moves from one phoneme to the next. In the case of a consonant-vowel syllable this would be a shift from the constriction to the steady state vowel.

2



## CHAPTER II

### REVIEW OF THE LITERATURE

This chapter contains a discussion of the theories of speech perception, a review of literature relevant to the perception of the consonants of the English language, a review of literature on the role of context in the perception of speech, and a discussion of time compression and speech interruption.

### THEORIES OF SPEECH PERCEPTION

A variety of opinions has been expressed concerning the nature of the perception of speech. The earliest idea concerning the speech code was that speech was a series of non-overlapping phonemes which were each self contained. It was assumed that traditionally speech was analyzed as a series of concatenated phonemes (Fant, 1962). Many of the early studies related to speech used this model.

Recently new models have been proposed in an attempt to explain the way in which man perceives speech. A brief discussion of these theories follows: Distinctive Feature Theory, Motor Theory of Speech Perception,

Co-articulation Theory of Speech Perception, the Phrase Theory of Perception, and the Contextual Theory.

### Distinctive Features Theory

It has always been considered that the phoneme was the smallest unit of speech which could be differentiated. On the basis of analysis of spectral representations of sound, Jakobson, Halle, and Fant (1961) were able to suggest that smaller perceptual and production units existed. These smaller units they called "distinctive features." They suggested that the individual phonemes were composed of bundles of smaller features. Twelve distinctive features were described by them. They further postulated that each of the features was perceived in a dichotomic manner, i.e. each feature was perceived as either being present or not being present. In the case of a phoneme, the listener decides not to what degree a feature is present but whether it is present or not. It was felt that in this manner the process of perception was considerably simplified. Other systems of distinctive features have been suggested by Chomsky and Halle (1968), Wicklegren (1966), and Miller and Nicely (1955).

### Motor Theory of Speech Perception

Motor theory suggests that the perception of speech is based upon awareness of articulatory movement of the

listener, attempting to produce unconsciously the same speech he is receiving. It suggests that the articulatory movements and their kinesthetic feedbacks of the listener mediate between the incoming acoustical events and the perception of speech. This theory was outlined by Cooper, et al (1952). Mattingly and Lieberman (1970) pointed out the same acoustical cues sometimes signal the presence of different phonemes while differing acoustical cues are sometimes present to make the same phoneme. They suggest that this makes clear the need for some mediational factors in the perception of the speech other than the acoustic information. They point out in the acoustical analysis of speech that different articulatory patterns are often represented by similar acoustical events and yet a listener is able to make the proper perceptual decision. Further, similar articulatory patterns are often signaled by varying and diverse acoustical representation. Still the listeners hear the correct sound. Lane (1965) points out that further support is lent to this theory because the perception of sounds more closely parallels the articulatory changes than it does the acoustical changes.

#### Coarticulation Theory of Sound Production

C. M. Harris (1953) conducted a study that recorded consonant-vowel syllables. The consonants were then spliced away from the original vowels with which they were recorded and spliced with new vowels. It was

found that this yielded unnatural and unintelligible speech. Harris suggested on the basis of his study that the use of individual phonemes as building blocks for speech was most likely an unrealistic assumption. He suggested that there was some type of interaction between sounds in combination.

The coarticulation theory of speech assumes that one or more of the sound features characterizing a sound segment may extend over several temporal segments of speech. This idea was suggested by Peterson and Barney (1952) and Hughes and Halle (1956); they concluded that it is not usually possible to find segments of the acoustical signal which can be uniquely identified as single and discrete entities during the flow of normal speech production. They speculated that this might mean that the unit of speech perception is larger than the phoneme. Stevens and House (1963) found evidence that the formant frequencies of vowels are modified in a systematic way by the consonantal contexts in which they occur and noted that the effects of coarticulation upon the vowel could often be found extending over one phoneme in either direction.

Stevens, House and Paul (1966) used Consonant-Vowel-Consonant utterances and found that the formant structures of certain vowels were modified by the consonant environments in which they were placed and found evidence of co-articulation over two phonemes.

This view was further supported by Ohman (1966) using Vowel-Consonant-Vowel syllables.

Two basic theories of coarticulation have evolved. One stresses that the perceptual unit is supra-phoneme in size. The other is explained on the basis of phoneme-sized units.

Several advocates of the supra-phoneme unit have proposed theories. Danilooff and Moll (1968) suggested that the unit of perception was a unit the size of the Consonant-Vowel-Consonant. They felt that this perception was not affected by word boundaries. Kozhevnikov and Chistovich (1965) agree that the unit of perception is larger than the phoneme but feel that the perceptual unit is the Consonant-Consonant-Vowel.

Henke (1966) has suggested that the unit of perception is still phoneme-sized but that the high level perceptual and production mechanism is responsible for the coarticulation effect. The process is one of higher level look-ahead and scanning.

Numerous studies have been conducted which have shown that in the production aspect of speech, coarticulation does occur (Fujimura, 1961; Moll, 1960, 1962; Danilooff and Moll, 1968; Amerman, et al., 1970).

### Phrase Theory

It has been argued by several that the spoken language is decoded by the listener in word groups called

phonemic clauses (Garrett, 1965; Garrett, Bever and Fodor, 1966; Fodor and Bever, 1965; Dittmann and Llewellyn, 1967; Fodor, Bever and Garrett, 1968; Bever, Lackner and Kirk, 1969; Bever, Lackner and Stolz, 1969). The general method of investigation was to place a click at various locations within a sentence. A point was marked at the place of the phrase juncture of the sentence. This position was called the zero point. Clicks were then recorded at the zero point and at points one syllable, one word, etc. away from the zero point in each direction. The findings were that the click or noise was generally mislocated and usually in the direction of the phrase juncture or zero point. Based upon the idea that a perceptual unit resists interruption, the general conclusion of the investigators was that since the noise or click migrated toward the phrase juncture, this indicated that the unit for decoding was the phrase or phonemic clause.

### Contextual Probability Theory

Lieberman (1967) proposed a model for the perception of ongoing speech. He suggests that the entire context of the sentence plays an important part in determining how essential any particular cue is for the perception of the message. As Black (1962) pointed out, words are not nonsense syllables. Words in any language have sounds that are not randomly assembled.

Brown and Hildum (1956) ran a study in which they constructed three sets of words. The first group of words were highly unfamiliar words of the English language, the second set were non-English words and the third set were words whose phonetic context could not occur in English. They found that the English words were remembered by native speakers of English about 69 percent of the time, the English non-words were recalled with 46.8 percent accuracy, while the non-English words were only recalled 9.7 percent of the time. They concluded that the perception of speech is directed by knowledge of the sequential probabilities.

Reed and Wang (1961) also provided evidence of this probability phenomenon. Three words were recorded in their study. They were bat, pat, and spat. The /s/ was then removed from the word spat and placed before bat and pat. Subjects perceived both s plus bat and s plus pat as being spat. The non-English combination of sb was not perceived even when actually present. They concluded that if the analysis were acoustical or motor rather than contextual probability, the listeners would have heard bat.

Harris (1960) has suggested that there are enough cues furnished by restriction placed on utterances by syntax and context to give the listener a good idea of the missing parts.

Lisker and Abramson (1967) conducted an experiment using the analysis of speech in isolated words and in ongoing speech. It was found that the time from the onset of the burst to the onset of periodicity was sufficient to classify the consonants as being voiced or voiceless. However, when the voice onset time was evaluated in ongoing speech, it became useless in predicting voiced and voiceless. They offered two possible explanations as to what happens. The first possibility suggested was that there is a general blurring of the distinctiveness of the sound. This blurring is compensated for by the linguistic constraints, thus taking up any slack in the intelligibility. The second alternative possibility offered by them was that there are certain cues in deliberate speech of single words which are merely redundant. Studies by Pickett and Pollack (1963) and Fry (1964) suggest that the blurring and linguistic constraint is the correct alternative.

Lieberman (1967) suggests that the perception of ongoing speech is based on both contextual and acoustical cues in combination. It is suggested that a listener first performs a preliminary analysis of the speech signal on those acoustical cues which are available to him. When the acoustical information has been processed, the listener postulates a hypothesis



regarding what the message might possibly be. This hypothesis is then referred to a process where it is tested using those semantic and syntactic constraints which exist in the listener's language system. If the hypothesis appears tenable in light of the constraints, the listener interprets the message as having been heard. Fry's model (1964) shown in Figure 1 illustrates this viewpoint.

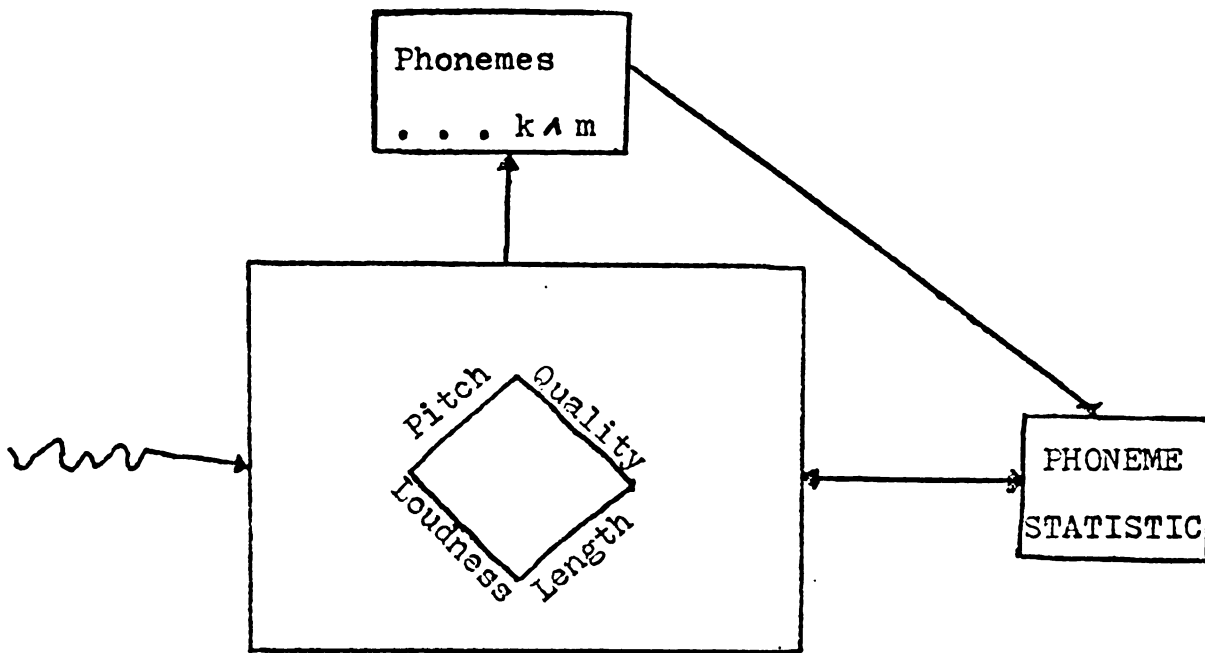


Figure 1.--Schematic diagram showing the place of an incoming perceptual pattern in a phoneme class with the aid of statistical information based on the immediate past of the message (Fry, 1964, p. 16).

In this model the incoming message is first analyzed according to the pitch, quality, loudness, and duration

of the component portions of the incoming message. Once as much information as possible has been gained from the acoustical message, on the basis of those phonemes which have been interpreted so far, a tentative hypothesis is formed. The hypothesis is then referred to the area called the phoneme statistic area. The probability of our hypothesis being correct is assessed; and if the assumption appears reasonable, the message is considered heard. If the hypothesis does not seem reasonable, further analysis of the message is made or the message is assumed not to have been heard. A further explanation using the example in the above model may make it clearer. Assume that because of the blurring which occurs in the acoustical properties during ongoing speech, according to the acoustical signal, we are able to determine that the sound which was heard was one of the nasals, i.e. /m/, /n/, or /ŋ/. The message can then be referred to the probability or statistical phoneme area. Here it is found that the most likely correct choice for that context is the /m/. This is then referred back to see if the assumption can be supported by what is further heard. The model as explained above would hold true if speech perception were based on the phoneme. However, it is very simple to expand this model to include all possible perceptual units. The expanded model is shown in Figure 2.

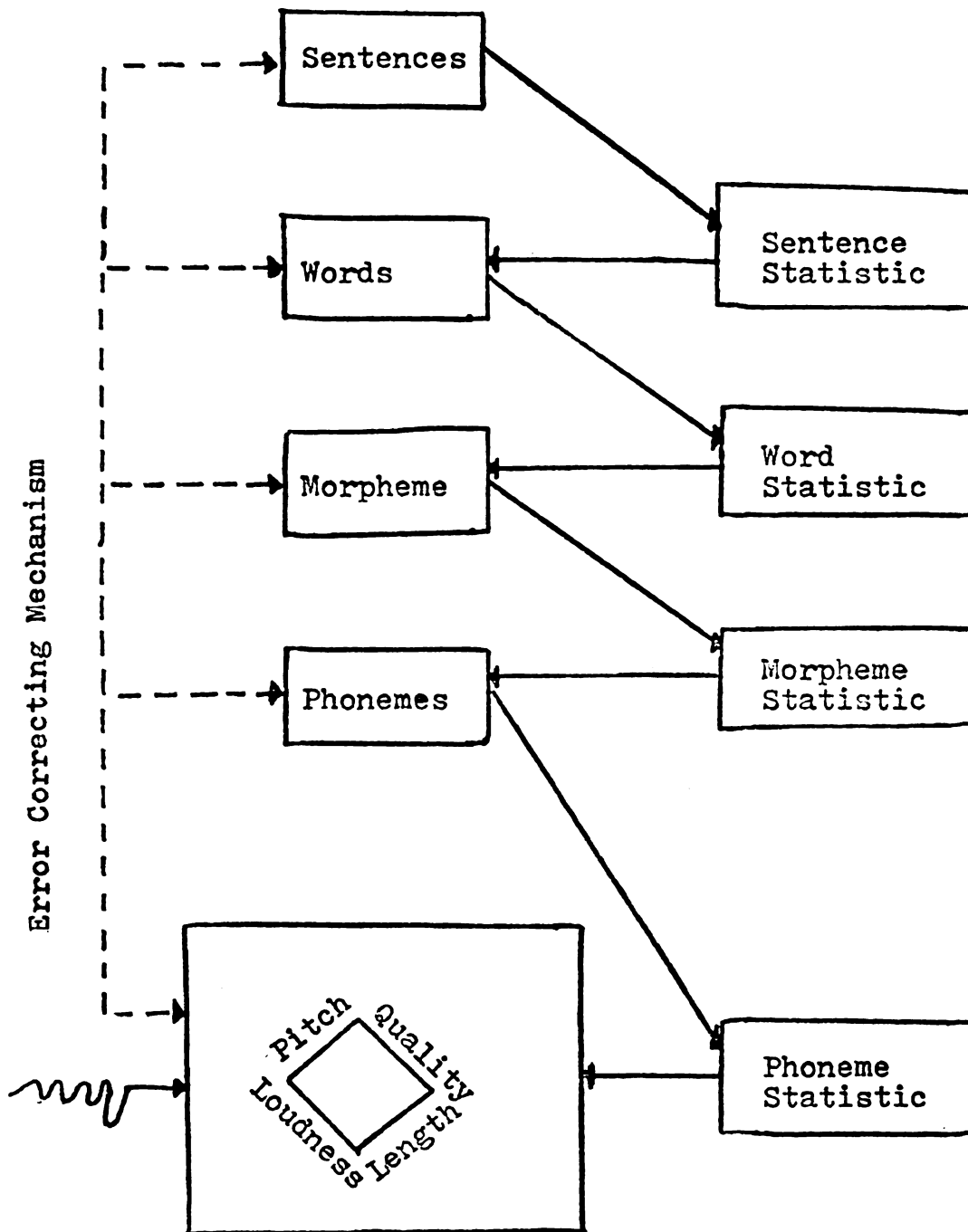


Figure 2.--The principle of statistically-based identification of linguistic units extending to all levels of operation (Fry, 1964).

At the present time, none of the above suggested models have been conclusively validated as being the model whereby perception occurs. It is further possible that under varying conditions, differing methods of perception are employed.

#### STUDIES RELATED TO THE PERCEPTION OF CONSONANTS

Three approaches have been employed in attempting to discover which elements of the speech signal allow for its perception. The first method is analysis. In this method, a recording of some speech signal is made. This sample is then analyzed by use of electrical equipment such as the spectrograph which allows the speech signal to be displayed in a spectrogram. The display is then examined for any consistency which might occur. The basic problem existing in this method is that one cannot be sure that those elements which are selected for the basis of analysis are the same elements which are used by the human auditory system in analyzing the same signal.

The second approach, and the one used in the current study, is distortion. In distortion the speech signal is altered in some way such as compression, clipping, or bandwidth limiting. When the distortion has occurred, the signal is then presented to listeners

in such a way that the human ear serves as the analyzer for the speech signal.

The third approach employed is synthesis. In synthesis the speech signal is in some way artificially created, and the signal is then presented to the human auditory system to ascertain what it perceives. The earliest method employed involved simplified handpainting of spectrograms. These are drawn and then are converted into speech sounds. This is the playback pattern system. The main type of synthesizers presently used is the analog-computer type. Each of the methods has some strong points and each has some inherent weaknesses.

Various phases of the consonants have also been explored. Liberman (1957) has suggested that cues for the perception of the consonants can be divided into three major classes: constriction, transition, and resonance. Constriction cues are those cues which are the noise of the consonants. In the case of the plosives, the constriction cues are the stops and bursts of sound. For the fricatives, the constriction cues are the frictional portion of the sound. The affricate sounds have both burst and friction constriction cues. The transitional cues are those based on the movement of the acoustic energy from the locus of the consonant to the steady state portion of the vowel. Resonant cues are used to classify the nasal sounds as nasal consonants

and all other consonants as being non-nasal. The majority of the studies which have been performed looking at these classes of cues have been done using either words or monosyllabic nonsense syllables.

Studies related to the importance  
of the constriction in speech  
perception

One of the earliest studies relating to the importance of the constrictive in the perception of sounds was conducted by Liberman, et al. (1952). Using speech synthesis, they found that the perception of the /t/, /p/ and /k/ was based upon the presence of noise at various frequency ranges. For the /p/ sound, the presence of noise in the area of the first formant was significant. For the /k/ and /t/, it was the presence of noise in the area of the second and third formant areas respectively.

C. M. Harris (1953) found that the /t/, /tʃ/, /ʃ/, /s/, /l/, and /r/ sounds were recognized on the basis of the constriction, regardless of the vowel context away from which the consonant was spliced.

Harris (1954, 1958) synthesized the fricative consonants /s/, /ʃ/, /θ/, and /f/. Based on the constrictive portion of the consonants, the listeners were able to divide the sounds into two groups, the /s/ and /ʃ/ group and the /θ/ and /f/ group. Further, on the basis of the constrictive component, the listeners were able to



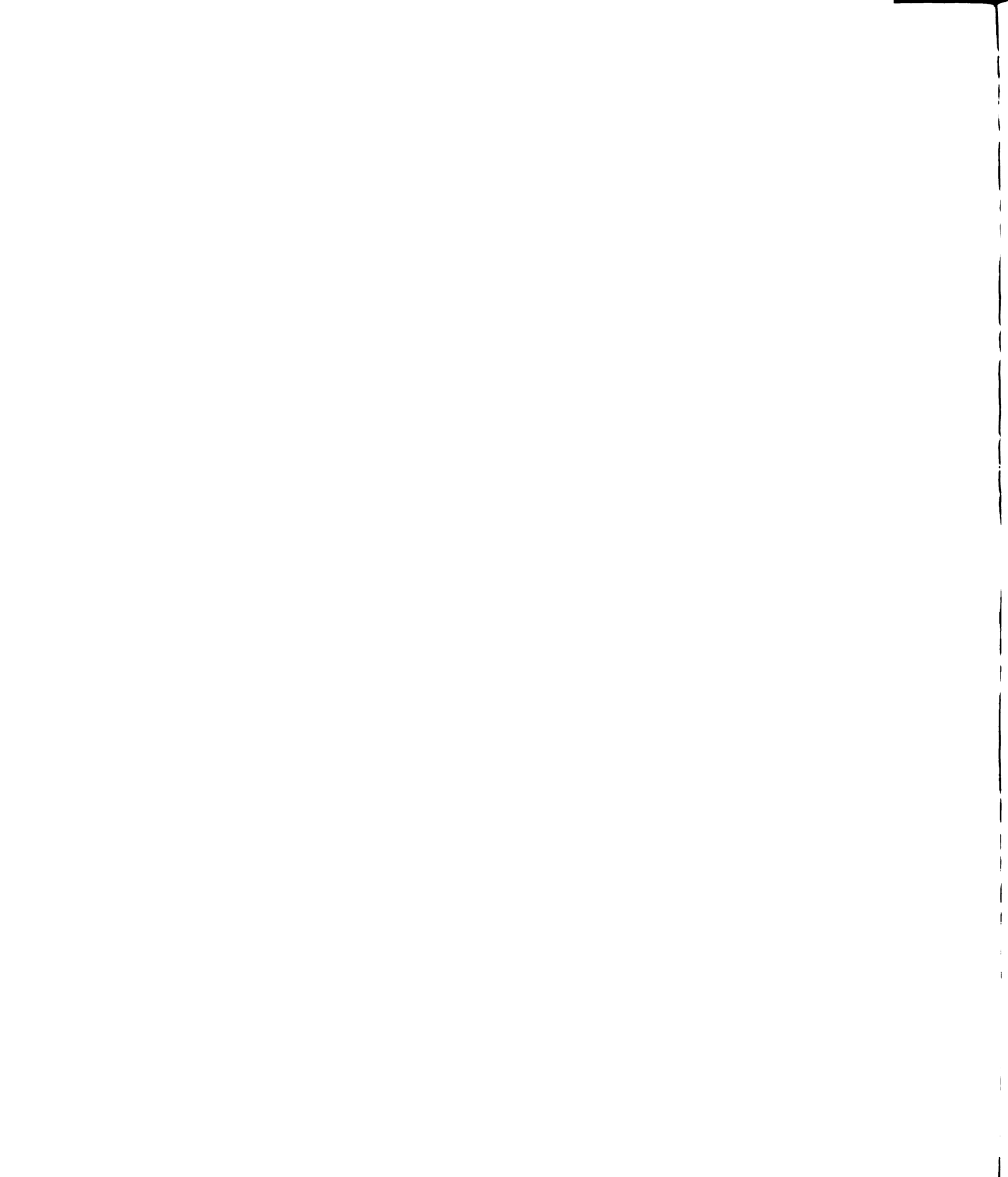
differentiate between /s/ and /ʃ/. The /s/ was perceived when the lower cut off frequency was at 3600 Hz. For the /ʃ/ it was best at a lower cut off frequency of 2000 Hz.

Cooper, Delattre, et al. (1952) speculated that it was possible to distinguish the /p/ and /k/ sound from the /t/ sound on the basis of the burst portion of the speech signal. These findings were based on work done with speech synthesis.

In 1954 Fischer-Jørgensen conducted a study of the acoustical elements of the Danish stop consonants. The results of this study were based upon spectrographic analysis of the stops /p/, /t/, /k/, /b/, /d/ and /g/ in the Danish. Several factors were found to be different in the constrictive or burst portion of the /p/, /t/, and /k/. Similar results were found for the cognates of each of the phonemes.

He found that the intensity of the explosion showed characteristic differences. It was found that /k/ was more intense than /p/, and /p/ was more intense than /t/. It was pointed out, however, that the results were somewhat compounded by the environment in which the sound occurred. Differences in duration were also present. The /g/ cognates were found to be longer than the /d/, which was longer than the /b/. Duration of the aspiration was also found to be a factor in the spectral analysis of the stop consonants. Several other variations were pointed





out which were spectrally present. In summarizing the acoustical characteristics of the stops, Fischer-Jørgensen suggests that adequate recognition of the plosive sounds is based on certain characteristics. The p/b cognates were found to have a fairly neutral explosion with relatively low resonance qualities. The t/d cognates were recognized by the high explosion and the relatively high resonance qualities. The k/g were recognized by the strong concentration of explosion in the area of the frequency of surrounding sounds and a resonance quality which is also greatly influenced by the presence of surrounding sounds.

An interesting observation was made by Fischer-Jørgensen regarding the influence of sounds which surround the sound under investigation:

From a phonemic point of view we would say that only differences between conditioned variants (or "allophones") must be phonetically constant, not differences between phonemes; thus the difference between /k/ and /t/ before /i/ must show some constancy if communication by means of speech is to be possible, but even this constancy is of limited kind, for there may be a bundle of differences which are all present in optimal cases, but which need not all be there. In some cases the frequency of the explosion might be the same but the duration might be different, or vica versa. Final /k/ and /p/ after /u/ may have the same formant bendings but different explosions or if they are unexploded, there must be differences of formant bending, etc. But there has to be some kind of consistency (p. 58).

In general, Fischer-Jørgensen felt that his findings were consistent with the findings of studies which had been conducted earlier at Haskins Laboratory using synthetic speech.

portion  
identifi-  
success  
point w  
the fri  
as /tʃa,  
the syl  
the bas  
duration  
suffici  
as bein

propert  
sounds  
relativ  
the con  
noted b  
compon  
counte

a conce  
than the  
/f/ and

in which

Gerstman (1956) found that duration of the noise portion of the fricative was significant in its identification. The syllable /ʃa/ was recorded and successive portions of the fricative were removed. At a point where approximately one-half of the duration of the friction had been removed, the syllable was perceived as /tʃa/. If all of the fricative duration were removed, the syllable was perceived as either /ka/ or /ta/. On the basis of his study, Gerstman suggested that the duration of the constriction portion of the consonant is sufficient to allow the listener to identify the sound as being a plosive, a fricative, or an affricate.

Hughes and Halle (1956) investigated the spectral properties of the constriction portion of the fricative sounds of English. The method used was analysis. The relative energy of the various frequency components of the consonants was obtained. A strong energy component was noted below 700 Hz for all of the voiced sounds, but this component was never present in any of the voiceless counterparts.

Further analysis showed that the /s/ cognates had a concentration of energy which was consistently higher than those present in the /ʃ/ cognates. Results for the /f/ and /v/ were somewhat confusing.

The authors devised a mathematical computation in which three measures were made. First, the energy

in the

subtrac

720 Hz

band fr

the ban

last me

region

energy

subtrac

the pea

procedu

procedu

each of

There a

percept

those s

more re

analys

they we

constr

inform

order

stops

transi

how th

in the frequency range between 4200 Hz and 10,000 Hz was subtracted from the energy present in the range between 720 Hz to 10,000 Hz. Second, the energy present in the band from 720 to 6500 Hz was computed and the energy in the band from 720 to 2150 Hz was deducted from this. The last measurement was made by locating the peak in the region from 1500 cps to 4000 Hz. Once this was done, the energy in dB in a band from 720 Hz to 1370 Hz was subtracted from energy in a 500 Hz band centered around the peak frequency. Figure 3 shows the identification procedure then proposed. Using this identification procedure, Hughes and Halle were able to identify correctly each of the sounds with no greater than a 10 percent error. There appeared to be no consistency in the error. A perceptual study was also conducted, and it was found that those sounds which were within the criteria range were more readily identified correctly by listeners.

Halle, Hughes and Radley (1957) conducted an analysis of the stop consonants. The first question which they were interested in was whether or not the burst, or constrictive, portion of stops contained sufficient information for the correct identification of the stop. In order to answer this question, they recorded each of the stops and then gated the monosyllables so that none of the transition was heard. No information was available as to how this gating was done. The portion of the burst heard

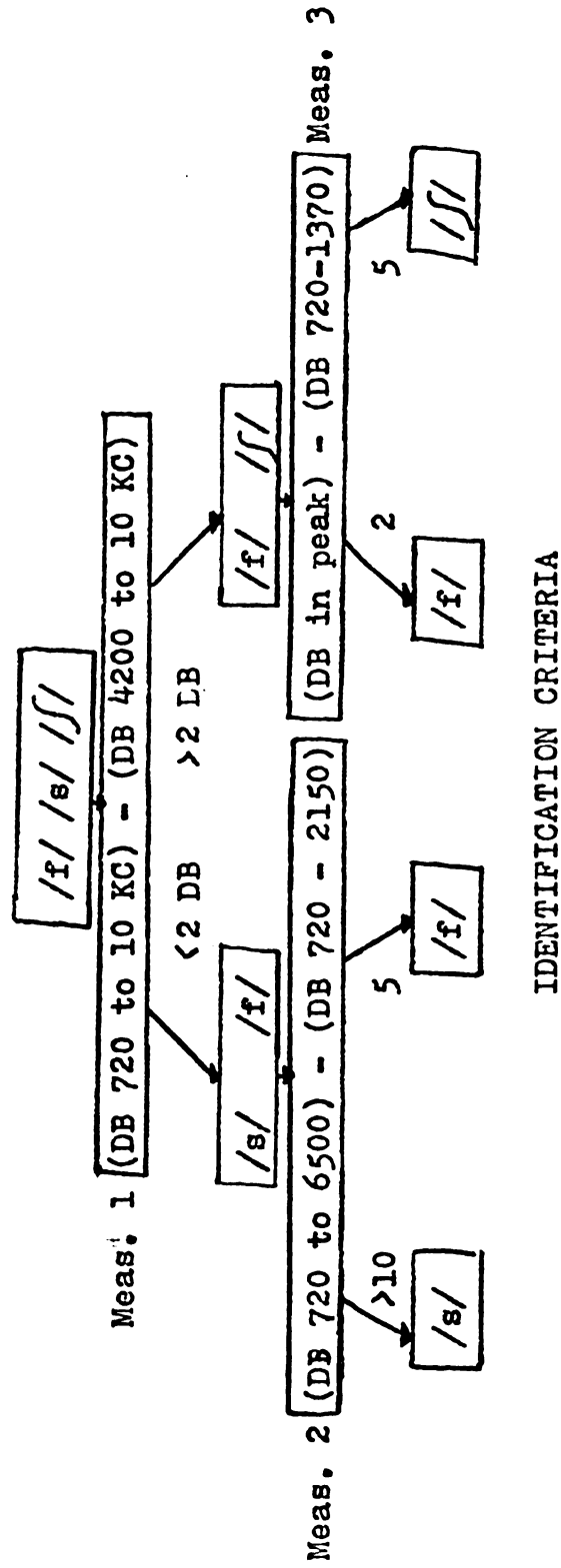


Figure 3.--Schematic diagram of identification procedure.

was 20 msec.

20 msec.

percenta

judges w

and 96 p

scores w

experien

that lea

was repo

the basi

undertak

various

concentr

500-1500

predomin

The vel

frequen

the fin

release

altered

also fe

conveyi

to comp

and /d/



was 20 msec. long. The taped stimuli containing the 20 msec. bursts were then presented to subjects. The percentage of correct identification for the five best judges was 65 percent, 70 percent, 75 percent, 80 percent, and 96 percent. They pointed out that the three best scores were obtained by people having had considerable experience in listening to isolated bursts. It was felt that learning could be a factor in these cases. Since it was reported that correct identification could be made on the basis of the burst, further analysis of the burst was undertaken. A frequency analysis of the energy in the various frequency bands showed that for the /p/ and /b/ the concentration of energy was in the frequency range between 500-1500 Hz. For the alveolar plosives there was a predominance of energy in the frequencies above 4000 Hz. The velar plosives had concentration in the middle frequencies. These results were in good agreement with the findings of Fischer-Jørgensen.

Malecot (1958) found that suppression of the release in identification of final stops drastically altered the perception of the voiceless stops. It was also felt that the release played a major role in conveying the manner of articulation.

Hoffman (1958) used a speech synthesizer in order to compare the perception of the voiced plosives /b/, /g/, and /d/. The results of his study yielded findings similar

to the r  
Cooper e  
exceptio  
speech s  
percepti

nasal co  
identifi

numerous  
that the  
to allow  
stops.

analysi  
that ot  
the ons  
serve a  
While n  
appear  
Fischer

duratio  
all fac  
between

voicela

to the results of the voiceless counterparts conducted by Cooper et al. (1952). He further found that with the exception of the /b/, the constrictive portion of the speech signal provided adequate information for the perception of the voiced plosives.

Nakate (1959) found that the duration of the nasal consonants was significant in making the correct identification.

Liberman (1957) in summarizing the results of numerous studies conducted at Haskins Laboratory stated that the position of the frequency burst was sufficient to allow the listener to distinguish within the voiceless stops. This seemed to agree with the findings of other analysis and synthesis studies. Liberman also speculated that other factors such as the duration and the nature of the onset of noise or the intensity of the noise might serve as cues for the perception of the consonant sounds. While no research data were presented, these speculations appear to agree with the analytical findings of Fischer-Jørgensen.

Ingeman (1960) suggested that low hiss sound, duration of friction and the intensity of the sound were all factors which were helpful in differentiating between /s/ and /z/.

Lotz, et al. (1960) compared the voiced and voiceless stops. It appeared that the lack of aspiration

following

be the M

a sound

interval

to inser

(Bastiar

Bastian

silence

resulte

rather

no cons

reason

it appe

basis o

a meani

found t

/f/, /s/

that t

in the

the ce

freque

/s/ re

betwee

with t

following the initial constrictive release seemed to be the major cue in causing the listener to evaluate a sound as being voiced rather than voiceless.

It has been found that the introduction of an interval of silence is enough to cause the listener to insert a plosive sound even when none was present (Bastian, et al., 1961). Using synthesized speech, Bastian created the words sore and slit. A period of silence was then introduced following the /s/. This resulted in the hearer perceiving the words as store rather than sore, and split rather than slit. While no consideration was made in the article regarding the reason for the selection of the stop which was inserted, it appears that the selection could have been made on the basis of meaning, since the stop inserted still maintained a meaningful word.

Heinz and Stevens (1958) using synthetic speech found that the center frequencies for the five consonants /f/, /s/, /ʃ/, /θ/, and /ç/ were different. He suggested that there is a consistent pattern which would be useful in the identification. It was found that for the /ʃ/, the center frequency was around 2000 Hz. The center frequencies were around 3500 and 5000 for the /ç/ and /s/ respectively. No consistent difference was found between the /f/ and /θ/. This would appear to agree with the findings of Harris (1958), findings which

suggest t

the /ɛ/ a

of the c

studied

the spec

intensit

fricativ

following

1.

2.

3.

4.

suggest that it is not possible to distinguish between the /f/ and /θ/ on the basis of the constrictive portion of the consonant.

Strevens (1961), using spectrographic analysis, studied nine fricatives. Line spectra were made from the spectrograms. It was found that the relative intensity and the frequency components of the nine fricatives used in the study varied. He reported the following findings:

1. ~~/ʒ/~~ /ʒ/ Lowest frequency at which energy is visible on the spectrogram is between 1600 and 1650 cps. Low peaks of energy tend to occur around 1800-2000 cps., 4000-4500 cps., and 5500 cps. Energy rarely above 6500 cps. Intensity ranking: lowest of 9.
2. /f/ Lowest frequency is around 1500-1700 cps. Low peaks of energy tend to occur around 1900 cps., 4000 cps., and occasionally 5000 cps. Upper limit of frequency is rarely below 7000 cps., usually around 7500 cps. In general, a higher upper limit than No. 1. Intensity ranking: 3rd in ascending order.
3. /θ/ Lowest frequency varies, but lies between 1400 and 2000 cps. Low peaks of energy tend to occur, the lowest being close to 2000 cps., the upper peaks varying somewhat, but tending to lie about 1000 cycles apart. Upper limit of frequency rarely below 7200 cps.; some speakers reach 8000 cps. In general, a somewhat higher upper limit than No. 2. Intensity ranking: 2nd in ascending order.
4. /s/ Lowest frequency almost always above 3500 cps. Peaks of energy tend to occur with no apparent pattern, except that they do not lie closer to one another than 1000 cycles. Upper limit of frequency exceeds 8000 cps. in most cases. Intensity ranking: 5th in ascending order.

5.

6.

7.

8.

9.



5. /ʃ/ Lowest frequency varies between 1600 and 2500 cps. Peaks of energy tend to occur not less than 1000 cycles apart and the aspect of amplitude cross-sections shows a sharp cut-off around 7000 cps. Intensity ranking: 7th in ascending order.
6. /ç/ Lower limit of frequency varies generally between 2800 and 3600 cps. Peaks of energy tend to appear at roughly 1000 cycle intervals; these peaks are sharper than those in No. 5. Upper frequency limit very variable, but usually between 6000 and 7200 cps., i.e. lower than for either No. 4 or No. 5. The general shape of the spectrum is like that of No. 4 /s/, but with all values transposed 1000 cps. down. Intensity ranking: greatest of all 9 items.
7. /s/ Lower limit of frequency usually between 1200 cps. and 1500 cps. There is always a strong peak of energy below 2000 cps., with others above about 3500 cps. The aspect of amplitude cross-sections gives a hint of formant-like structure; the low peak is steeper than the upper peaks, which are often double peaks, some 500-600 cycles apart. Upper frequency limit is very variable, usually between 5000 cps. and 7500 cps. A considerable variety of different sound qualities was obtained from the subjects, as was to be expected. Versions judged in phonetic terms to have a more back place of articulation tended to approach more closely to a formant-like structure. Intensity ratings: 6th in ascending order.
8. /x/ Lower limit of frequency varies between 700 cps. and 1200 cps. All spectra bear a marked resemblance to vowels, with a "formant" of one or two high peaks between 1000 cps. and 2400 cps. Sometimes there are 3 or even 4 "formants" altogether, with 1 or 2 of them having rather high peaks of intensity, in the region from 3000 cps. to 6000 cps. At first glance the spectrographic pattern is that of a vowel rather than a voiceless fricative. Upper limit of frequency variable between 6000 cps. and 7000 cps. Intensity ratings: 4th in ascending order.
9. /h/ Lower limit of frequency usually varies between 400 cps. and 700 cps. The peaks of intensity which occur are so marked as to suggest a multi-formant vowel. One major peak occurs around 1000 cps., one around 1700 cps. At least

signifi  
signal  
perceiv  
burst  
consis  
cited.  
with o  
stops,  
enviro  
sound,  
the sp  
were a  
identi  
of Hof  
suffic  
/g/ ap  
burst  
in dis  
and Ha  
of Ain  
effect

5 major peaks occur in each pattern; spectra for women subjects exhibit more of these peaks than for men. Upper limit of frequency is usually around 6500 cps. Intensity ranking: 8th in ascending order, but for technical reasons the data (and thus the ranking) for this item are suspect (pp. 211-213).

In the most recent study conducted on the significance of the constrictive portion of the speech signal, Ainsworth (1968) found that the phoneme /t/ was perceived on the basis of the frequency of the noise burst rather than any other factor. Again, this remains consistent with the findings of many of the other studies cited. The results on the /p/ and /k/ were also consistent with other findings cited. However, for the voiced stops, the results were not as clear, as the effect of environment seemed to show up more strongly. For the /b/ sound, the presence of the interval of voicelessness, the spread of the transitions, and the noise probabilities were all important in order for this sound to be correctly identified. This shows some agreements with the findings of Hoffman (1958), who found that the burst was not sufficient to identify the /b/. The recognition of the /g/ appeared to be based on the frequency of the noise burst near the second formant. This view appears to be in disagreement with the findings of Delattre (1955) and Harris (1958). One of the most important findings of Ainsworth's study was the implication regarding the effects of the environment in determining factors

important for perception. It was found that the vowel following the voiced consonants had an effect upon which factors were important for perception and in what way the important factors were affected. A case in point can be made from looking at the /d/. In the situation where the /d/ precedes one of the back vowels, it appeared that the formant transitions were the most important cues for perception. When the /d/ was presented in conjunction with one of the front vowels, the constrictive element of the sound was the cue which provided for perception. The above generalizations did not hold true for two vowels, however. The two vowels which showed deviation from the rule were /æ/ and /ɛ/. Ainsworth explained this digression for /æ/, citing the fact that with this particular phoneme almost no second formant transition is present and, therefore, the listener was taking cues from the third formant transition. With /ɛ/, the frequency of the burst is located very near the locus of the transition such that it is possible to perceive /d/ on the basis of either the burst or the transition. The hypothesis formulated regarding these two exceptions was not investigated.

While several people cited earlier have suggested that high frequency components of fricatives are sources of identification, it has been shown (Lawrence and Byers,

1969) that persons with high frequency hearing losses are able to identify fricatives correctly.

Studies related to the importance  
of the transition in speech  
perception

The suggestion that speech perception might be dependent in part upon the consonant-vowel transition was first made by Potter, Kopp and Green (1947) based upon the extensive spectrographical studies undertaken by these authors.

Early exploration of the question was undertaken by Coopen, et al. (1952) and Liberman, et al. (1954). They looked at the role of the transition in the perception of the English consonants /b/, /p/, /d/, /t/, /k/, and /g/. They found that the direction of the second formant allowed the listener to make certain judgments. On the basis of a rising second formant, the listener heard the sound as a /b/ or as its voiceless cognate. A falling transition caused the sound to be heard as either of the other four sounds. Further results showed that the difference between the voiced and voiceless sounds could be made on the basis of the first formant transition along with the presence or absence of the voice bar. These findings, along with those previously mentioned regarding the constrictive portions of the same sounds, yield the following criteria for selecting the plosive sound: The labial plosive is

signal

rising

energy

have a

between

of the

the co

identi

et al.

simpli

enviro

exampl

by /i/

which

found

import

impor

recor

cop. 9

was ti

spoker

signalled by the presence of a low burst of sound and a rising transition; the alveolar plosives have a high energy burst and a falling transition; the velar plosives have a characteristic low burst and falling transition.

Harris (1954, 1958) found that differentiation between the /f/ and /θ/ could be made on the basis of the second formant transition but not on the basis of the constrictive portion of the sound.

Fischer-Jørgensen (1954) felt that the identification criteria for the stops suggested by Cooper, et al. (1952) and Liberman, et al. (1954) were overly simplified because they didn't take into account the environment in which the sound occurred. They found, for example that the /t/ had a rising formant when followed by /i/, /e/, and /ɛ/.

Gerstman, et al. (1954) undertook to determine which factors were important in the transition. They found that the duration of the transition was more important than the rate at which the transition occurred.

Schatz (1954) found that the transition had an important role in the perception of the /k/. Actual tape recordings were made of a speaker saying the words keep, cop, coop, heap, hop, and hoop. Each of the consonants was then spliced away from the context in which it was spoken. Initial findings indicated that as long as the

transition was present, the isolated consonant was correctly identified but that if the transition was removed, the perception changed.

Delattre, Cooper, and Liberman (1955) used synthetic speech to investigate the locus of the transition for the voiced stops. They found that for these plosives, loci could be determined. The second formant locus for the /d/ was at 1800 Hz, for the /b/ at 720 Hz and for the /g/ at 3000 Hz. For the /g/, this occurred only when the adjoining vowel has its second formant above 1200 Hz. It also was shown that a silent period prior to the transition resulted in better perception of the sound than when the transition began right at the locus. There needed to be a silent period before the onset of voicing. The steady state level of the first formant appeared irrelevant except when there was a straight second formant transition occurring about halfway between the locus for the /g/ and the /d/. In that situation, a rising first formant caused one to perceive the /d/, whereas a falling one was perceived as /g/.

Malecot (1956) used distortion in order to determine the role of the transition in the perception of the nasal sounds. Two methods of preparing the stimulus were attempted. In the first procedure, the tape was edited using the human ear to determine the points where



cuts were

the use

transitions

method of

purposes

this method

the source

monosyllabic

Malecot

for the

stops, :

important

these cases

speech

duration

40 msec

the transition

it was

in the

for the

vowels

perception

upon

not at

the /

cuts were to be made. The human judgment was aided by the use of the oscilloscope. This method of determining transitions was found to be unreliable. The second method employed used the spectrograms for editing purposes. This method was found to be suitable. Since this method involved the use of the peak amplitude of the sound wave, it would have application only for the monosyllables said in isolation. Using this technique, Malecot determined that the transition was important for the perception of nasal sounds.

Liberman, et al. (1956) again using the voiced stops, found that the duration of the transition was an important factor in allowing for the distinction between these consonants and certain semivowels. Using the speech synthesizer, it was found that by increasing the duration of the transition for the /b/ sound beyond 40 msec., the listener perceived the stop as a /w/. When the transition for the /g/ was increased beyond 60 msec., it was perceived as a /j/. Further increase resulted in the listener perceiving a /u/ for the /b/ and an /i/ for the /g/. This was found to be true in front of the vowels /ɛ/, /i/, /e/, /a/, /ʊ/, and /o/.

Stevens and House (1958) found that the perception of the /f/ from the /θ/ was dependent upon the second formant transition. The transition did not appear important in recognizing the difference between the /s/ and /ʃ/.

Harris, et al. (1958) examined the importance of the third formant transition in the speech perception process. It was not possible to find third formant criteria for the perception of sounds, and it was found that the presence of the third formant did help to make for more accurate and consistent perception of phonemes.

O'Conner, et al. (1957) suggested that the extent and direction of the second formant transition was an important cue for distinguishing among the /w/, /j/, /r/, and /l/. This was again based on synthesis studies.

Green (1958) applied speech analysis to the study of the importance of the transition in consonant recognition. The long vowels (/i/, /aɪ/, /o/, /ə/, and /æ/) were combined with 22 English consonants to form monosyllables. The monosyllables were then presented in the following fashion: /bi:bi:bi/. Through spectrographic analysis of the consonants, Green was able to establish loci for all but four of the consonants. The loci were determined as follows:

/p/	400-	
/b/	600-	= 460- average
/m/	300, 100, 700, 650, 500-	
/w/	450, 550, 450-	= 480- "
/f/	1300-	
/v/	800, 350, 150, 700-	= 660- "
/θ/	1650, 1500, 1400-	
/ð/	1250, 1200, 1250, 1400-	= 1380- "
/t/	1600, 1600, 1700, 1850, 1850-	
/d/	1850, 2000-	= 1780- "
/n/	1850-	
/l/	1700, 1650, 1700, 1800-	= 1710- "

/s/	1800, 1850, 1750-	
/z/	1750, 1700, 1400-	= 1620- average
/r/	1100-	
/l/	2000-	= 2000- "
/ʒ/	2000-	
/j/	2300, 2500, 2400-	= 2400- "
/k/	----	
/g/	----	= ---- "
/ŋ/	----	
/h/	----	= ---- "

(p. 59)

Information on the transitions further showed that the /p/, /b/, and /m/ phonemes had negative second and third formants; the /k/, /g/, and /n/ had positive second and negative third formants; and /t/, /d/, and /n/ had both positive and negative second formants but positive third formants. It appears, then, that these sounds can be classed according to place of articulation on the basis of the transition. Time lag between the locus and the onset of the transition also appeared to be a useful cue in assessing place of articulation. It also appeared that it might be helpful in distinguishing between cognate pairs.

Nakata (1959) found that the duration of the transition was important in identification of the nasals. The transition alone was adequate to provide consistent identification between the /v/ and the /ʒ/ (Delattre, Liberman, and Cooper, 1960). Kelly, Anthony and Uldall (1963) found that the locus of the transition was the important feature for perception. Lindblom and Studdert-Kennedy (1967) found that the recognition of

monosyllables could be altered significantly by altering the rate and direction of the adjacent formant transition.

Grimm (1966) created 42 CV syllables. These were then altered by removing an increasing number of 10 msec. segments from the syllables. They found that perception was random until the segment containing the noise-non-noise portion was present.

Delattre (1969) suggested that the first formant transition may be important for the perception and recognition of sounds with regard to the manner of articulation, whereas the second formant may be important for the determination of place. This shows agreement with the analytic study conducted by Green (1958).

In a study looking at initial and final consonants, Ahmen and Patechand (1959) used 16 consonants and 8 vowels in order to make up 80 monosyllabic consonant-vowel-consonant meaningless words and 80 vowel-consonant meaningless words. A large-amplitude pulse was placed as near the start of the word as possible for the purpose of determining the beginning of the word. Various portions of these words were then gated out and a perceptual task was undertaken. The main interest of this study was the effect of duration. The duration of the sounds was determined by the aural method. The authors played the tape, and the gate was operated

manually. The duration of the gate was adjusted until the listeners agreed that only the consonant was heard without a trace of the vowel. This was called the "pure consonant." The "vowel contaminated consonant" was determined by adjusting the gate to a point where only the vowel-consonant was heard or to that point where the consonant disappeared completely. Similar procedures were employed to determine the length of the "pure vowel" and the "contaminated vowel." The duration of the "pure vowel" was found to be from 100 msec. to 460 msec. with an average duration of 300 msec. The "contaminated vowel" was found to average 440 msec. The average initial consonant length was 70 msec. They found that it was possible to remove the first 30 msec. of the fricative and plosive with no loss in intelligibility, whereas 40 msec. could be removed from the semivowels. It was further found that the portion of the consonant nearest the vowel provided more information than the initial portions. All of this tends to point to the importance of the transition in speech perception.

The results of the studies on speech perception are often confusing, and it has been shown that possibly any one of several cues may provide adequate information for the perception of certain consonants. Grimm (1966) has stated:

(1) The perception of the initial consonants was dependent upon the different elements of the syllable-- i.e., the plosive bursts, the noise between the plosive burst and the juncture of the noise and nonnoise portions, and the transition area itself. Exceptions to this were with the syllables /di/ and /bɔ/. These syllables were perceived with a high probability even when all of these elements were excised. (2) No rule could be established to show that any one of the elements of the different plosive syllables contributed uniformly to the perception of all syllables. (p. 1460)

Similar statements could be made for the fricatives.

#### STUDIES RELATING TO THE ROLE OF CONTEXT IN PERCEPTION

Fischer-Jørgensen (1954) found that the vowel which followed a consonant tended to affect whether the transition of the vowel was rising or falling.

Brown and Hildum (1956) looked at the effect of a listener's expectation in his ability to recall words. Three word lists were constructed. The first word list contained English words which were infrequent in occurrence and usage. The second list was composed of nonsense syllables whose form would be possible in the English language. The third group were nonsense syllables which would not be possible in English. They found that the mean recall for the words was 69.2 percent; for the English non-words, 46.8 percent; and for the non-English syllables, 9.7 percent. The authors concluded that the perception of speech is directed by the knowledge of the sequential probabilities.

It has further been shown that the recognition of the final consonant is improved by the knowledge of the initial consonant of a nonsense monosyllable (Ahmen and Fatehchand, 1959).

Reed and Wang (1961) found that when listeners heard non-English syllables, they rejected them in favor of the English alternative.

Shearme and Holmes (1962) studied the first and second formants in ongoing speech and isolated words. They found that in ongoing speech the formants of the vowels were located in a much more neutral position than the formants in isolated carrier words.

Swaffield, Shearme and Holmes (1961) suggested that the differences in formant transitions for conversational speech and monosyllables based on analysis of the acoustics are striking. They note two major differences. The first was that the set of vowel transitions in ongoing speech tended to be much more neutral, and the second difference was that many times the position of the transition in ongoing speech didn't even correspond to the region for the monosyllable.

Pickett and Pollock (1963) found that the intelligibility of ongoing speech depended on the duration of the excerpt listened to. They recorded conversational speech and then gated out all of the message except for a single word. They found the single



word to be unintelligible. They then allowed the gate width to increase until the initial word was perceived. For example, if the word to be heard was of, the listener first heard of, then of the, then of the world. It was found that one word was usually inappropriately recognized and the correct perception occurred only after a certain length of signal was allowed to pass.

In 1964 they replicated the study. However, the interest this time was to determine whether the results obtained in the previous study were a result of contextual cues or the increasing acoustical information available. Subjects read the sentence which they were going to hear. One word was omitted from the sentence. This word was then presented in the method previously described for the 1963 study. It was determined that even if contextual cues were available, correct perception of isolated words gated out of ongoing speech was difficult and that correct identification continued to improve to the point where three words were contained within the gated segment.

Lisker and Abramson (1967) attempted to study the time between the burst and the onset of periodicity (voice onset time, or VOT) in order to determine its effectiveness in classifying sounds as voiced and unvoiced. They used both isolated words and ongoing

1

speech and found that in isolated words the VOT time was useful in making the classification. In the ongoing speech it was found not to provide the differentiation. They hypothesized that this was due to the fact that the linguistic constraints take up any change in the information lost because of the "blurring" of ongoing speech.

Mattingly, Liberman, Syrdal, and Halives (1969) found that the perception of stop consonants depends on whether or not the cue is presented in a speech-like context or in isolation.

It has also been shown that a sound can be removed from ongoing speech and still be perceived as being there. This has been termed Phonemic Restoration (PhR). Warren and Obusek (1970, 1971) found that if the /s/ sound were removed from a word, the listeners still replaced it. They removed the /s/ out of the word legislatures in a sentence. The /s/ was replaced by either a cough, a loud buzz, a soft buzz, a period of silence, or a loud 1,000 Hz tone presented at 8 dB above peak intensity of the sentence and one presented at the peak intensity of the sentence. The listeners were asked to locate where the noise or silence occurred in the sentence. In all conditions except for the silence, PhR occurred and the inserted sound was difficult to locate. In the case of the silence, the PhR was made

less  
identi

found  
inform  
in ore  
presen  
...  
save,  
inform  
selec

sylla  
with  
nasal  
conso  
spect  
relia  
wheth  
the n  
nine  
liste  
perce  
tenda  
of th  
the

less often and location was more often correctly identified.

In a preliminary study by Sherman (1971) it was found that the listener was able to maintain acoustical information long enough to wait for subsequent context in order to perceive a message correctly. The study presented a sentence beginning, "There was time to ave . . . . ." The missing fragment could have been shave, save, wave, etc. It was found that the subsequent information allowed the listener to make the correct selection.

Ali, et al. (1971) recorded consonant-vowel syllables. Each of these syllables was then combined with either a plosive, a fricative, an affricate, or a nasal. The tapes were then processed so that the final consonant including the transition was removed using spectrograms which Malecot (1956) had found to be most reliable. The data were then analyzed to determine whether it was possible for the person still to perceive the nasal sound. It was found that five or more of the nine nasals were perceived correctly by all persons who listened. The statistical analysis showed that the perception of nasals was better than chance. Plosives tended to be perceived more often as nasals. The vowel of the preceding syllable also appeared to influence the listener's perception of a nasal.

Two tentative conclusions appear warranted based upon the studies reviewed. The first conclusion is that the presentation of phonemes in a contextual environment appears to blur some of the acoustic elements which are present when the phoneme is produced either in isolation or in syllables. The second factor appears to be that the listener to a language with which he is acquainted applies constraints to what he hears based upon his knowledge of the language system employed. This constraint causes him to select certain choices even though acoustically the choice may be unclear.

#### STUDIES RELATING TO TIME COMPRESSED AND INTERRUPTED SPEECH

The question of how much speech could be removed and still not cause a significant loss in correct perception has been the subject of considerable interest. Perhaps one of the earliest studies dealing with interrupted speech was carried out by Miller and Licklider (1950). They considered three factors in looking at the intelligibility of ongoing speech. Factors considered were the number of interruptions per second, the proportion of time the speech was on, and the regularity of the interruptions. The PB word lists of Egan were employed. They found that it was possible to remove up to 50 percent of the signal and still have

better than 90 percent intelligibility. It was further found that as long as the interruption occurred more than 10 times per second, intelligibility was not impaired. The regularity of the interruption was not a factor in the study.

As a result of this study by Miller and Licklider, the interest was aroused as to what might happen if the speech segments were removed rather than just having periods of silence. For this reason, studies in time compression began.

Two methods of speech compression have been used. The earliest method was that of speeding up the speech. The message was recorded at a certain rate and then played back at a much more rapid rate. This method has been employed by several researchers (Fletcher, 1929; Foulke, 1966; Garvey, 1953; Klumpp and Webster, 1961; Kurtzrock, 1957, and McLain, 1962). The second method is the sampling method. This method entails removing portions of speech from the signal. The first method used was that of Miller and Licklider (1950); they used a switching arrangement which permitted a recorded speech signal to be turned off periodically during the playback of the recording. Garvey (1953) manually removed portions of the tape and then spliced the tape together, thus resulting in time compression. Fairbanks, et al. (1954) devised an electro-mechanical machine for removing speech. The machine

reproduced periodic segments of the speech signal while discarding other segments. Computers also have recently been used to sample the speech signal (Scott, 1965).

Several factors have emerged as being important in the perception of time compressed speech. One of the factors is the method of compression used.

Klumpp and Webster (1962) used the speed changing method of compression and found that when this method was employed, the intelligibility dropped by 40 percent or more with only a 33 percent savings in time. This agrees with findings by Garvey (1953), Kurtzrock (1957), and Foulke (1966).

Using the manual sampling method, words can withstand considerably more compression and still be intelligible. With the speech compressed by 60 percent, the intelligibility still remained approximately 90 percent (Garvey, 1953). However, the intelligibility dropped to 50 percent when the compression was increased to 75 percent.

Using the electromechanical method appears to give results similar to the sampling-by-hand method. Fairbanks, Guttman and Miron (1957) performed a study in which the message was compressed at five different compression levels. The material was factual information about which it was hoped foreknowledge would be small. At 50 percent compression, the reduction of message effectiveness was small. The comprehension was still





about 90 percent of that for subjects who heard the original version of 141 words per minute. At 60 percent compression, however, the perception of the speech was only at 50 percent of maximum. A sharp decline in accurate perception occurred between 50 percent and 60 percent compression. Again, these results showed good agreement with the findings of Kurtzrock (1967), Fairbanks and Kodman (1957), and Dukelskiy (1967).

Another factor which appears to be important in the perception of the compressed speech is the duration of the discarded portion. For example, it is possible to get 50 percent time compression by discarding 10 msec. of a 20 msec. portion of speech or by discarding a 30 msec. portion of a 60 msec. segment. As reported earlier, Miller and Licklider (1950) found that as long as ten samplings per second occurred, there was not any loss of intelligibility.

Garvey (1953) used discard intervals of 40, 60, 80, and 100 msec. intervals. The resultant intelligibility scores were 95, 96, 95, and 86 percent. Fairbanks and Kodman (1957) also investigated this question. Their results indicated that if intervals as large as .16 or .24 seconds were discarded, intelligibility would never reach 100 percent. They found a considerable loss of intelligibility when the compression sample removed was greater than 80 msec.

The frequency and duration of sampling and discarding also appears to make a difference. Fairbanks and Kodman (1957) found that if the portion removed was less than 10 msec., intelligibility was also decreased.

Foulke (1966) has suggested that masking may occur if sampling occurs with sufficient frequency. He illustrates this by presenting the following example. He found that if sampling occurred every 2.5 msec., a 400 Hz tone was generated. This then could serve to mask the signal.

It was further found that the length of the utterance made a difference in the perception. Henry (1966) found that the more the number of phonemes, the greater the intelligibility. He used words with from 3 to 9 phonemes and found a continued improvement as the number increased. Klumpp and Webster (1961) found that short phrases were more intelligible than single words.

In comparing the results of studies on time compression with the study by Miller and Licklider (1950) on interrupted speech, Garvey (1953) found that with up to 50 percent time compression or interruption, the results were similar. However, when 62 percent of the signal was discarded, the interrupted technique yielded 40 percent better results than the time compression.

Ahmend and Fatechand (1959) in their study using clipped speech made the following observations. They found

that where the interruption was made had an effect on the perception. If the initial portion of the consonant were allowed to pass unaltered, the next segment of the tape then gated out or suppressed, and then the remainder of the consonant allowed to pass, the recognition for the plosives, fricatives, and semivowels was 52, 60, and 93 percent respectively. If, however, the initial segment was suppressed, the perception for the consonant-vowel-consonant nonsense words was 36, 60 and 80 percent.

The results of time compression study have shown that up to 60 percent of the signal can be removed with only slight deterioration in the listener's ability to perceive the message. Frequency of removal and the duration of removal also appear to be a factor. It was also shown that the more contextual cues that were available, the more one was able to compress speech without loss of information. A further important suggestion which has been made is that the place of the interruption was significant.

## CHAPTER III

### EQUIPMENT, MATERIALS, SUBJECTS, AND PROCEDURES

The following sections contain a detailed description of the subjects, experimental materials, and procedures used in this study.

#### EQUIPMENT

The following equipment was used during the various phases of the study:

- Tape Recorder I (Ampex AG 440 - recording)
- Tape Recorder II (Ampex AG 600 - dubbing)
- Tape Recorder III (Ampex AG 500 - playback)
- Microphone (635A Electrovoice)
- Level Recorder (Bruel and Kjaer 2305)
- Commercial Test Room (Industrial Acoustic Company, Inc., Single-walled booth, series 402A)
- Magnetic Recording Tapes (type 201, Scotch Brand)
- Permanent Magnet (Alnico V 3/16" diameter x 1" long)
- Sound Spectrograph Model PV10A (Voiceprint Laboratory)
- Sound Level Meter (B & K 2204S)
- Earphones (TDH 39 300 Ohms)
- Filter (B & K 1613)
- Measurement Amplifier (Hewlett-Packard 450 A, 20 dB gain)

#### MATERIALS

Six sentences were constructed to meet the following criteria:

5

- a. At least seventy-five percent of the consonants contained in each of the sentences were either stops or fricatives. The remaining consonants were nasals, affricates, glides, and semi-vowels. This criterion was established to allow for some comparisons between the current study and the previous literature written dealing with perception of consonants in words and syllables.
- b. All of the sentences used in the study were five words in length. Miller (1956) suggests that the human brain is able to store up to seven items for recall with little difficulty. Because this study was not to be a test of short term memory but rather one of perception the use of five words was felt to be well within the recall limits of the subjects.
- c. All of the sentences were of the simple, declarative type in order to not introduce linguistic complexity as a variable.
- d. All words in the sentences were mono-syllables selected from the mono-syllable word lists of Moser (1969).

Using these criteria, the following sentences were constructed for use in this study:

1. Big dogs ate the food.
2. The boys bought five fish.
3. Cups should go up there.
4. She gave us six pigs.
5. His wife baked this cake.
6. You pushed the door shut.

These 30 words then served as the basis for constructing 6 word lists of 5 words each. The word lists were randomly constructed and are as follows:

1. cups six go the boys
2. five cake their fish you've
3. gave baked the door bought
4. dogs this she the food
5. pushed ate should pigs big
6. shut his up wife us

#### PREPARATION OF MATERIALS

The master tape was recorded in the Speech Science Laboratory at Michigan State University by one male speaker who spoke with a General American Dialect. The speaker is a member of the faculty in the Department of Audiology and Speech Sciences at Michigan State University and has had extensive phonetic training as well as voice and diction training. The stimulus material was recorded using the Ampex AG 400 tape recorder employing an Electrovoice 635A microphone in a single-walled sound-treated booth. Each of the sentences was



recorded as naturally as possible to assure that stress and intonation patterns remained normal. The sentences were read at the rate of about two and one-half words per second. This rate is at approximately the lower limits for normal reading. Care was taken in reading the sentences to assure that each of the phonemes was articulated. Following the taping of the sentences, the sound spectrograph PV10A (Voiceprint Laboratory) was employed to obtain spectrograms of each of the sentences as a check to make certain that each of the consonants was visually present. In one of the sentences all of the phonemes were not present and so the sentence was re-recorded and again visually displayed. The second recording had all of the phonemes present and was, therefore, used. The master tape was then played to three listeners who had some training in speech pathology, and these listeners were asked whether they detected anything unnatural about the sentences. One stated that the speech was slightly exaggerated. All of them felt that rate, stress, and intonation were normal.

This master tape provided the stimulus material for the two additional tapes. Each of the additional tapes, herein referred to as sub-master tapes 1 and 2, was dubbed using the Ampex AG 440 in conjunction with the Ampex AG 600. Each of the sub-masters was then placed on the sound spectrograph. Use of the spectrograph

follows the suggestion of Malecot (1956). Spectrograms were made of each of the sentences. Each spectrogram was then removed from the recording drum of the Voiceprinter and the beginning of the consonants was marked on the spectrogram. Criteria used for establishing this point are as follows:

Voiceless stops. The VC boundary was taken to be the point where the last trace of the formant transition of the initial vowel was visible in the sound spectrograms. The CV boundary was located at the first transient spike of the release.

Voiced stops. The VC boundary was identified with the point where the F1 first reached the low frequency of the so-called voicing bar of the closure segment. The boundary was put just before the voice pulse in which the higher formants reappeared in the segment after the closure segment.

Voiceless Fricatives. The friction noise of all stimulus words containing these consonants was preceded and followed by brief aspirative segment, the middle of which was taken as the VC and CV boundaries.

Nasals. The segments representing the nasal murmur of the nasal consonants were easily delimited from the adjacent nasalized vowel segments by observing the very different spectrum characteristics of these two segment types. The boundaries were put where the spectrum change was most abrupt.

Sonorants. The segmentation of the intervocalic /l/ presented no problems since the lateralized closure of this consonant is associated with a very rapid change of spectrum character at the VC and CV boundaries (Ohman, 1966, p. 984).

These provided a general guide for determining the point to be marked. Potter, Kopp and Green's (1947) information was used to aid in determining the steady state portion of the vowels as well as the spike and

fill (constrictive) portions of the consonants. Areas between the steady state portion of the vowel and the constrictive element of the consonant were considered to be the transition. Spectrograms were made of the sentences and steady state vowel, transitions and constrictives were marked by the experimenter on the spectrograms. A second person, an expert and recognized authority in spectrography, was then asked to examine the spectrograms and agreement was made regarding the segmentation of the spectrograms. Only in two cases was there a discrepancy, and in these cases an agreement was reached as to where the removal should be made. In each case the discrepancy was a minor difference.

Once the segmentation of the sentences had been determined, the tapes were again placed on the voice-printer and spectrograms made. For locating the onset of the consonant constrictions the spectrograms were left in place on the recording drum and the recording drum of the voiceprinter was rotated until the printing stylus was in direct alignment with the mark on the spectrogram indicating the onset of the consonant.

Because the rotation of the recording drum and the playback head of the scanning drum are synchronized, it was possible to find the onset of the constriction portion of the consonant on the tape using this synchronization.

By locating the revolving playback head beneath the portion of the tape which was looped about the scanning drum of the voiceprinter, the location of the onset of the consonant on the magnetic tape was located. The same procedure was used for determining the onset of the transition, the onset of the steady state portion of the signal, and the onset of the vowel-consonant transition. These points were marked on the back of the tape with a felt tip pen.

Removal of the segments of the tape was accomplished by passing a permanent magnet (Alnico V which had a  $3/16$ " diameter) between the lines marked on the magnetic tape, thus erasing the portions of the tape contained between those lines.

To check that the correct portion of the speech signal had been removed, a second spectrogram was made upon the completion of the deletion. The two spectrograms were then placed upon each other and the portion which was removed was compared to the portion which should have been removed to determine that the correct portion of the sounds had been removed.

Sub-master 1 was altered initially so that all of the consonant constrictions contained in each of the six sentences were removed. This left only the transitional section of the consonants and the steady state portion of the vowels remaining. Sub-master 1 was

then dubbed twice yielding experimental tape 1 and 2. Experimental tape 1 was left unaltered at this point. Experimental tape 2 was spliced so that each word was contained in a separate segment of spliced tape. Words were then randomly assigned to six word lists of five words each and the tape was respliced. Experimental tape 1 was a meaningful sentence with consonant constrictives removed. Experimental tape 2 was first order approximation sentences with consonant constrictives removed.

Sub-master 1 was then further altered by removing the transitional element, leaving only the steady state portion of the vowel remaining. Two experimental tapes were then dubbed from sub-master 1, yielding experimental tapes 3 and 4. Experimental tape 3 was left and was a meaningful sentence with the consonant constriction and transition removed, leaving only the steady state vowel. Experimental tape 4 was spliced and the words reordered as in experimental tape 2, resulting in a first order approximation sentence with consonant constriction and transition removed.

Sub-master 2 was altered by initially removing the transitional segment of all of the words, leaving the consonant constrictives and steady state portions of the vowel. Two experimental tapes were then dubbed. Experimental tape 5 was a meaningful sentence with transition removed. Experimental tape 6 was spliced

and the words reordered, resulting in a first order approximation with transition removed.

Sub-master 2 was then further altered by removing the steady state portion of the vowel, leaving only the consonant constriction. Experimental tape 7 and 8 were then dubbed from sub-master 2. Experimental tape 7 was a meaningful sentence containing only consonant constrictions with the transition and steady state portion of the vowel having been deleted. Experimental tape 8 had the same segmentation effect as experimental tape 7 but was rearranged in first order approximation. Figure 4 shows the procedure involved in developing the eight experimental tapes. Table 1 lists the tapes and explains which portions were removed, the order of presentation and the portions which the subjects actually heard.

#### SUBJECTS

The 80 subjects used in this study were all adults ranging in age between 19 and 30 years old. The mean age of the subjects was 23.8 years and the median age was 24. There were 20 male and 60 female subjects. All of the subjects were residents of East Lansing and Lansing, Michigan, and all of the subjects had completed high school and had some college education. All were native English speaking and spoke general American English. None of the subjects had clinically significant hearing

1

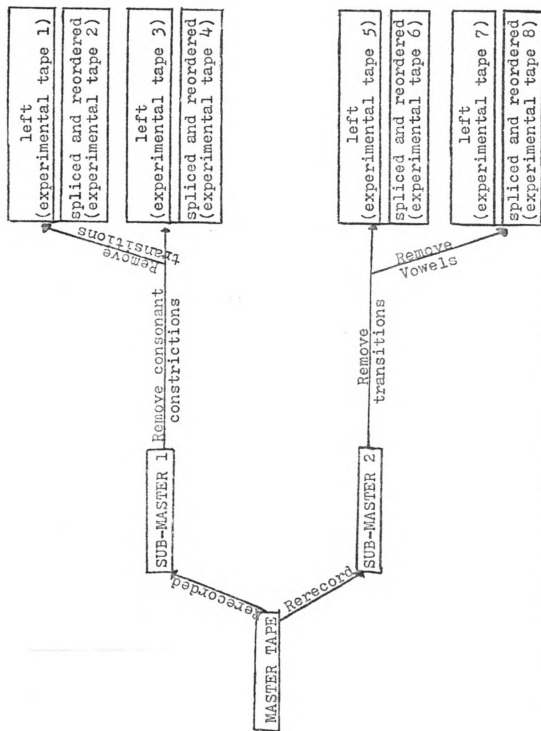


Figure 4.--Diagram showing preparation of tapes to form the eight experimental tapes used in this study





TABLE 1.--Contents of the experimental tapes

Experi- mental Tape	Order of Presenta- tion	Portions Removed	Portions Subjects Heard
1	Meaningful Sentence	Consonant Constrictions	Transitions and Steady State Vowels
2	First Order Approximation	Consonant Constrictions	Transitions and Steady State Vowels
3	Meaningful Sentence	Consonant Constrictions and Transit- ions	Steady State Vowels
4	First Order Approximation	Consonant Constrictions and Transit- ions	Steady State Vowels
5	Meaningful Sentence	Transitions	Consonant Constrictions and Steady State Vowels
6	First Order Approximation	Transitions	Consonant Constrictions and Steady State Vowels
7	Meaningful Sentence	Steady State Vowel and Transitions	Consonant Constrictions
8	First Order Approximation	Steady State Vowel and Transitions	Consonant Constrictions

losses as determined by a pure tone screening test administered at 20 dB ISO. The subjects were tested just prior to their participation in the present study. The three listeners used to validate the master tape were subjected to the same criteria as the experimental subjects. Each of the subjects used in the study heard only one of the eight tapes and was not aware at the time of testing specifically how the tape had been altered. Assignment to groups was done randomly.

#### PROCEDURES

The tapes were played to each of the subjects. Each group of subjects heard one of the four taped conditions. Each of the subjects was seated in a room. The tape was then presented at 70 dB SPL to the right ear of each of the listeners. The calibration of the tape for presentation was accomplished by averaging the peak rms value of the tape from the graphic level recording. A 1000 Hz puretone calibration tone was then placed at the beginning of the tape and was used for adjusting the volume of presentation. The tape was played on the Ampex AG500 with output to a Dynaco 60-watt amplifier and was then fed to 12 sets of TDH 39 300 Ohms telephonic headphones.

The following instructions were given to the subjects who heard the meaningful sentences:

You are going to hear six sentences of five words each. Each sentence will be preceded by the phrase, "Sentence number . . . ." Portions of the sentence have been removed from the tape recording. Your task is to write the sentence you hear in the appropriate space on your answer sheet. For example, sentence number one will be written in space number one. After the completion of the sentence you will have fifteen seconds. If you are uncertain of what you have heard, you are urged to make the best response possible, that is, leave as few blanks as possible. Record your response only after hearing the entire sentence. Are there any questions?

If there were any questions, they were answered.

The only question raised was regarding guessing. To answer this question, the portion of the original statement about uncertainty was reread.

For the subjects who heard the first order approximations, the following directions were given:

You are going to hear six lists of five words each. Each list will be preceded by the phrase, "List number . . . ." Portions of the list have been removed from the tape recording. Your task is to write the list you hear in the appropriate space on your answer sheet. For example, list number one will be written in space number one. After the completion of the list you will have fifteen seconds. If you are uncertain of what you have heard, you are urged to make the best response possible, that is, leave as few blanks as possible. Record your response only after hearing the entire list. Are there any questions?

The tape was then played and the listeners recorded their responses.

#### ANALYSIS OF THE DATA

The total number of words correctly identified by each listener was tabulated and an initial preview

of the data was conducted by using a two-way analysis of variance. It was determined that the F Distribution or F- ratio would be used to determine whether the null hypotheses could be rejected. This test was accepted because it analyzes the significance of the difference between variances. A critical value at the .01 level of confidence was accepted as the level that must be reached in order to reject the null hypothesis. If significance was reached, a Sheffe post hoc comparison would then be employed to determine where the significant difference occurred.

## CHAPTER IV

### RESULTS AND DISCUSSION

The purpose of this study was to determine the effects which context had in the recognition of meaningful and non-meaningful sentences. A further factor considered in the study was the importance of the transition and the constriction in the recognition of words spoken in context but heard both in context and out of it. To aid in guiding the study, the following questions were formulated:

1. Does the use of context provide cues for the perception of words which have been altered by the removal of various acoustic segments?
2. How much intelligibility is maintained because of co-articulation when only a vowel sound segmented from a word is heard both presented in context and out of context?
3. Is there a significant difference in the perception of contextual speech when transitions are removed and when constrictive elements are removed?
4. If significant differences occur in ongoing contextual speech, do the same differences also appear in non-contextual sentences?

The study was conducted using 80 normal hearing adults. The subjects were divided into eight groups of ten, and each group heard one of the taped conditions.

Scores for each person were computed based upon the number of words which he was able to perceive correctly.

The results of the study appear to provide, within the confines of the experimental procedure, significant information relative to the perception of language and the phoneme.

#### RESULTS OF COMPARING CONDITIONS

Initially, means and standard deviations for each of the taped conditions were computed. The means were based on the average number of words correctly identified by the ten people who heard each of the conditions. Table 2 on the following page presents the means and standard deviations for each of the eight conditions.

Table 2 shows that the range of standard deviations for the eight conditions is  $0.2 < s.d < 3.7$ , being the lower end for experimental tape 4 and the upper end for experimental tape 2. The means for the eight conditions ranged  $0.6 < \text{mean} < 26.1$  experimental tape 4 to experimental tape 5 respectively. The F distribution, or F ratio, was used to further analyze the significance of the differences in variance of the scores obtained. Table 3 summarizes the results of the analysis. The differences for the columns (order effect), the differences for rows (segmentation effect) and the interaction effect were all significant at the 0.01 level

TABLE 2.--Group means and standard deviations of ten subjects for each of eight test conditions ranked in descending order

Experimental Tape	Mean (in number of words)	Standard Deviation
5	26.1	2.6
1	19.2	3.1
6	11.2	3.4
2	9.0	3.7
7	7.2	2.3
3	3.6	2.2
8	2.5	1.5
4	0.6	0.2



TABLE 3.--Analysis of variance

Source	SS	df	MS	F
Rows (order of presentation)	1,371.2	1	1,371.2	133.1*
Columns (segmentation condition)	3,637.2	3	1,212.4	117.7*
Interaction	414.5	3	138.2	13.4*
Error (within cell)	735	72	10.3	
Totals	6,157.9	79		

\* = significance at 0.01 level

of confidence. An  $F$  of 2.09 was needed for significance at this level. The magnitude of the segmentation and the order effect allow us to conclude with considerable certainty that both of these factors are significant in speech perception. The  $F$  for segmentation was 117.7 and for order, 113.3. The  $F$  for interaction was only 13.4.

In order to examine the interaction effect which was also found to be significant, Figure 5 presents a visual display of the results obtained in the study relative to the interaction effect. As can be seen in this figure, the interaction which occurs between the segmentation condition and the order effect is that context seems to become less significant as more and more of the signal is removed. It might be hypothesized that as more and more information is removed from the contextual sentences through the segmentation process, the contextual cues which had aided in understanding the meaningful sentences are lost and therefore, that the performance which occurs parallels more closely the performance of the non-contextual or first order approximation sentences.

As it was a concern of this study to know which of the individual comparisons contributed to the overall significance, a post hoc comparison according to Scheffe (Hays, 1963) was employed. The Scheffe comparison is suitable for this analysis because it allows one to

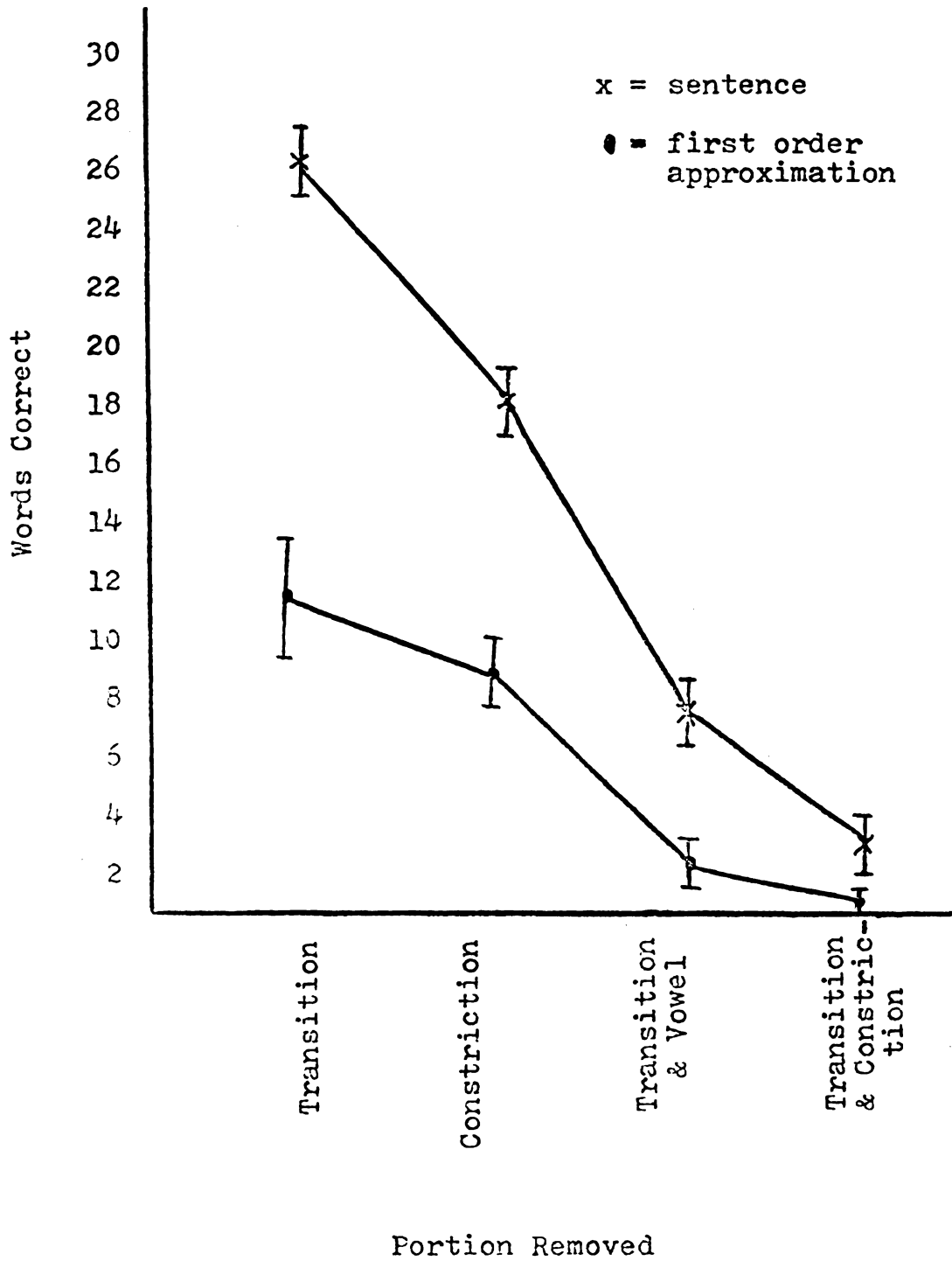


FIGURE 5.--Standard deviations and mean number of words correctly identified based on ten listeners under each condition

look at all possible pairwise comparisons. The Scheffe analysis resulted in the following confidence interval:

$$\hat{\psi}_g - 6.8 \leq \psi_g \leq \hat{\psi}_g + 6.8$$

To test for significance, Table 4 was constructed to compare each of the eight conditions with every other condition. In order for a comparison to be significant, the difference as shown on the table must be outside the limits of the confidence interval.

### Effects of Context

The importance of context in perception of the spoken word can be seen by looking at the number of times each word was correctly identified when placed in a meaningful context compared to the number of times the same word was correctly identified in a first order approximation sentence. Table 5 shows the number of times each word was identified correctly under each of the eight presentation conditions.

As can be seen by examining Table 5, the context has a general effect across all words. Particular attention will be paid to experimental tape 1 (meaningful sentence minus the consonant constrictions), experimental tape 2 (first order approximation minus the consonant constrictions), experimental tape 5 (meaningful sentence minus the transitions), and experimental tape 6 (first order approximation minus the transitions). Contrasting experimental tape 1 with experimental tape 2, the words

TABLE 4.--Differences in averages contrasting each of the conditions with all other conditions

	Experimental Tape							
	Mean	2	5	6	3	4	7	8
Mean		9.0	26.1	11.2	3.6	.6	7.2	2.5
Experimental Tape								
1	19.2	10.2*	-6.9*	8.0*	15.6*	18.6*	12.0*	16.7*
2	9.0		-17.1*	-2.2	5.4	8.4*	1.8	6.5
5	26.1			14.9*	22.5*	25.5*	18.9*	23.6*
6	11.2				7.6*	10.6*	4.0	8.7*
3	3.6					3.0	3.6	1.1
4	.6						-6.6	1.9
7	7.2							4.7

\*Significant at the 0.01 level

TABLE 5.--Number of correct identifications for each of  
the 30 words under each of the  
experimental conditions

Word	Experimental Tape							
	1*	2	3*	4	5*	6	7*	8
big	6	8	1	--	8	--	1	--
dogs	1	--	--	--	9	6	1	--
ate	10	8	7	2	10	8	1	--
the	30	22	16	4	30	23	22	12
food	7	8	2	--	8	6		2
boys	8	10	--	--	10	10	6	3
bought	10	9	--	--	9	6	2	--
five	10	10	1	--	10	7	1	--
fish	3	10	1	--	10	4	1	--
cups	1	--	--	--	6	--	--	--
should	3	--	--	--	8	2	--	--
go	10	7	1	--	9	--	--	--
up	10	8	4	--	10	8	--	--
there	10	9	--	--	8	--	--	--
she	--	--	--	--	10	2	5	2
gave	6	--	--	--	9	--	1	--
us	6	2	--	--	8	--		--
six	6	--	--	--	10	--	--	--
pigs	6	--	1	--	6	1	1	--
his	9	--	--	--	7	2	--	--
wife	9	--	--	--	7	2	--	--
baked	7	2	--	--	10	5	1	--
this	--	--	--	--	7	--	--	1
cake	8	--	1	--	10	--	--	--
you've	2	--	--	--	4	--	--	--
pushed	2	--	--	--	8	--	8	--
door	6	3	1	--	10	10	10	4
shut	6	--	--	--	10	7	9	1

\*These are the contextual conditions

in context were consistently recognized more often. In the case of 20 of the 30 words, the words in the first order approximation sentence were recognized only half as often as the same words in the sentential presentation. Only in one case did experimental tape 2 exceed experimental tape 1 and that was for the word food. In this case, the word was recognized seven times in context but eight times in the first order approximation sentence. In one further case the word was not correctly recognized in either condition. This was the word she. In the sentential order this was perceived as he. Here the linguistic constraints known to an English speaker of that sentence would have allowed for the use of either word.

In comparing experimental tape 5 and experimental tape 6, the same results were obtained as for experimental tape 1 and experimental tape 2. Words in experimental tape 5 were correctly identified more frequently than the same words in the first order approximation sentence used in experimental tape 6 for all but two cases. In these two cases the words boys and door were both correctly identified by all of the subjects who heard these conditions. In 21 of 30 words presented, experimental tape 5 words were recognized at least twice as frequently as experimental tape 6 words.

Further support can be found for the influence of linguistic constraint by looking at the influence of

word position. Soderberg (1967) looked at the information values of words in clauses of varying length. The information value was determined by the extent to which college students could predict what that word was to be. Table 6 presents Soderberg's results. Of particular interest to this study are the findings for five-word clauses.

TABLE 6.--Soderberg's information values of words in clauses classified by length

No. of Words in clause	1	2	3	4	5	6	7	8	9	10	Totals
2	70	65									135
3	75	56	51								182
4	71	79	40	23							213
5	59	69	67	64	46						305
6	76	56	52	77	52	41					354
7	76	44	71	77	35	43	61				407
8	82	77	74	46	77	53	66	39			514
9	61	56	62	75	50	43	62	35	64		508
10	97	54	68	17	62	15	50	85	59	10	517
Totals	667	556	485	379	322	195	239	159	123	10	3,135

In this current study, using sentence position as a variable, the results in Table 7 were obtained.

It can be seen that the score for experimental tape 5 and experimental tape 1 both follow the general distribution of Soderberg's findings with positions one and five being the poorest and the medial positions being better. The first order approximations, however, do not follow this order. In neither case is position one the poorest, and in both cases the last position is the



TABLE 7.--Total number of correct responses for each of the five sentence positions

Experimental Tape	Word Position				
	1	2	3	4	5
5	49	51	56	57	52
1	28	29	53	42	40
6	20	15	20	32	22
2	10	10	31	17	23

second best. It would appear that in the case of the sentential ordering, the recognition of words is in some way related to or parallels the information value of 5-word clauses as found by Soderberg. It would appear that subjects were using some of the predictive value of linguistic constraint just as Soderberg's college students did. In the first order approximation sentences, however, subjects had no linguistic information so that correct responses were based upon the acoustic cues alone and word position was no longer a factor in whether a word was correctly perceived. An example of linguistic context affecting perception is shown to demonstrate this. In sentence one, Big dogs ate the food, food was perceived as bone. The /b/ for /f/ and /n/ for /d/ were both rare confusions in Miller and Nicely's (1955) study. All of this would again support the fact that much of the information of ongoing speech is determined by linguistic

constraints as opposed to the acoustic constraints which function in the first order approximations.

### Effects of segmentation

A comparison of the effects of temporal segmentation was undertaken using Table 4 (page 74). First of all, if the effects of segmentation upon the contextual sentences are compared, it is found that those contextual sentences in which only the transition had been removed were superior to all of the other segmentation conditions. This significant difference between the condition in which only the transition had been removed and in which only the consonant constriction had been removed will warrant further consideration. It can also be noted that the condition in which only the consonant constriction had been removed, while being inferior to the experimental tape 5 conditions (without transitions), was superior to the other two conditions, experimental tape 3 (without constriction and transition) and experimental tape 7 (without transition and vowel). Between the experimental conditions on experimental tape 7 and experimental tape 3 there were no significant differences. Among first order approximation sentences no significant differences were noted between experimental tape 6 (without transitions) and experimental tape 2 (without consonant constrictions), nor between experimental tape 8 (without vowels and

transitions) and experimental tape 4 (without constrictions and transitions). However, there were significant differences between experimental tape 6 (first order approximation without transitions) and experimental tape 8 (first order approximation without vowels and transitions), between experimental tape 6 and experimental tape 4 (first order approximation without consonant constrictions and transitions), between experimental tape 2 (first order approximation without consonant constrictions) and experimental tape 8, and also between experimental tape 2 and experimental tape 4.

Before a more indepth comparison is attempted between experimental conditions in tape 5 and tape 1, which appear of particular importance in this study, some general comments are deemed necessary.

The first question which will be raised at this point is the question of what factor might account for the differences which were found in the segmentation effect. In terms of the literature dealing with time compression, the simplest explanation which could be made would deal with the amount of the signal which was removed. In accounting for the differences between conditions in which only one portion of the speech signal was removed compared to those in which two portions were removed, it is clear that in these comparisons approximately twice as much of the signal was removed

when two of the speech elements (either transition, constriction, or vowel) was removed as when only one was deleted. The significant differences between experimental tape 1 as compared to 3 or 7 and experimental tape 5 as compared to 3 or 7 are most easily explained on the basis of this time removal factor. The fact that no differences occur between conditions on tapes 3 and 7 can be explained in the same way. Since approximately the same amount of information was removed, the results would be expected to be the same. However, when one compares the meaningful sentences with the transitions removed with those in which the constrictions were removed, one finds again approximately the same amount of speech signal was removed. This was determined by measuring the amount of the spectrogram removed under each condition. In comparing the amounts deleted, one finds that slightly more was removed when all of the transitions were removed and, yet, this condition was superior to the condition in which the constriction was removed. The findings here that suppression of the initial part of the word is more detrimental to the speech signal is in agreement with the findings of Ahmand and Fatechand (1959) who found that with the transitions removed perception was at 52-93 percent and with constrictions removed it was at 36-80 percent. One possible explanation of why this is so might be that the removal of the initial portion of words makes it difficult for the listener to segment the

acoustical signal into individual words. While no evidence of this was seen in the contextual situation, several examples of this were seen in the non-contextual or first order approximation sentences where subjects tended to run words together, hearing she pig as sheep egg.

#### Amount of information contained in vowels

In the present study it was found that in ongoing speech as well as in word lists, almost no information is carried in the vowel itself. If one looks at the findings in Table 5, it is seen that in the meaningful sentence condition where only the vowel was remaining (experimental tape 3) subjects were only able to give 36 correct responses. Even in these figures some doubt exists as to the correct responses given. Sixteen of the 36 correct responses were for the word the. However, during this particular tape, subjects reported hearing 42 the's so that it is possible that subjects, knowing they were to hear sentences simply guessed that many of the sentences would have the in them and therefore placed them in. The other word which was heard frequently was ate which was correct seven times. Here again, while ate only occurred one time in the six sentences, the subjects guessed this an average of 2.4 times each or a total of 24 times. The fact that ate occurs so commonly in English may have led the subjects to guess ate any time they heard /e/ plus an apparent plosive. The fact

remains, however, that even under the most favorable interpretation of these data the subjects who heard only the vowels in meaningful sentences were only able to identify an average of 1.2 words or a recognition of 4 percent.

In the non-contextual sentences similar results were obtained. The and ate accounted for all of the correct responses; but again, both words were guessed a disproportionate number of times. The results of this study in the area of vowels appears clear: Vowels carry very little meaning when said alone.

#### Perceptual differences in ongoing contextual and non-contextual speech

Since it appears that the major differences which were significant occurred in the case of the meaningful sentences in which the transition had been removed and the meaningful sentences in which the constrictions had been removed, a confusion matrices of errors for these conditions was constructed. The first order approximations of these two conditions were also included so that some further comparisons of perceptual confusion might be undertaken. Because of the limited number of times each phoneme was used, it was felt that an error matrix which would allow for grouping the phonemes would be more helpful. Therefore, the phonemes were grouped according to three major categories. In order to allow for some comparisons with Miller and Nicely's (1955) study, the same taxonomy system was

employed although two of the smaller classifications were deleted. The phonemes were grouped according to the following categories:

1. Voicing.--Phonemes in which the vocal cords are in vibration were contrasted against those in which they do not vibrate. The voiced consonants in this study were the /b/, /d/, /g/, /v/, /z/, /m/, and /n/. The voiceless ones against which they were contrasted were the /p/, /t/, /k/, /f/, /s/, and /h/.
2. Affrication.--If the flow of air is halted completely so that the consonant is a stop or nasal or if a turbulence is only forced, resulting in a friction noise, we get the distinctions between /p/, /t/, /k/, /b/, /d/, /g/, /m/, /n/ and /f/, /s/, /v/, and /z/.
3. Place of articulation.--The sounds were also compared for errors based upon the place in the mouth in which either the plosion or the friction took place. For this category, three places of articulation were considered: Those consonants which are produced in the front of the mouth, /p/, /b/, /f/, /v/, and /m/; those which are produced in the middle of the mouth /t/, /d/, /s/, /z/, and /n/; and those which are produced at the back of the oral cavity /k/, /g/, and /h/.





Table 8 shows the classification of the consonants used to construct the confusion matrices.

As no errors were noted in duration which Miller and Nicely used to distinguish the /θ/ from the /s/ and the /ʃ/ from the /z/, this classification was deleted from the study. The number of errors of each type was then tabulated. Each time a consonant was used in place of another consonant, the error was analyzed and counted. For example, if a /p/ was heard as a /t/, this was considered an error in place of articulation. If a /p/ was heard as an /f/, this was considered as an error in affrication. If a /p/ was heard as a /b/, this was considered an error in voicing. If a /t/ was heard as a /v/, this was considered as an error in place, affrication and voicing. The number of errors made in each category are shown in Table 9. Several omissions of words occurred, and these could not be taken into account in the error matrices. Figure 6 shows a graphic presentation of this material. Miller and Nicely (1955) found that place was most often incorrectly perceived, with affrication being next and voicing most often correctly perceived. When one looks at the data for first order approximations in this current study, one finds that the same relationship holds true. However, if one looks at the information that is available on the contextual sentences which were altered, it appears that there is a

TABLE 8.--Classification of the consonants used to  
construct the confusion matrices

Consonant	Voicing	Affrication	Place
p	0	0	0
t	0	0	1
k	0	0	2
f	0	1	0
θ	0	1	1
s	0	1	1
ʃ	0	1	2
b	1	0	0
d	1	0	1
g	1	0	2
v	1	1	0
ð	1	1	1
z	1	1	1
h	0	1	2
m	1	2	0
n	1	2	1

Voicing:	0 = unvoiced	Place:	0 = front
	1 = voiced		1 = middle
Affrication:	0 = plosive		2 = back
	1 = fricative		
	2 = nasal		

TABLE 9.--A confusion matrix for meaningful sentences with transitions omitted or constrictions omitted and first order approximations with transitions omitted or constrictions omitted

	Meaningful Sentences		First Order Approximations	
	Transi- tions Omitted	Constric- tions Omitted	Transi- tions Omitted	Constric- tions Omitted
Voicing	3	12	1	7
Affrication	10	37	22	23
Place	10	36	23	42

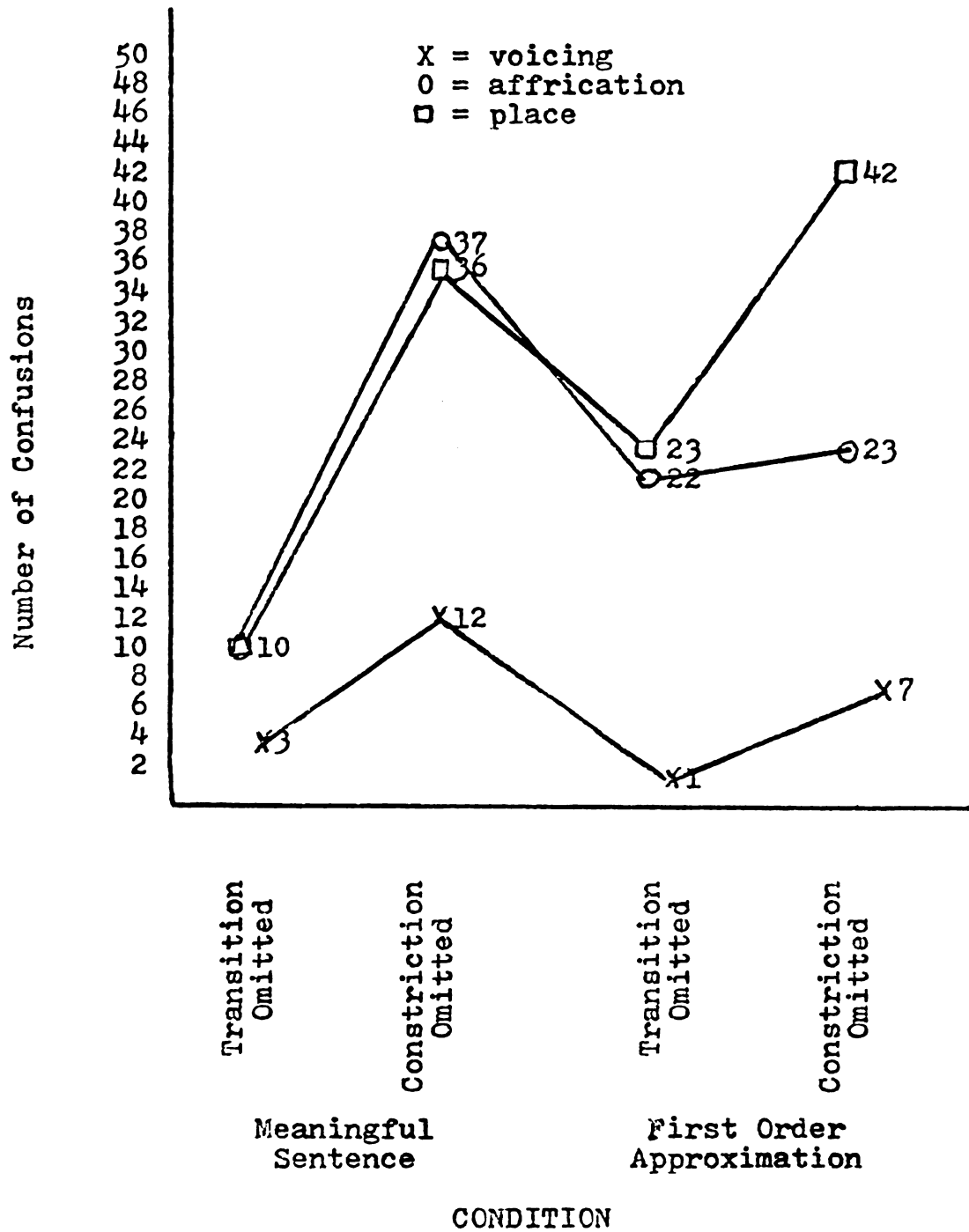


FIGURE 6.--Confusion errors for contextual and non-contextual speech

much higher percentage of affrication errors than in the first order approximation sentences and in Miller and Nicely's study. It must be remembered that the types of alterations done in Miller and Nicely's study were not the same as for the present study; therefore, direct comparisons with the current study are not possible. However, that the same results occurred in the first order approximations in the current study but not in the contextual sentences appears to be significant and casts further doubt upon the hypothesis that speech in single words is perceived in the same way as speech in context. It would appear that Fry's (1964) suggestion that auditory information coming in is compared to the person's linguistic model has a valid basis. These two types of information then result in perception. Further, results on perceptual studies using single words should not be extended to ongoing speech and vice versa.



## CHAPTER V

### SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Eight experimental tapes were produced for the purposes of this study. Each of the tapes presented thirty stimulus words under differing conditions. Four of the tapes presented words in a meaningful sentence and four of the tapes presented the words in random order. Under each of these conditions various portions of the speech signal were removed. The portions removed were the constrictions, the constrictions and the transitions, the transitions, and the transitions and the vowels. Eighty listeners were then randomly divided into eight groups and one of the groups heard each of the experimental tapes. Each person was asked to write down what he heard.

The mean number of words correctly identified for each of the eighty experimental conditions was then computed and the standard deviations were determined. An analysis of variance showed that a significant difference existed both in terms of the portion of speech removed and in terms of the order of presentation. Interaction effect was also significant.

A post hoc comparison was performed which showed that the following conditions contributed to the significance of the portion removed:

1. The comparison between sentences with the constriction removed vs. the sentences with the transition removed
2. The comparison between sentences with the constriction removed vs. the sentences with the transition and vowel removed
3. The comparison between sentences with the transition removed vs. the sentences with the constriction and transition removed
4. The comparison between sentences with the transition removed vs. the sentences with the transition and vowel removed
5. The comparison between first order approximation sentences with the constriction removed vs. first order approximation sentences with the constriction and transition removed
6. The comparison between first order approximation sentences with the constriction and transition removed vs. first order approximation sentences with the transition removed
7. The comparison between first order approximation sentences with the transition removed vs. the first order approximation sentences with the transition and vowel removed;

All comparisons contributed to the significance in contextual and non-contextual word order except the following:

1. Sentences with transition and vowel removed vs. first order approximation sentences with constriction and transitions removed
2. Sentences with the transition and vowel removed vs. first order approximation sentences with the transition removed



3. Sentences with the transition and vowel removed vs. first order approximation sentences with the constriction removed
4. Sentences with the transition and vowel removed vs. first order approximation sentences with the transition and vowel removed
5. Sentences with the constriction and transition removed vs. first order approximation sentences with the constriction and transition removed
6. Sentences with the constriction and transition removed vs. first order approximation sentences with the constriction and transition removed
7. Sentences with the constriction and transition removed vs. first order approximation sentences with the transition and vowel removed.

Each individual word was examined and it appeared that the differences were across all of the words.

A confusion matrix was constructed to examine the type of perceptual errors which were made under each of the orders of presentation. It was found that, in classifying errors as errors of place, voicing and affrication, differences in these three perceptual categories did exist.

The vowels alone were found to contain little information although the subjects did appear to apply some linguistic constraints to the material as shown by the high number of the's that the subjects guessed. In terms of five word sentences, one would expect to find a the in each of the sentences.

All of the findings of the study suggest that contextual perception is a different process than when

one hears the words only as single entities. Further, it appeared that the segments of the speech signal removed seemed to make a difference.

### CONCLUSIONS

Within the limitations of the instrumentation employed and the design of this study, the following conclusions are warranted:

1. The use of a magnet to remove selected portions of a speech signal in research on speech perception appears to offer potential for future research.
2. The context in which a word is embedded has a positive influence upon perception when either the transitions or constrictions have been removed.
3. When two of the three portions, either the transitions, the constrictions, or the vowel, are removed, contextual listening situations are not significantly more intelligible than non-contextual speech. It is possible that with so much information removed context is lost.
4. Linguistic rules appear to override auditory cues. Big being heard as the at the beginning of a sentence indicates that linguistic rules are being applied.

5. The total amount of the speech signal removed has an effect upon perception. The more of the signal removed, the less listeners are able to understand correctly. This applies to both contextual and non-contextual situations.
6. Despite the fact that approximately the same amount of the speech signal is removed, the removal of the constriction is more detrimental to speech than the removal of the transition when the speech signals are presented in a meaningful linguistic context, but no significant difference occurs in non-contextual situations.
7. Apparently in a situation where a forced choice is not required, the vowels carry little or no information. That is, if subjects hear only a string of vowels and are not forced to decide on the vowel as a word, they appear to have no information.
8. The perceptual errors in contextual speech differ from the perceptual errors in isolated words. Errors of perception based on affrication appear much higher in contextual speech than would be expected on the basis of previous studies using isolated words or than were found in the non-contextual lists in the current study. This is

particularly true when the constrictive portion is removed.

9. Perception of contextual material is a different task than perception of isolated words.

#### RECOMMENDATIONS FOR FURTHER RESEARCH

In view of the findings of this research, the following recommendations for additional research are presented:

1. This study should be replicated to see whether the method of removal is adequate to obtain similar results.
2. Sentences of various linguistic transformations should be used to allow for further comparisons in context.
3. Various orders of approximation could be employed to allow for an analysis of varying degrees of context.
4. A large sampling of selected consonants should be employed in contextual and non-contextual situations to allow for more definitive results in terms of perceptual confusions.
5. In looking at the amount of information available in vowels, comparisons should be made between forced choice and non-forced choice situations.

6. Various speakers should be employed to determine whether speaker variation results in any differences in terms of effect of context or perceptual removal.
7. An attempt should be made to compare perception of the same listener on contextual and non-contextual information to see whether one can predict performance on one task based on their performance on the other.
8. The study should be replicated on subjects of various ages to see whether the influence of context is affected by age. This would be useful in understanding language acquisition.
9. Studies using similar materials could be employed with clinical cases exhibiting various pathological conditions: e.g., hearing impairment; articulation disorders; language delayed children.

## BIBLIOGRAPHY

## BIBLIOGRAPHY

- Ahmend, R. and Fatehchand, R., Effects of sample duration on the articulation of sounds in normal and clipped speech. J. Acoust. Soc. Amer., 31, 1022-1029 (1959).
- Ainsworth, W. A., Perception of stop consonants in synthetic CV syllables. Language and Speech, 11, 139-156 (1968).
- Ali, Latif, Gallagher, Goldstein J., and Daniloﬀ, R., Perception of coarticulated nasality. J. Acoust. Soc. Amer., 49, 538-540 (1971).
- Amerman, J., Daniloﬀ, R., and Moll, K., Lip and jaw coarticulation for the phoneme /æ/. J. Speech Hearing Res., 13, 147-161 (1970).
- Bastian, J., Delattre, P.C., and Liberman, A. M., Silent intervals as a cue for the distinction between stops and semivowels in the medial position. J. Acoust. Soc. Amer., 31, 1568 (1959).
- Bever, T. G., Lackner, J. R., and Kirk, R., The underlying structures of sentences are the primary units of immediate speech processing. Perception & Psychophysics, 5, 225-234 (1969).
- Bever, T. G., Lackner, J. R., and Stolz, W., Transitional probability is not a general mechanism for the segmentation of speech. J. exp. Psychol., 79, 387-394 (1969).
- Black, J. W., Predicting the intelligibility of words. In A. Sovyarir and P. Aalto (Eds.), Proceedings of the 4th International Congress of Phonetic Science (1962).
- Brady, P. T., Perception of speech-like sounds. J. Acoust. Soc. Amer., 32, 1501 (1960).
- Brady, P. T., House, A. S. and Stevens, K. S., Perception of sounds characterized by a rapidly changing resonant frequency. J. Acoust. Soc. Amer., 33, 1357-1362 (1961).

- Brown, R. W. and Hildrum, D. C., Expectancy and the perception of syllables. Language, 32, 411-419 (1956).
- Chomsky, N., and Halle, M., The Sound Patterns of English. New York: John Wiley and Sons (1968).
- Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J., Some experiments on the perception of synthetic speech sounds. J. Acoust. Soc. Amer., 24, 597-606 (1952).
- Daniloff, R., and Moll, K., Coarticulation of lip rounding. J. Speech Hearing Res., 11, 707-721 (1968).
- Delattre, P. C., First formant transition as a cue to place of articulation. J. Acoust. Soc. Amer., 46, 110 (1969).
- Delattre, P. C., Liberman, A. M., and Cooper, F. S., Second and third formant transitions in English fricatives. J. Acoust. Soc. Amer., 32, 1501 (1960).
- Delattre, P. C., Liberman, A. M., and Cooper, F. S., Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Amer., 27, 769-773 (1955).
- Dukelskij, N. I. Cited in J. W. Falter and K. W. Ottan, Cybernetics and speech communication: A survey of Russian literature. IEEE Trans on Audio-Electroacoustics, AV15, 27-36 (1967).
- Fairbanks, G. and Kodman, F., Jr., Word intelligibility as a function of time compression. J. Acoust. Soc. Amer., 29, 636-641 (1957).
- Fairbanks, G., Gutman, N., and Miron, M. S., Effects of time compression upon the comprehension of connected speech. J. Speech Hearing Dis., 22, 10-19 (1957).
- Fant, G. G. M., Descriptive analysis of the acoustic aspects of speech. Logos, 5, 3-17 (1962).
- Fischer-Jørgensen, E., Acoustic analysis of stop consonants. Miscellanea Phonetica, 2, 42-59 (1954).



- Fletcher, H., Speech and Hearing. New York: Van Nostrand (1929).
- Fodor, J. A., and Bever, T. G., The psychological reality of linguistic segments. Journal of Verbal Learning and Verbal Behavior, 4, 414-420 (1965).
- Foulke, E., Comparison of comprehension of two forms of compressed speech. Exceptional Children, 33, 169-173 (1966).
- Fry, D. B., The correction of errors in the reception of speech. Phonetica, 11, 164-174 (1964).
- Fujimura, O., Effects of vowel context on the articulation of stopconsonants. J. Acoust. Soc. Amer., 33, 842 (1961).
- Garrett, M., Bever, T., and Fodor, J., The active use of grammar in speech perception. Perception and Psychophysics, 1, 30-32 (1966).
- Garvey, W. D., The intelligibility of speeded speech. J. exp. Psychol., 45, 102-108 (1953).
- Gerstman, L. J., Noise duration as a cue for distinguishing among fricative, affricate, and stop consonants. J. Acoust. Soc. Amer., 28, 160 (1956).
- Gerstman, L. J., Liberman, A. M., Delattre, P. C., and Cooper, F. S., Rate and duration of change in formant frequency as cues for the identification of speech sounds. J. Acoust. Soc. Amer., 26, 952 (1954).
- Green, P. S., Consonant-vowel transitions: A spectrographic study. Studia Linguistica, 12, 57-105 (1958).
- Grimm, W. A., Perception of segments of English-spoken consonant-vowel syllables. J. Acoust. Soc. Amer., 40, 1454-1461 (1966).
- Halle, M., Hughes, G. W., and Radly, J. P. S., Acoustical properties of stop consonants. J. Acoust. Soc. Amer., 29, 107-116 (1957).
- Harris, C. M., A study of the building blocks in speech. J. Acoust. Soc. Amer., 25, 962-969 (1953).
- Harris, K. S., Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1, 1-7 (1958).

- Harris, K. S., Cues for the identification of the fricatives of American English. J. Acoust. Soc. Amer., 26, 952 (1954).
- Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., and Cooper, F. S., Effects of 3rd formant transitions on the perception of the voiced stop consonants. J. Acoust. Soc. Amer., 30, 122-126 (1958).
- Hays, W. L., Statistics. New York: Holt, Rinehart and Winston, Inc. (1963).
- Heinz, J. M., and Stevens, K. N., On the properties of voiceless fricative consonants. J. Acoust. Soc. Amer., 33, 589-596 (1958).
- Hendrick, M., Are phonemes really realized? In E. Zwirner and W. Bethge (Eds.), Proceedings of the 5th International Congress of Phonetic Science, 426-430 (1965).
- Henke, W. R., Dynamic articulatory model of speech production. 1967 Conference on Speech Communication and Processing. Cambridge: The MIT Press (1967).
- Hoffman, H. S., Study of some cues in the perception of voiced stop consonants. J. Acoust. Soc. Amer., 30, 1035-1041 (1958).
- Hughes, J. W., and Halle, M., Spectral properties of fricative consonants. J. Acoust. Soc. Amer., 28, 303-310 (1956).
- Ingemann, P., Formants as a voicing cue in the perception of /z/s. J. Acoust. Soc. Amer., 32, 1501 (1960).
- Jakobson, R., Fant, G., and Halle, M. Preliminaries to Speech Analysis. Cambridge: MIT Press (1963).
- Kelly, J., Anthony, J. K., and Uldall, E., Tempo and transition. J. Acoust. Soc. Amer., 35, 1113 (1963).
- Klumpp, R. G., and Webster, J. C., Intelligibility of time-compressed speech. J. Acoust. Soc. Amer., 31, 265-267 (1961).

- Kozkevnikov, V. A., and Chistovich, L. A., Rech Artikuliatsieu i Vospruatie. Moscow-Leningrad. (Translation Speech: Articulation and Perception) Washington, D. C.: Joint Publications Research Service (1965).
- Kurtzrock, G. H., The effects of time and frequency distortion upon word intelligibility. Speech Monographs, 24, 94 (1957).
- Ladefoged, P., Three Areas of Experimental Phonetics. London: Oxford University Press (1967).
- Ladefoged, P. and Broadbent, D. E., Information conveyed by vowels. J. Acoust. Soc. Amer., 29, 98-104 (1957).
- Lane, H., The motor theory of speech perception: A critical review. Psychol. Review, 74:6, 431-461 (1967).
- Lawrence, D. L. and Byers, V. W., Identification of voiceless fricatives by high frequency hearing impaired listeners. J. Speech Hearing Res., 12, 426-434 (1969).
- Liberman, A. M., Some results of research on speech perception, J. Acoust. Soc. Amer., 29, 117-123 (1957).
- Lieberman, P., Intonation, Perception and Language. Cambridge: The MIT Press (1967).
- Liberman, A. M., Delattre, P. C., and Cooper, F. S., The role of selected stimulus-variables in the perception of the unvoiced stop consonants. Amer. J. of Psychol., 65, 497-516 (1952).
- Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J., The role of consonant vowel transitions in the perception of the stop and nasal consonants. Psychol. Monog., 68:379, 1-13 (1954).
- Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S., Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. exp. Psychol., 52, 127-137 (1956).
- Lindblom, B. E. F., and Striddert-Kennedy, M., On the role of formant transition in vowel recognition. J. Acoust. Soc. Amer., 42, 830-843 (1967).

- Lisker, L. and Abramson, A. S., Some effects on context on voice onset time in English stops. Lang. Speech, 10, 1-28 (1967).
- Lotz, J., Abramson, A. A., Gerstman, L. J., Ingeman, F., and Nemser, W. J., The perception of English stops by speakers of English, Spanish, Hungarian, and Thai: A tape-cutting experiment. Lang. Speech, 3, 71-77 (1960).
- Malécot, A., Acoustic clues for nasal consonants: An experimental study involving a tape splicing technique. Language, 32, 274-284 (1956).
- Malécot, A., The role of releases in the identification of released final stops. Language, 32, 274-284 (1956).
- McLain, J., A comparison of two methods of producing rapid speech. International Journal for the Education of the Blind, 12, 40-43 (1962).
- Mattingly, I. G. and Liberman, A. M., The speech code and the physiology of language. In K. N. Leibovic (Ed.), Information Processing in the Nervous System. Springer Verlag (1970).
- Mattingly, I. G., Liberman, A. M., Syrdal, A. K., and Malives, T., Discrimination of  $F_2$  transition in speech context and in isolation. J. Acoust. Soc. Amer., 45, 314-315 (1969).
- Miller, G., The magical number seven, plus or minus two: some limits to our capacity for processing information. Psychol. Rev., 63, 81-97 (1956).
- Miller, G. and Licklider, J., The intelligibility of interrupted speech. J. Acoust. Soc. Amer., 22, 167-173 (1950).
- Miller, G. A. and Nicely, P. E., An analysis of perceptual confusion among some English consonants. J. Acoust. Soc. Amer., 27, 338-352 (1955).
- Moll, K. L., Cinefluorographic techniques in speech research. J. Speech Hearing Res., 3, 227-241 (1960).
- Moll, K. L., Velopharyngeal closure on vowels. J. Speech Hearing Res., 5, 30-37 (1962).

- Moser, H. M., One-syllable Words. Columbus, Ohio: C. E. Merrill Publishing Company (1969).
- Nakata, K., Synthesis and perception of nasal consonants. J. Acoust. Soc. Amer., 31, 661-666 (1959).
- O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., and Cooper, F. S., Acoustic cues for the perception of initial /w, j, r, l/ in English. Word, 13, 24-43 (1957).
- Öhman, S. E. G., Perception of segments of VCV utterances. J. Acoust. Soc. Amer., 40, 979-988 (1966).
- Peterson, G. E., and Barney, H. L., Control methods used in a study of the vowels. J. Acoust. Soc. Amer., 24, 175-184 (1952).
- Pickett, J. M. and Pollack, I., Adjacent context and the intelligibility of speech excised from fluent speech. J. Acoust. Soc. Amer., 35, 807- (1963).
- Pollack, I. and Pickett, J. M., Intelligibility of excerpts from fluent speech: Auditory vs. structural context. Journal of Verbal Learning and Verbal Behavior, 3, 79-84 (1964).
- Potter, R. K., Kopp, G. A., and Green, H. C., Visible Speech. New York: Van Nostrand (1947).
- Reed, J. A. and Wang, W. S-Y., The perception of stops after s. Phonetica, 6, 78-81 (1961).
- Schatz, C. D., The role of context in the perception of stops. Lang. Speech, 30, 47-56 (1954).
- Shearne, J. N. and Holmes, J. N., An experimental study of the classification of sounds in continuous speech according to their distribution in the formant 1 and formant 2 plane. In A. Sovyarvi and P. Aalto (Eds.), Proceedings of the Fourth International Congress of Phonetic Science (1962).
- Sherman, G. Cited in R. M. Warren and C. J. Obusek, Speech perception and phonemic restoration. Perception and Psychophysics, 9, 358-362 (1971).
- Soderberg, G. A., Linguistic factors in stuttering. J. Speech Hearing Res., 10, 801-810 (1967).

- Stevens, K. N. and House, A. S., Perturbations of vowel articulation by consonantal context: An acoustical study. J. Speech Hearing Res., 6, 111-128 (1963).
- Stevens, K. N., House, A. S., and Paul, A. P., Acoustic description of syllable nuclei: An interpretation in terms of a dynamic model of articulation. J. Acoust. Soc. Amer., 40, 123-132 (1966).
- Straus, O. H., Phonetics and communication. J. Acoust. Soc. Amer., 22, 709-711 (1950).
- Stevens, P., Spectra of fricative noise in human speech. Lang. Speech, 3, 33-49 (1960).
- Swaffield, J., Shearne, J. N., and Holmes, J. N., Some measurements of vowel sounds of conversational speech. J. Acoust. Soc. Amer., 33, 1683 (1961).
- Warren, R. M., Perceptual restoration of missing speech sounds. Science, 167, 392-393 (1970).
- Warren, R. M. and Obusek, C. J., Speech perception and phonemic restoration. Perception and Psychophysics, 9, 358-362 (1971).
- Wickelgren, A., Distinctive features and errors in short-term memory for English consonants. J. Acoust. Soc. Amer., 39, 388-398 (1966).

APPENDIX  
RAW SCORES FOR 10 SUBJECTS  
IN THE EIGHT EXPERIMENTAL  
CONDITIONS

## EXPERIMENTAL CONDITION 1

(Meaningful Sentence with the Consonant  
Constrictions Removed)

Subject	Sentence						Total
	1	2	3	4	5	6	
1	4	4	3	4	2	3	20
2	4	4	3	4	4	1	20
3	3	4	3	0	4	0	14
4	4	4	4	4	4	1	21
5	2	5	4	4	4	4	23
6	3	5	3	4	4	3	22
7	3	4	3	0	4	1	15
8	4	3	3	0	5	5	20
9	3	4	4	3	4	3	21
10	<u>3</u>	<u>4</u>	<u>3</u>	<u>0</u>	<u>3</u>	<u>3</u>	<u>16</u>
Total	33	41	33	23	38	24	192



## EXPERIMENTAL CONDITION 2

(First Order Approximation with the  
Consonant Constrictions Removed)

Subject	Sentence						Total
	1	2	3	4	5	6	
1	3	2	1	1	1	1	9
2	2	1	4	2	1	2	12
3	2	2	2	2	2	1	11
4	2	2	1	2	1	1	9
5	1	2	1	0	1	1	6
6	1	2	0	2	1	1	7
7	3	2	0	2	1	1	9
8	2	2	1	2	1	0	8
9	2	2	2	2	1	1	10
10	<u>1</u>	<u>2</u>	<u>1</u>	<u>2</u>	<u>2</u>	<u>1</u>	<u>9</u>
Total	19	19	13	17	12	10	90

## EXPERIMENTAL CONDITION 3

(Meaningful Sentence with the Consonant  
Constrictions and Transitions  
Removed)

Subject	Sentence						Total
	1	2	3	4	5	6	
1	2	0	0	0	0	0	2
2	3	1	1	0	1	0	6
3	1	1	1	0	0	0	3
4	0	1	0	0	0	0	1
5	0	2	0	0	0	0	2
6	0	1	0	0	0	0	1
7	2	3	1	0	0	0	6
8	2	1	0	0	0	0	3
9	2	1	0	0	0	2	5
10	<u>3</u>	<u>1</u>	<u>2</u>	<u>0</u>	<u>0</u>	<u>1</u>	<u>7</u>
Total	15	12	5	0	1	3	36

## EXPERIMENTAL CONDITION 4

(First Order Approximation with the Consonant  
Constrictions and Transitions  
Removed)

Subject	Sentence						Total
	1	2	3	4	5	6	
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	1	0	1	0	2
4	0	0	0	0	0	0	0
5	0	1	0	0	0	0	1
6	0	0	1	0	0	0	1
7	0	0	0	0	0	0	0
8	1	0	0	0	1	0	2
9	0	0	0	0	0	0	0
10	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
Total	1	1	2	1	1	0	6

## EXPERIMENTAL CONDITION 5

(Meaningful Sentence with Transitions Removed)

Subject	Sentence						Total
	1	2	3	4	5	6	
1	5	5	5	5	5	4	29
2	4	5	4	3	5	4	25
3	5	4	3	5	3	4	24
4	5	5	4	2	3	3	22
5	3	5	3	5	4	4	24
6	5	5	5	5	2	4	26
7	5	5	5	4	5	5	29
8	5	5	5	5	5	5	30
9	5	5	3	4	5	5	27
10	<u>3</u>	<u>5</u>	<u>4</u>	<u>5</u>	<u>5</u>	<u>3</u>	<u>25</u>
Total	45	49	41	43	42	41	261

EXPERIMENTAL CONDITION 6  
 (First Order Approximation with  
 Transitions Removed)

Subject	Sentence						Total
	1	2	3	4	5	6	
1	3	3	3	3	0	3	15
2	3	2	4	4	1	3	17
3	2	0	2	0	1	0	5
4	2	1	4	3	1	3	14
5	2	0	3	3	2	2	12
6	2	2	4	2	1	2	13
7	2	0	2	0	0	2	6
8	2	2	4	2	1	2	13
9	2	2	2	3	1	0	10
10	<u>2</u>	<u>0</u>	<u>2</u>	<u>0</u>	<u>1</u>	<u>2</u>	<u>7</u>
Total	22	12	30	20	9	19	112

## EXPERIMENTAL CONDITION 7

(Meaningful Sentence with the Transitions  
and Steady State Vowel Removed)

Subject	Sentence						Total
	1	2	3	4	5	6	
1	0	2	0	1	0	4	7
2	0	2	0	1	0	4	7
3	2	0	0	3	0	2	7
4	0	1	0	0	0	3	4
5	0	2	0	1	0	4	7
6	5	3	0	0	0	4	12
7	1	4	0	1	0	4	10
8	1	1	0	0	0	4	6
9	0	2	0	0	0	4	6
10	<u>0</u>	<u>2</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>4</u>	<u>6</u>
Total	9	19	0	7	0	37	72

## EXPERIMENTAL CONDITION 8

(First Order Approximation with the  
Transitions and Steady State  
Vowel Removed)

Subject	Sentence						Total
	1	2	3	4	5	6	
1	1	0	1	1	0	0	3
2	0	0	2	0	0	0	2
3	0	0	0	0	0	0	0
4	0	0	2	0	0	1	3
5	0	0	3	0	0	0	3
6	1	0	0	0	0	0	1
7	1	0	2	0	0	0	3
8	1	0	2	0	0	0	3
9	2	0	0	0	0	0	2
10	<u>2</u>	<u>0</u>	<u>0</u>	<u>3</u>	<u>0</u>	<u>0</u>	<u>5</u>
Total	8	0	12	4	0	1	25

1000  
1.1.1.1  
2.1



MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03058 0249