

## INFORMATION TO USERS

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

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

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

### University Microfilms

300 North Zeeb Road  
Ann Arbor, Michigan 48106

A Xerox Education Company

72-24,149

ROSENTHAL, Ruth D., 1944-  
THE EFFECTS OF NARROW BANDS OF LOW  
FREQUENCY SPEECH ENERGY ON INTELLIGIBILITY.

The City University of New York, Ph.D., 1972  
Speech

University Microfilms, A XEROX Company , Ann Arbor, Michigan

THE EFFECTS OF NARROW BANDS OF LOW FREQUENCY SPEECH  
ENERGY ON INTELLIGIBILITY

by

RUTH D. ROSENTHAL

A dissertation submitted to the Graduate  
Faculty in Speech in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy, The City University  
of New York.

1972

This manuscript has been read and accepted for the Graduate Faculty in Speech in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

May 1, 1972      James K. Long  
date      Chairman of Examining Committee

May 1, 1972      Harry Swett  
date      Executive Officer

James W. Abel  
John Rose  
Supervisory Committee

Norma S. Rees

The City University of New York

PLEASE NOTE:

Some pages may have  
indistinct print.

Filmed as received.

University Microfilms, A Xerox Education Company

## ACKNOWLEDGMENTS

The writer wishes to gratefully acknowledge the members of her advisory committee, Professors James Abel, James Lang, and Harry Levitt, for their guidance in formulating this study, and for their aid in the preparation of the final manuscript.

The writer is particularly indebted to the chairman of her committee, Professor James Lang, without whose inspiration, help, and encouragement this manuscript could never have been written.

## TABLE OF CONTENTS

### Chapter

|      |                                              |     |
|------|----------------------------------------------|-----|
| I.   | INTRODUCTION AND STATEMENT OF THE PROBLEM... | 1   |
|      | Introduction                                 |     |
|      | The Problem and the Null Hypothesis          |     |
| II.  | REVIEW OF THE LITERATURE.....                | 7   |
|      | The Frequencies of Importance for the        |     |
|      | Intelligibility of Speech                    |     |
|      | The Intelligibility of Single- and           |     |
|      | Multi-Pass Band Systems                      |     |
|      | The Effects of Ear Dominance on              |     |
|      | Intelligibility                              |     |
|      | Techniques for Measuring Intelligibility     |     |
|      | Cues for Consonant Discrimination            |     |
| III. | METHODS AND PROCEDURES.....                  | 51  |
|      | Preparation of the Stimulus Materials        |     |
|      | Basic Instrumentation                        |     |
|      | Calibration                                  |     |
|      | Subjects and Testing Procedures              |     |
|      | Data Reduction                               |     |
| IV.  | RESULTS AND DISCUSSION.....                  | 72  |
|      | Results                                      |     |
|      | Discussion                                   |     |
| V.   | SUMMARY.....                                 | 97  |
|      | Introduction and Statement of the Problem    |     |
|      | Subjects and Procedures                      |     |
|      | Results                                      |     |
|      | Discussion                                   |     |
|      | APPENDIX.....                                | 102 |
|      | BIBLIOGRAPHY.....                            | 106 |

## LIST OF TABLES

| Table                                                                                                                                                                          | Page |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 1. Experimental Design.....                                                                                                                                                    | 70   |
| 2. Analysis of Variance Summary Table of Main Effects and First Order Interactions.....                                                                                        | 73   |
| 3. Analysis of Variance Summary Table of Main Effects and First Order Interactions. For this analysis, the data were collapsed over Factor A (right-left ear differences)..... | 75   |
| 4. Errors according to Articulatory Feature.....                                                                                                                               | 84   |
| 5. Errors according to Sound Type.....                                                                                                                                         | 85   |
| *A. Sample of Answer Sheet on which Subjects Recorded their Responses.....                                                                                                     | 103  |
| B. Summary of the Order of Presentation of Experimental Conditions for each Subject...                                                                                         | 104  |

\* Tables designated by letters appear in the Appendix.

## LIST OF ILLUSTRATIONS

| Figure                                                                                                                                                                                                 | Page |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 1. Block Diagram of Apparatus used in Making the Test Tape Recordings.....                                                                                                                             | 53   |
| 2. Block Diagram of Apparatus used for Presenting the Stimulus Material.....                                                                                                                           | 55   |
| 3. Block Diagram of Apparatus used to Determine the Frequency Response Characteristics of the System Used in Recording the Master Tapes.....                                                           | 56   |
| 4. Block Diagram of Apparatus used to Calibrate System Used in Making the Test Tape Recordings.....                                                                                                    | 59   |
| 5. Frequency Response of the System Used to Produce the Test Tapes. The frequency response shown in this graph was produced when the filter was set to produce an octave band centered at 80 Hz.....   | 60   |
| 6. Frequency Response of the System Used to Produce the Test Tapes. The frequency response shown in this graph was produced when the filter was set to produce an octave band centered at 160 Hz.....  | 61   |
| 7. Frequency Response of the System Used to Produce the Test Tapes. The frequency response shown in this graph was produced when the filter was set to produce an octave band centered at 315 Hz.....  | 62   |
| 8. Frequency Response of the System Used to Produce the Test Tapes. The frequency response shown in this graph was produced when the filter was set to produce an octave band centered at 1600 Hz..... | 63   |
| 9. Block Diagram of Apparatus Used to Calibrate the Gain Control of the Playback System.....                                                                                                           | 65   |

| Figure |                                                                                                                     | Page |
|--------|---------------------------------------------------------------------------------------------------------------------|------|
| 10.    | Block Diagram of Apparatus Used to Determine the Frequency Response Characteristics of the Playback System.....     | 66   |
| 11.    | Articulation Scores for the Monotic and Dichotic Listening Conditions for the Three Low Frequency Speech Bands..... | 76   |
| 12.    | Articulation Gain Functions for the Three Low Frequency Speech Bands.....                                           | 80   |

## CHAPTER I

### INTRODUCTION AND STATEMENT OF THE PROBLEM

#### Introduction

The major purpose of the present research was to evaluate the effects of narrow bands of low frequency speech information on intelligibility.<sup>1</sup> The specific bands examined in this research were 50-100, 110-220, and 200-400 Hz, respectively. Other variables which have been shown in previous research to influence intelligibility were also manipulated; namely, the intensity of the low frequency bands, the mode of presentation (monotic or dichotic), and ear dominance.

The goals of the present research can be stated in the form of a question as follows:

Are the effects on intelligibility of the low frequency speech bands contingent in any way on a) the intensity at which the low frequency speech bands are presented, b) the mode of presentation (monotic or dichotic) of the high and low frequency speech bands, and

---

<sup>1</sup>Intelligibility was measured by articulation testing. By this technique, one manipulates the variable of interest and determines the effect of this manipulation by measuring the number of correct responses. This number, expressed as a per cent, is called an articulation score.

c) the ear to which the high frequency speech band is presented.

The narrow bands of frequencies presented in the present research were a high band of 1100-2200 Hz and the following low bands: 50-100, 110-220, and 200-400 Hz. A basic assumption of this research was that the frequency region of approximately 1000-2000 Hz is important for intelligibility. However, the importance of the low frequencies to intelligibility appears uncertain. Fletcher (1929) reported that elimination of frequencies below 500 Hz has only a negligible effect on intelligibility. Steinberg (1929) noted that nonsense syllables containing stop and fricative consonants have no frequencies of great importance below 1000 Hz. But, when he eliminated all frequencies above 1000 Hz, articulation scores for these syllables were still 50-70 per cent.

Investigating the effects of high-pass and low-pass filtering in a background of noise, Pollack (1948) showed that the intensity of low frequency bands of speech is an important variable affecting speech intelligibility. At high intensity speech levels, articulation scores were higher when frequencies below 350 Hz were eliminated. However, at low speech levels, the presence of frequencies below 350 Hz improved articulation scores. The present writer was unable to determine by reviewing the literature or by prior experimentation the level at which the low frequency bands should be presented

for maximum contribution to intelligibility. Therefore, several intensity levels for the low frequency bands have been sampled in the present research.

The effect of the mode of presentation (monotic or dichotic), on intelligibility has been shown to be related to the intensity at which the low frequency speech bands are presented. Franklin (1969, 1970) found that the presentation of a low frequency speech band (240-480 Hz) at 0 or 20 dB SL to the same or opposite ear as a high frequency speech band (1020-2040 Hz) resulted in no significant difference in articulation scores. By contrast, when the low frequency speech band was presented at 40 dB SL, articulation scores were higher when the high and low frequency speech bands were in opposite ears. The present research attempted to replicate this finding by, in one instance, presenting the low frequency bands at several intensity levels and the high frequency band to the same ear, and in the second instance, presenting the low frequency bands and the high frequency band to opposite ears.

The effect of ear dominance is another important variable. In a dichotic listening condition, Shankweiler and Studdert-Kennedy (1967) indicated that the articulation scores of the right ear are significantly better than those of the left ear for consonant-vowel (CV) syllables. Kimura (1961) likewise found in a dichotic listening situation that articulation scores for digits

were significantly better in the right ear than in the left. Therefore, in the present research, half of the subjects received the high frequency speech band (the band assumed to be important to intelligibility) in the right ear and half received this band in the left ear.

The present research was designed to systematically control the variables discussed above--the center frequency of the low frequency speech bands, the intensity of the low frequency speech bands, the mode of presentation (monotic or dichotic), and ear dominance--which have been shown to affect articulation scores in dichotic listening experiments and in filtered speech experiments.

#### The Problem and the Null Hypotheses

The present research was designed to investigate the effects of narrow bands of low frequency speech energy on intelligibility. Previous research has identified other variables--the intensity of the low frequency energy, the mode of presentation (monotic or dichotic), and ear dominance--as significant determinants of intelligibility. This research attempted to manipulate these variables in such a way as to reveal the existence of any interactions among them. The goals of the present research can be stated in the form of a set of null hypotheses as follows:

Intelligibility scores, for a high frequency band

of speech (1100-2200 Hz) will not be significantly altered by:

a) presentation of a low frequency band of speech (50-100 Hz) at 0, 10, 20, 30, and 40 dB NSL,<sup>2</sup> when

1. the high frequency speech band is presented to the right ear and the low frequency speech band is presented to the left ear

2. the high frequency speech band is presented to the left ear and the low frequency speech band is presented to the right ear

3. both high and low frequency speech bands are presented to the right ear

4. both high and low frequency speech bands are presented to the left ear.

b) presentation of a low frequency band of speech (110-220 Hz) at 0, 10, 20, 30, and 40 dB NSL, when

1. the high frequency speech band is presented to the right ear and the low frequency speech band is presented to the left ear

2. the high frequency speech band is presented to the left ear and the low frequency speech band is presented to the right ear

---

<sup>2</sup>NSL stands for Normal Speech Level in a given band. Zero dB NSL corresponds to the speech level in the narrow bands under the experimental conditions being equal to the speech levels in the same frequency bands for broad band speech.

3. both high and low frequency speech bands are presented to the right ear

4. both high and low frequency speech bands are presented to the left ear.

c) presentation of a low frequency band of speech (200-400 Hz) at 0, 10, 20, 30, and 40 dB NSL, when

1. the high frequency speech band is presented to the right ear and the low frequency speech band is presented to the left ear

2. the high frequency speech band is presented to the left ear and the low frequency speech band is presented to the right ear

3. both high and low frequency speech bands are presented to the right ear

4. both high and low frequency speech bands are presented to the left ear.

## CHAPTER II

### REVIEW OF THE LITERATURE

#### The Frequencies of Importance for the Intelligibility of Speech

This section reviews literature concerning the frequencies of importance for speech intelligibility. In many studies, background noise has been employed. Although the use of noise was not part of the present research, such studies are described because their results are applicable to studies carried out without noise (Palva, 1965). As is shown in this section, at least two variables influence the frequencies of importance for intelligibility; namely, the signal-to-noise ratio and the stimulus material.

One method for determining the frequencies important for intelligibility is to measure the masking effects of various pure tones and narrow bands of noise. Miller (1947) measured the shift in the threshold of perceptibility due to a) broad band noise (20-4000 Hz), b) narrow bands of noise (covering the range 135-4000 Hz), and c) pure tones. Results indicated that at high masking levels the low frequencies were more effective maskers than the high frequencies. Specifically, Miller found that a 300 Hz tone presented at 100 dB SPL shifted

the threshold of perceptibility by 42 dB; whereas, a 1000 Hz tone presented at the same intensity shifted the threshold of perceptibility by only 18 dB. Similarly, when the narrow bands were sampled, complete masking of the speech was attained only with the noise bands 135-400 Hz and 350-700 Hz; when noise bands above 1000 Hz were presented, the speech could still be perceived. This is consistent with the results reported by Hirsh and Bowen (1954), who likewise investigated the effects of bands of noise on the threshold of perceptibility. Their findings indicated that of the noise bands sampled the band 670-1000 Hz was the most effective masker.

It thus appears that at high noise levels the low frequencies are more effective maskers of speech than the high frequencies. Conversely, however, at low noise levels, the high frequencies have been found to be more effective maskers (Miller, 1947). This apparent inconsistency was explained by Miller by referring to the energy distribution of speech as seen in the long term average speech spectrum. The spectrum shows that at least half of the energy for speech is concentrated in the frequency range below 1000 Hz. Thus, the presentation of low frequency noise at low intensity levels will have little effect on the intense low frequency components of speech. As the intensity is increased, the low frequency noise will be strong enough to mask the low

frequencies of speech, and the masking effect of the low frequencies will spread to the high frequencies. Similarly, the speech spectrum shows that the high frequency components of speech are weak and therefore easily masked. Consequently, the presentation of high frequency noise at low intensity levels will mask the high frequencies of speech.

Another method of determining the frequencies important to intelligibility is to measure the intelligibility of filtered speech. Egan and Weiner (1946) attempted to find the frequency region of band-pass filtered speech that was most important to the intelligibility of speech in broad band noise. Their findings indicated that the frequency of greatest importance was between 1100-2000 Hz. Consistent with Miller (1947), Egan and Weiner also found that the frequencies important to intelligibility were a function of the level of the speech relative to the level of the noise (the signal-to-noise ratio). Specifically, they found that at high signal-to-noise ratios the center frequency was about 2000 Hz and that frequencies as high as 6500 Hz may be important for intelligibility; at poor signal-to-noise ratios, the center frequency shifted down to 1000 Hz, and frequencies above 3900 Hz were of little importance for intelligibility.

Pollack (1948) reported similar findings regarding

the relation between the frequencies important to intelligibility and the signal-to-noise ratio. Pollack determined the "importance frequency" for various intensity levels of phonetically balanced words presented against a constant level of white noise (1000-7000 Hz at an equivalent free field pressure of 43 dB SPL). The "importance frequency," as defined by Webster (1964) refers to that frequency which divides the speech spectrum so that frequencies above and frequencies below the importance frequency contribute equally to intelligibility. Pollack found that at +10 dB orthotelephonic gain<sup>1</sup> the importance frequency was 800 Hz. However, this frequency shifted, respectively, to 1010, 1300, 1430, and 1620 Hz for the levels 20, 30, 40, and 50 dB orthotelephonic gain. Thus there was a progressive shift in the relative contribution to intelligibility from the low frequencies to the high frequencies as the intensity of speech was raised relative to a constant background noise. In addition, Pollack observed that at high speech levels articulation scores were higher when the frequencies below 350 Hz were eliminated. However, at low speech levels, articulation scores increased when the frequencies below 350 Hz were added. This is again consistent with results reported

---

<sup>1</sup>According to Stevens (1951), orthotelephonic gain refers to "the transmission of a system relative to that of a standard system consisting of a talker and a listener who face one another at a distance of 1 meter in open air."

above (Miller, 1947; Egan and Weiner, 1946). As a possible explanation, Pollack suggested that at low intensity levels the frequencies below 350 Hz were probably effective because they made the speech audible, but at high levels, they may have masked the weak high frequency sounds. This explanation appears to imply interband masking, which Kryter (1960) doubts is significant at normal speech levels.

Kryter (1960) presented phonetically balanced words at an overall level of 75 dB SPL against a background of noise. The speech material was filtered in such a way that several passbands each 500 Hz wide were produced. When two of the 500 Hz bands were presented, Kryter noted that "the proper position of the low band is somewhat dependent upon the center frequency of the second band it is to be paired with." For example, at a signal-to-noise ratio of 20 dB, when a band with a center frequency of 750 Hz was paired with a second band positioned around either 1750 or 2500 Hz, comparable articulation scores were obtained; whereas, articulation scores were not as high when a first band centered at 500 Hz was paired with a second band centered at frequencies higher than 1750 Hz.

Kryter also observed this same kind of interaction when three 500 Hz bandwidths were presented. He found that maximum intelligibility scores resulted when

the center frequencies of the 500 Hz bandwidths were about 500 Hz for the low band, 1500 Hz for the mid band, and 2500 Hz for the high band.

Palva (1965), devising a filtered speech test for the diagnosis of central auditory disorders, investigated the effects of filtering on intelligibility. Phonetically balanced, filtered Finnish words were presented monaurally at 50 dB SL. The results of octave band filtering covering the range 240-5720 Hz indicated that maximum articulation scores were obtained for the bands 900-1800 Hz (52.0 per cent  $\pm$  12.6 per cent) and 1080-2160 Hz (54.0 per cent  $\pm$  12.7 per cent). Combining results of octave bands with the results of other bandwidths studied ( $1/4$ ,  $1/2$  and  $1/3$  octave bands), Palva found, first, that the mean frequency bands yielding the highest articulation scores were in the area of 1300-2000 Hz, and second, that between 500-700 Hz was another area of great importance for intelligibility. These results obtained by Palva for filtered speech in a quiet background agree quite well with those reported by Kryter (1960) for filtered speech in noise.

Fletcher (1929), one of the first researchers to systematically study the effects of high- and low-pass filtering on intelligibility, reported that the "elimination of frequencies below 500 Hz produced only a small effect on intelligibility."

Steinberg (1929) also studied the effects of high- and low-pass filtering on various groups of speech sounds, his material being CVC nonsense syllables. Results indicated that 200-3000 Hz was the important frequency range for vowels; whereas, consonant discrimination depended upon the frequency region above 500 Hz. Specifically, 500-4000 Hz was found to be the important frequency range for nasal consonants, 750-5000 Hz for stop consonants, and 1000-8000 Hz for fricative consonants.

Steinberg further noted that high- and low-pass filtering produced different effects. High-pass filtering of stops and fricatives showed that these sounds have no frequencies of great importance below 1000 Hz. However, when all frequencies above 1000 Hz were eliminated, articulation scores were still 50-70 per cent. Steinberg explained that this finding resulted from the fact that a given consonant may be identified because of the information contained in the beginnings and endings of adjoining vowel sounds.

The different effects of high- and low-pass filtering were likewise observed by Palva (1965). Finnish speech materials were filtered and presented at 50 dB SL. Low-pass filtering indicated that hearing the fundamental frequency alone was not sufficient for intelligibility. Only after the first formant entered at about 600 Hz did intelligibility increase and continued to increase until the second formant was heard.

In contrast, by high-pass filtering, Palva found that the elimination of the fundamental frequency and the first formant did not affect intelligibility greatly. Only when the cut-off frequency was about 1000 Hz did intelligibility decrease, thus indicating that the majority of vowels could be identified without the first formant. Palva suggested that this unusual finding may be due to the attenuation rate of the filters, which was only 60 dB per octave. It was therefore possible that the first formant (which Palva noted to be 600-700 Hz) was heard by the subjects. Another possible explanation for this finding is that the ear may have provided the fundamental frequency in accordance with the concept of the missing fundamental or residue phenomenon (Schouten, 1940).

The different effects of high- and low-pass filtering noted by Steinberg (1929), and Palva (1965), above, were also seen by Miller and Nicely (1955). With low-pass filtering, the perception of voicing and nasality were least affected, the perception of affrication and duration being somewhat superior to that of place of articulation, but far inferior to that of voicing and nasality. These results were comparable to those obtained when unfiltered nonsense syllables were presented in a background of noise. Miller and Nicely suggested that "the correspondence seems reasonable if we think of the high frequency components of speech as

relatively weak and therefore most susceptible to masking of the uniform spectrum of noise. That is to say, the uniform noise spectrum should mask high frequencies more than low, so it is in effect a kind of low pass system."

Using high-pass filtering, Miller and Nicely found that intelligibility scores for all linguistic features, with a minor exception for duration, decreased similarly as low frequencies were eliminated. This homogeneity resulted from the fact that the errors did not fall into obvious patterns, but were distributed almost randomly. Miller and Nicely observed that "when an error occurs with high-pass filtering, there is little chance of predicting what the error will be." Thus there appears to be an important difference between high- and low-pass filtering; "low pass filtering affects the several linguistic features differentially, leaving the phonemes audible but similar in predictable ways, whereas high-pass filtering removes most of the acoustic power in the consonants, leaving them inaudible and consequently producing quite random confusions."

In addition, Miller and Nicely determined the importance frequency by dividing the frequency scale into two segments making equal contribution to intelligibility. This frequency was found to be 1550 Hz. When discussing the meaning of this frequency, Miller and Nicely noted that this frequency varied for the different artic-

ulatory features of voicing, nasality, duration, affrication, and place of articulation. Thus, the importance frequency for a given vocabulary will vary, depending upon how it is loaded with these features. In the past, the distinguishing fact that the different articulatory features have different importance frequencies had not been noted. For example, French and Steinberg (1947) determined the importance frequency for CVC nonsense syllables. At 10 dB orthotelephonic gain, the importance frequency was found to be 1900 Hz; i.e., equal articulation scores (68 per cent) resulted when the frequency range above or below the importance frequency was filtered out.

Hirsh et al. (1954) determined the effects of high- and low-pass filtering on the intelligibility of various kinds of speech materials, including phonetically balanced words, nonsense syllables, and disyllabic words manifesting various stress patterns. Results indicated that the importance frequency was the same for all words sampled. However, the articulation scores at this frequency increased as the number of syllables per word increased.

Attempting to define a performance-intensity function<sup>2</sup> for each of several high- and low-pass

---

<sup>2</sup>The performance-intensity function is a graph depicting intelligibility (as measured by articulation testing) as a function of speech intensity.

frequency bands, Speaks (1967) presented synthetic sentences representing third order approximations of actual English sentences. When the stimulus material was presented at 20 dB SPL, results indicated that the importance frequency was in the region of 725 Hz, which resulted in almost perfect articulation scores (99.6 per cent). However, presentation of the same material at a slightly lower level (16 dB SPL) resulted in approximately a 50 per cent decrease in articulation scores at this frequency.

In summary, it appears that the importance frequency for sentences is about one octave lower than that for isolated words at high signal-to-noise ratios. However, when words are presented at poor signal-to-noise ratios, the importance frequency is observed to decrease and become comparable to that for sentences. Thus there appears to be an interaction between the importance frequency, the type of stimulus material, and the intensity at which the material is presented.

Interaction between the stimulus material and the level of presentation also influences intelligibility. This was demonstrated in a study by Speaks, Jerger, and Jerger (1966) in which the intelligibility of filtered (low-pass, 500 Hz) first, second, and third order sentence approximations were compared. The level necessary for 50 per cent identification of the sentences was

found to decrease from 20.4, to 19.7, to 19.3 dB SPL, as more contextual restraints were imposed.

With filtered third order sentence approximations (high-pass, 7000 Hz or low-pass, 125 Hz), Speaks (1967) found that performance "always reached 100 per cent if the material was presented with sufficient intensity." This finding was thought by Speaks to be "due to the redundancy which characterizes sentences."

When comparing performance on sentences with performance on phonetically balanced words, a striking contrast was seen. In a second experiment, Speaks (1967) presented third order synthetic sentence approximations at 40, 60, 80, and 100 dB SPL and phonetically balanced words at 60, 80, and 100 dB SPL. All materials were filtered (low-pass, 125 Hz) and presented monaurally. As before, performance on the sentence material increased rapidly to nearly 100 per cent and remained constant throughout the range tested. By contrast, maximum performance (6 per cent) for the phonetically balanced words occurred at 80 dB SPL and did not increase with added intensity.

These findings are consonant with those reported by Pollack (1948) for phonetically balanced words presented at various levels against a background of noise. Pollack noted that there is "no amount of gain which can compensate for the loss of more than a limited

range of high and low frequencies." He did observe, however, a difference in slope of the articulation-gain functions for high- and low-pass filtering. Based on this difference, Pollack noted that "it is easier to compensate for the loss of the low frequencies by an increase in gain than it is for a proportionate loss of high frequencies."

Linden (1964) monaurally presented filtered spondee words at 30, 40, 50, and 60 dB SL. Intelligibility as a function of sensation level was compared for two filter bands; namely, 560-715 and 1800-2200 Hz. For the low filter band (560-715 Hz), articulation scores were low at 30 and 40 dB SL but rose rapidly with increased intensity. For the high filter band (1800-2200 Hz), articulation scores were low at low intensities but increased only slightly with added intensity. The difference in articulation-gain functions reported by Linden for high and low frequency bands of spondee words coincide with findings reported above for high- and low-pass filtered phonetically balanced words (Pollack, 1948). Linden also noted that the intelligibility curve for spondees, in spite of marked filtering, can be quite steep, and that a difference of only a few decibels can result in great differences in intelligibility scores. He suggested, therefore, that the evaluation of intelligibility with spondees presented at only one intensity level may be difficult.

Based on the above review, certain generalizations may be drawn which have direct bearing on the present research. Firstly, the frequency range around 1000-2000 Hz appears to be of great importance for intelligibility (Egan and Weiner, 1946; Kryter, 1960; Palva, 1965). Therefore, the present research employed a high frequency speech band of 1100-2200 Hz.

Secondly, it appears that the intensity level at which the stimulus material is presented determines, in part, the frequencies important to intelligibility (Miller, 1947; Egan and Weiner, 1946; Pollack, 1948). In addition, there appears to be an interaction among the level of presentation of the stimulus material, the type of stimulus material, and the frequencies important to the intelligibility of the material (Miller and Nicely, 1954; Speaks et al., 1966; Speaks, 1967). Because of this complex interrelationship among intensity level, test vocabulary, and the frequencies important to intelligibility, the low frequency speech bands employed in the present research were presented at several intensity levels.

#### The Intelligibility of Single-and Multi-Pass Band Systems

The literature reviewed in this section concerns the effects of single and multi-pass band systems on intelligibility. Specifically, these studies compare the effects on intelligibility of the presentation of a) a

high frequency speech band (single pass band system), and b) a high and low frequency speech band (multi-pass band system). Since the mode of presentation (monotic or dichotic) affects the results of such research, this variable is likewise discussed.

Matzker (1956), while devising a test for the diagnosis of central auditory disorders, presented phonetically balanced German words which were filtered into two bands; namely, a low band of 500-800 Hz and a high band of 1500-2400 Hz. At intensity levels of 30-40 dB SL, articulation scores did not exceed 26 and 30 per cent, respectively, when each band was presented alone. When both bands were presented monaurally, articulation scores rose to 90 per cent; when the high band was presented to one ear and the low band was presented to the other ear (dichotic listening condition), articulation scores rose to 97 per cent.

Linden (1964) measured articulation scores for spondee words filtered to produce two bands; namely, 560-715 Hz and 1800-2200 Hz. In contrast to the articulation scores obtained by Matzker for comparable bandwidths, Linden found that at 40 dB SL articulation scores for the low band were only 14 per cent and for the high band, 11 per cent. However, when both bands were presented dichotically, articulation scores rose to 89 per cent. No significant differences were found between articulation

scores for the dichotic listening condition and the monotic listening condition.

To examine binaural hearing aids, Huizing and Taselaar (1961) measured articulation scores for phonetically balanced words presented monotically and dichotically. The speech material was filtered to produce two octave bands, a low band of 140-280 Hz and a high band of 1128-2256 Hz. The low band alone was unintelligible to the normal ear at any level; whereas, the high band yielded a 50 per cent articulation score at an intensity level of 34 dB.<sup>3</sup> In the monotic condition, the 50 per cent articulation score was reached when the low band was presented at 75 dB and the high band at 26 dB. In the dichotic listening condition, with the low band presented at the same level (75 dB), the 50 per cent articulation score was reached when the high band was presented at 12 dB. Thus, in this situation, performance in the dichotic listening condition was superior by 14 dB to that in the monotic condition.

Similarly, Franklin (1969, 1970) presented stimulus material from the Fairbanks Rhyme Test (1958) filtered in such a way that two bands were produced; namely, a low band of 240-480 Hz and a high band of 1020-

---

<sup>3</sup>Huizing does not state the reference for his speech level; i.e., it could be 34 dB HL or 34 dB SPL.

2040 Hz. Results indicated that the presentation of the low band at 0 or 20 dB SL to the same or opposite ear as the high band improved articulation scores by as much as 21 percentage points, as compared with the presentation of the high band only. The presentation of the low band at 40 dB SL resulted in a decrease in articulation scores, as compared to scores obtained when this band was presented at 20 dB SL; scores were observed to decrease more in the monotic condition than in the dichotic condition. The advantage noted by Franklin of the dichotic condition when the low band was presented at a high intensity level is similar to that noted by Linden (1964).

One of the problems associated with reporting and interpreting the results of these experiments relates to the fact that articulation scores are not additive. Since in most of these studies intelligibility is measured by articulation scores, the comparison of the intelligibility of single and multi-pass band systems is difficult; i.e., although intelligibility (articulation scores) is observed to increase with the presentation of additional passbands, it is difficult if not impossible to quantitatively determine the amount of this improvement, because articulation scores cannot be arithmetically combined.

French and Steinberg (1947) drew attention to this problem, and it was for this reason that they devised and

advocated the use of the articulation index (AI),<sup>4</sup> as opposed to articulation scores, as a measure of intelligibility. Specifically, they noted that the articulation score "is not an additive measure of the importance of the speech components in the different frequency regions (and thus) for the purpose of establishing relationships between the intelligibility carrying capacity of the components of speech and their frequencies and intensities, a more fundamental index free of the above defects is needed." In this way, an articulation index (AI) was devised which "is based on the concept that any narrow band of speech frequencies of a given intensity carries a contribution to the total index which is independent of the other bands with which it is associated and that the total contribution of all bands is the sum of the contribution of the separate bands."

For devising the AI, CVC nonsense syllables were filtered into 20 filter bands covering the range 250-7000 Hz. When each of the bands is presented at its optimal level, it contributes equally (5 per cent) to the AI; thus the presentation of all 20 bands at their optimal levels results in an AI = 1.00, which is perfect intelligibility.

---

<sup>4</sup>The articulation index is a computational device designed to supplement or replace articulation testing. According to Stevens (1951), this procedure "involves weighting the contribution to intelligibility of the speech components as a function of their frequency and intensity."

Since perfect intelligibility can be achieved with frequencies limited to the range 250-7000 Hz, there is the implication that frequencies outside this range (below 250 Hz and above 7000 Hz) have little importance for intelligibility. Further, as noted above, for maximum contribution to the AI (5 per cent), each band has to be presented at its optimal level. If the band is presented below this level, its contribution to the AI is less than 5 per cent. This implies that as the intensity of the bands is increased up to their optimal level intelligibility of the bands also increases.

Though not stated above, a third assumption of the AI is that articulation scores and the AI are monotonic functions; i.e., an increase in the AI implies an increase in articulation scores.

Several investigators have investigated the applicability of the AI in various experimental situations. The following review serves to point out many of the variables affecting the validity of the AI and several limitations imposed upon its use.

Black (1954) compared the 20 bands contributing to the AI for nonsense syllables (French and Steinberg, 1947) with those which he determined for phonetically balanced words in a) a multiple choice test, and b) a write down test. Results of the write down test were comparable to those reported by French and Steinberg.

The multiple choice test yielded slightly different results. From the lowest band (150-230 Hz) to the 1500-1700 Hz band, the multiple choice test yielded an AI of .5, while the results of French and Steinberg for a comparable frequency range (250-375 Hz to 1515-1720 Hz) were .40 to .45. The frequencies above 3500 Hz contributed more to the AI in the French and Steinberg study than in the multiple choice test, where their contribution was 10-20 per cent. It thus appears that the type of task involved (multiple choice or write down task) has an influence on the relative contribution of the high and low frequency bands to the AI; in this situation, the multiple choice task appears to rely more heavily on the low frequencies than does the write down task.

Hirsh et al. (1954) suggested that the type of stimulus material was another variable that affects the validity of the AI. Specifically, they noted that "the articulation index, which is calculated from physical characteristics alone, will predict best the intelligibility of items that depend exclusively on physical characteristics, e.g., nonsense syllables." They further noted that "the intelligibility of words is a function of physical characteristics plus such factors as word probability in the language, and similar contextual clues."

Kryter (1956), discussing some of the problems associated with the AI, pointed to Miller and Nicely's

findings (1955) "that speech articulation consists of 4 or 5 independent acoustic factors which are differently related to frequency and noise masking." Kryter felt that these findings "undoubtedly are the cause of some of the unexpected results in attempting to predict both the effects of frequency distortion and noise masking by dividing, as do the 20-band and octave-band methods,<sup>5</sup> the frequency scale into a fixed number of bands of fixed width and location on the frequency scale."

Palva (1965), commenting on the use of formulae such as the AI for predicting intelligibility, noted that because of the number of variables involved in intelligibility testing the use of formulae may be impossible and that intelligibility may have to be defined experimentally. Palva believed that this was particularly true when two bands from different parts of the speech spectrum are presented simultaneously (multi-pass band system).

Kryter (1962), validating the use of the AI as a predictor of intelligibility, noted several instances where the AI was not applicable. Most important among them for the present review were multi-pass band systems, i.e., the AI was found to underestimate the intelligibility of the multi-pass band systems and therefore to be

---

<sup>5</sup>The 20-band and octave-band methods are two procedures for calculating the AI.

inappropriate for use with such systems. Similarly, in a paper by Flanagan and Levitt (1969), the AI was reported not to be a satisfactory predictor of intelligibility when speech contained irregularities in its spectrum i.e., when the peaks and valleys exceeded, on the average, 18 dB per octave. This situation exists in most multi-pass band systems, and as such, precludes the use of the AI as a means of predicting the intelligibility of such systems. In view of the limitations of the AI specified above, the writer felt that this procedure was not appropriate for this study.

After reviewing the studies concerning the effects of multi-pass band systems on intelligibility, the writer felt that the presentation of a low frequency speech band in combination with a high frequency speech band contributed positively to intelligibility. However, as Kryter (1960) indicated, it appears that there is not enough information available to show which frequency bands combine most efficiently for maximum contribution to intelligibility. Therefore, the present research was designed to further investigate this by pairing each of three low frequency speech bands (50-100, 110-220, and 200-400 Hz) with a high frequency speech band (1100-2200 Hz).

Further, in reviewing these studies, the writer noted that the mode of presentation (monotic or dichotic) affected the intelligibility of multi-pass band systems.

The data reported in the literature concerning the relative efficacy of the monotic and dichotic modes of presentation appear contradictory. Where Matzker (1956) and Huizing and Taselaar (1961) found an improvement in intelligibility when speech bands were presented dichotically rather than monotically, Linden (1964) found no difference in intelligibility for speech bands so presented. Consequently, the present research attempted to further investigate the effects on intelligibility of the monotic and dichotic modes of presentation.

#### The Effects of Ear Dominance on Intelligibility

Ear dominance refers to a finding of some researchers that in complex listening situations the right ear yields significantly higher articulation scores than the left ear for verbal material. The present research, which likewise involved complex listening situations (filtering conditions), attempted to further investigate the effects of ear dominance.

Licklider (1948), while investigating the effects of interaural phase relations on the masking of speech, was one of the first researchers to note asymmetry between the right and left ears with respect to intelligibility. Using binaurally presented white noise (at 80 dB SPL) as the masker and monaurally presented phonetically balanced words (at -10 dB s/n) as the speech signal, Licklider found that average articulation scores were higher when

the speech signal was presented to the right ear than when presented to the left ear.

This phenomenon of asymmetry between the ears, with respect to intelligibility, was not further studied until Kimura (1961a, 1961b, 1963) and Milner (1962) began investigating the effects of unilateral temporal lobe lesions on auditory perception. Kimura (1961a) presented digits in such a way that each ear simultaneously received a different auditory stimulus (dichotic mode of presentation). Results indicated that unilateral temporal lobe lesions were accompanied by lower articulation scores in the ear contralateral to the side of lesion. In addition, Kimura found that subjects with lesions in the left temporal lobe had scores which were significantly lower than subjects with lesions of the right temporal lobe. She interpreted these findings to mean that "although each ear sends fibers to the auditory cortex on both sides, the crossed connections are more important than the uncrossed." Each temporal lobe was said to receive maximum contribution from the contralateral ear. Since the left hemisphere in most cases "plays a more important part than the right in the perception of spoken material," it was suggested that the right ear advantage for perceiving verbal material is a consequence of the dominant left hemisphere receiving "stronger or more numerous" crossed fibers from the right cochlea.

Kimura (1961b) further supported this hypothesis in another study in which she compared articulation scores from the right and left ears in subjects with speech represented in the right hemisphere.<sup>6</sup> Results indicated that articulation scores were higher for verbal material arriving at the left ear.

In the studies described above, the stimulus materials were verbal in nature. Subsequent studies attempted to investigate asymmetry of auditory perception for nonverbal material. Kimura (1963) presented two kinds of auditory materials to normal subjects. In one case, digits were presented as the stimulus material and in the other case, melodies. Both conditions involved a dichotic listening situation. Results indicated that digit scores were higher for the right than the left ear, which is consistent with previous findings (Kimura, 1961a, 1961b). When melodies were presented, scores were higher for the left ear. Kimura believed "these findings were related to the different roles of the right and left hemispheres of the brain in verbal and nonverbal preception."

Kimura's conclusion had been supported in a previous study by Milner (1962), who investigated the

---

<sup>6</sup>Sodium amytol was used to select subjects with speech represented in the right hemisphere. When injected into one of the carotid arteries, it temporarily disrupts the functioning of that hemisphere (Wada and Rasmussen, 1960). Dysphasia occurs only after injection of the dominant hemisphere.

effects of unilateral temporal lobe lesions "to determine whether left temporal lobe lesions produced a general impairment of auditory discrimination or whether the verbal or nonverbal nature of the task was the critical factor." Using various subtests of the Seashore Measures of Musical Talents, subjects with right and left temporal lobe lesions were studied. Results indicated that after right temporal lobectomy there was a decrease in scores, especially on the timber and tonal-memory subtests. Such deficits did not occur after left temporal lobectomy.

The difference in the perception of verbal and nonverbal material was again emphasized in a study by Shankweiler and Studdert-Kennedy (1967), who compared recognition of CV syllables and steady state vowels in a dichotic listening situation. The right ear gave significantly higher scores for the CV syllables but not for the steady state vowels. Shankweiler and Studdert-Kennedy observed that "the finding of a left ear advantage for musical melody recognition, as against a right ear advantage for spoken digits, the neutral status of steady state vowels, midway as it were, between speech and music, is perhaps not surprising." In addition, the evidence strongly suggested that the left hemisphere dominance for speech perception operated at the speech sound level. Laterality was greater for CV pairs differing in two articulatory features (voicing and place of articulation) than for pairs differing in only one

feature. This suggests that the perception of CV syllables involves a process of analysis by feature.

In a later study, Studdert-Kennedy and Shankweiler (1970) confirmed and extended their previous results. The speech stimuli were spoken CVC syllables formed by pairing each of the six English stops with various vowels, all presented dichotically. For the initial and final stop consonants, significantly higher articulation scores came from the right ear. Analysis of the correct and incorrect responses confirmed their previous findings that the consonantal features of voicing and place of articulation were perceived independently and significantly better in the right ear. As had Kimura (1961a), Studdert-Kennedy and Shankweiler attributed the right ear advantage to left cerebral dominance and to functionally stronger contralateral pathways. They further suggested that "while the general auditory system may be equipped to extract the auditory parameters of the speech signal, the dominant hemisphere is specialized for the extraction of linguistic features from those parameters."

Higher articulation scores have been found for verbal material arriving at the right ear when a message stimulating one ear is in competition with another message stimulating the other ear. Franklin (1969, 1970) has also demonstrated a right ear advantage for verbal material in another complex listening situation, i.e., in a fusion condition. She presented a high frequency speech band

(1020-2040 Hz), which was believed to contain a great deal of consonant information, to one ear and a low frequency speech band (240-480 Hz), which theoretically contains little consonant information, to either the same or the opposite ear. Average articulation scores were higher when the high frequency speech band (the band containing a great deal of consonant information) was presented to the right ear than when it was presented to the left ear.

Bocca (1955) monaurally presented low-pass (1000 Hz) filtered Italian disyllables to subjects with unilateral, temporal lobe lesions. Results indicated that articulation scores were significantly lower in the ear contralateral to the side of lesion, which is consistent with findings reported above (Kimura, 1961a). However, inconsistent with above findings, comparable losses were found for cases of both right and left temporal lobe lesions. Bocca felt that this finding did not support the claim of "the prevalence of one hemisphere for the reception and expression of speech." In addition, the loss in the contralateral ear, which had been found to exist pre-operatively, disappeared 20-30 days after lobectomy. This too is inconsistent with the findings reported by Kimura (1961a). However, Kimura tested patients approximately two weeks after lobectomy, and therefore, it is possible that the results were confounded by direct effects of the operation.

Calearo and Antonelli (1963) monaurally presented speech materials to normal subjects and to subjects with unilateral temporal lobe lesions. The materials were degraded by filtering (low-pass, 800 Hz) and by interruptions (up to 20 interruptions per second). Results for normal subjects showed no significant differences in articulation scores for the right and left ear. However, subjects with unilateral temporal lobe lesions showed a loss in the ear contralateral to the side of lesion, and this loss was of the same magnitude for lesions of either the right or left hemisphere. This finding agrees with that reported above by Bocca (1955).

The studies reviewed above tend to support many researchers (Jerger, 1960; Ohta et al., 1960; Hayashi et al., 1966) who distinguish between binaural separation and binaural fusion tasks saying they represent two different binaural mechanisms. In the binaural separation task, the right ear advantage for speech perception has been quite consistently demonstrated (Kimura, 1961a, 1961b, 1963; Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970). In the binaural fusion condition, results appear uncertain. To the writers knowledge, Franklin (1969, 1970) is the only investigator who has reported a right ear advantage in a fusion condition. The present research attempted to further investigate the right ear advantage in a fusion situation.

### Techniques for Measuring Intelligibility

This section concerns techniques for measuring intelligibility. Rather than being an exhaustive review of intelligibility testing materials and procedures, this section discusses the most commonly used techniques for collecting intelligibility data and indicates why the present writer chose the particular procedures and test vocabulary used in her research.

A great number of experimenters have studied the effects of various kinds of distortion on speech intelligibility. Results are usually obtained by articulation testing. As stated in Chapter I, this technique involves manipulating the variable of interest and determining the effect of this manipulation by measuring the per cent of correct responses, i.e., the articulation score. Some researchers, however, have used the threshold of perceptibility as the measure of intelligibility. To obtain the threshold of perceptibility, the interfering signal is set at a fixed intensity, and the intensity of the stimulus is so adjusted that the listener can, with effort, understand the material to be received.

Egan, Carterette, and Thwing (1954) compared articulation testing and the threshold of perceptibility as measures of intelligibility. In their study, the intelligibility of a speech message presented in the presence of a similar interfering speech signal was measured both by

articulation testing and by the threshold of perceptibility. In various experimental conditions, high- and low-pass filtering of a) the speech message or b) the interfering signal were employed. Since the similarity between the two signals in this situation was known to be an important factor in intelligibility, the authors anticipated that the elimination of low frequency components (high-pass filtering) in either the speech message or the interfering signal would improve intelligibility. Using the threshold of perceptibility as the measure of intelligibility, this prediction was not supported. Using articulation testing as the measure of intelligibility, the authors showed that high-pass filtering of either the speech message or the interfering signal yielded significantly greater intelligibility. It was suggested that since the intensity differences between the two signals is an important aid to intelligibility, the threshold of perceptibility was probably "too insensitive to measure the advantage of filtering."

Many different test vocabularies have been used in intelligibility testing, Hirsh et al. (1954) showing that intelligibility scores obtained with one vocabulary cannot necessarily be generalized to another. When deciding on the vocabulary to use in the present research, various kinds of test materials were considered.

Nonsense syllables (CV and CVC forms) have been used quite extensively for intelligibility testing (French

and Steinberg, 1947; Miller and Nicely, 1955). Lehiste and Peterson (1959) questioned, however, the value of using nonsense syllables for communication measuring, noting that "phonemes within a language are characterized not only by their phonetically distinctive features, but also by their distributional properties." They further indicated that the composition of nonsense syllables "by definition does not follow the normally allowed sequences of a language, and thus the sounds uttered in nonsense syllables are not appropriate representations of the phonemes of the language." Because of the great influence of adjacent speech sounds on intelligibility, "the acoustical characteristics of vowels and consonants within nonsense syllables are indeed different from those within normal linguistic sequences." Lehiste and Peterson suggest instead the use of sentences and longer linguistic units for measuring intelligibility.

But Egan (1948) noted a number of drawbacks in the use of sentences for intelligibility testing. Most important among them, for the present research, is the influence of what Egan calls psychological factors that make the results difficult to analyze and interpret. To avoid this difficulty and yet to follow the suggestion of Lehiste and Peterson, many experimenters have employed the technique of a carrier phrase, which has the additional advantage of preparing the subject for the presentation of the test item and thus reducing the variability in artic-

ulation scores which result from inattention and distractability (Egan, 1948).

The phonetically balanced word list is another widely used test of intelligibility. Lehiste and Peterson (1959) stated that on the basis of spectrographic analysis, "it is not possible to have lists of words that are phonetically balanced." They note that it is possible to have lists of words that are, in certain respects, phonemically balanced,<sup>7</sup> but that analysis of the Harvard PB word lists, for example, shows they do not meet this criterion. For this reason, Lehiste and Peterson devised their own CVC word lists which are more nearly phonemically balanced.

In 1958 Fairbanks devised the Rhyme Test, which is an open response CVC test.<sup>8</sup> In this test, the response is confined to the initial consonant and consonant-vowel transition. With respect to the construction of the test,

---

<sup>7</sup>According to Lehiste and Peterson (1959), phonetics is concerned with the physiological and acoustical properties of speech; whereas, phonemics or "perceptual phonetics" is concerned with the listeners' linguistic backgrounds and its effect on the listeners' perception of speech.

<sup>8</sup>In an open response test, the only limit placed on the responses a subject can give is the subject's vocabulary. This contrasts with a closed response test in which the subject is given a set of choices which restrict him more than would his vocabulary or the statistical properties of the language.

Fairbanks notes that "the transitions are adequately represented and there are no undue biases."

House, Williams, Hecker, and Kryter (1965) devised a closed response CVC test similar in format to the Fairbanks Rhyme Test. As explained by the authors, it differs from the Rhyme Test in that no given test is phonetically balanced, nor is word familiarity taken into account. The authors note that the test provides several advantages for intelligibility testing, the most important among these being that the test forms are phonetically similar and represent the major classes of speech sounds. Therefore, the responses to the material can be analyzed in detail for diagnostic appraisal.

The information contained in the foregoing section provided the present writer with a basis for selecting the method of intelligibility testing and the test vocabulary. Since filtering and intensity were important variables in this research, in accordance with the suggestion of Egan, Carterette, and Thwing (1954), intelligibility was measured by articulation testing, as opposed to the threshold of perceptibility. The Fairbanks Rhyme Test was the vocabulary used, because it samples consonants and consonant-vowel transitions. Since it has been shown that consonant-vowel transitions contain cues for consonant perception (see pp. 41-50), the present writer felt that this factor should be considered. Consonant discrimination was of interest in the present research because consonants

contain major information about the speech message. In addition, it is often assumed that the low frequencies of speech are of negligible importance for consonant perception. The validity of this assumption is tested in the present research.

### Cues for Consonant Discrimination

This section concerns the acoustic cues for consonant perception. Studies involving synthetic speech and real speech are discussed. The results of these studies demonstrate the importance of consonant-vowel transitions for consonant discrimination. Consequently, the Fairbanks Rhyme Test (1958) was the stimulus material employed in the present research.

In addition, the studies reviewed in this section indicate that much of the information for consonant discrimination is located in the higher frequency regions (second and third formant transitions and high frequency noise bursts). Accordingly, the presentation of the low frequency speech bands in the present research should contribute little to intelligibility, i.e., consonant perception, in this situation. This is one of the null hypotheses that the present research was designed to investigate.

It has long been suggested, based on spectrographic analysis of voiceless plosives, that one of the cues for identifying the place of articulation is the

position along the frequency scale of the noise burst which characterizes the voiceless plosives. In an attempt to investigate this hypothesis, Cooper et al. (1952) presented a series of synthetic patterns consisting of a burst-plus-vowel. The frequency of the noise burst was systematically varied. Results indicated that the frequency position of the noise burst provided the cue for distinguishing among /p,t,k/. It was found that with all vowels high frequency bursts were heard as /t/. Bursts at lower frequencies were perceived as /k/ when at the same frequency or when slightly above the frequency of the second formant of the vowel,<sup>9</sup> otherwise, they were heard as /p/. This indicates that for /p/ and /k/ identification depends "not solely on the frequency position of the burst of noise, but rather on this position in relation to the vowel." The authors therefore noted that "the perception of these stimuli, and also, perhaps, the perception of their spoken counterparts, requires the consonant-vowel combination as a minimal acoustic unit."

This observation led to further investigation of the consonant-vowel transitions to determine whether these transitions are important cues for the perception of the

---

<sup>9</sup> Liberman et al. (1954) defines a formant as "a relatively high concentration of acoustic energy within a restricted-frequency region."

consonant-vowel combination (Cooper et al., (1952).<sup>10</sup>

Using synthetic consonant-vowel patterns, the voiced and voiceless plosives were presented in combination with various vowels. The frequency of the second formant transition was systematically varied. Results indicated that a rising second formant transition was heard as /b/ or /p/; whereas, a falling transition tended to be perceived as /g,k/ or /d,t/, depending upon the vowel. Thus the authors concluded that "the cognate relationship between /b,d,g/ and /p,t,k/ are effectively cued by the second formant transition."

Using the synthetic patterns described above, a second experiment was performed with the nasal consonants /m,n,ŋ/ (Cooper et al., 1952). Results tended to parallel those reported for plosives; i.e., the second formant transitions, which were regularly heard as /b,p/ in the preceding experiment, were now perceived as /m/, and those previously perceived as /d,t/ or /g,k/ were now perceived as /n/ or /ŋ/, respectively.

These findings were confirmed by Liberman et al. (1952), Liberman et al. (1954), and Delattre et al. (1955). The locus for the second formant transition for /p,b,m/ was located at approximately 720 Hz, for /t,d,n/

---

<sup>10</sup>The term transition refers to the shift in frequency of formants in the region where the consonant and vowel join. This frequency shift is presumably related to movement of the articulators.

at 1800 Hz, and for /k,g,ŋ/ the locus of the second formant transition was observed to depend on the vowel environment. In addition, the first formant transition was identified as an important cue for the manner of articulation and was located between 0-240 Hz for the plosives and nasals.

More recently, however, Menon et al. (1969), using spectrographic analysis of real speech, further investigated the importance for /p,t,k/ of the frequency of the noise burst. These consonants were placed in VCC syllables consisting of the consonant clusters /sp, st, sk/ in combination with various vowels. Menon et al. found that "a clear acoustic distinction between /t/ and /p/ or /k/ may be found in the burst frequency." For the alveolar stop, the energy of the noise burst was always above 3000 Hz; whereas, for the bilabial and velar plosives, the bursts were approximately at 1350 and 1750 Hz, respectively. Though these results are not inconsistent with previous findings, these investigators seem to be implying that the frequency of the noise burst is sufficient for the discrimination of these consonants. This interpretation does not appear consistent with previous findings.

The second formant transitions for nasal consonants were further investigated in a series of experiments by Malecot (1956). These studies were undertaken to determine whether the results reported above for synthetic speech

could be reproduced when using real speech. The stimulus materials were nasal consonants presented in a) unaltered CV and VC syllables, b) CV and VC syllables composed exclusively of steady-state vowels (no transitions), and c) isolation.

For the unaltered syllables, /m/ was most often correctly identified in the initial position, with /n/ following. In the final position, all three consonants were equally well identified. For the stimuli composed of steady-state vowels, there was a loss in identifiability as compared with that for the unaltered syllables, Malecot attributing this loss to the absence of transitions. The perception was, however, not entirely random; in the initial position, the /m/ was most often correctly identified, and in the final position, correct /n/ identifications increased. When the nasal resonances were presented in isolation, results corresponded with those reported for the steady-state vowels. Malecot felt that these findings suggest "that some of the resonances may themselves contain a small amount of place information."

This hypothesis led to a second experiment which was designed to compare the relative contribution of resonances and transitions to the identification of the place of articulation (Malecot, 1956). In this experiment, each of the /m,n,ŋ/ resonances was paired with different nasal second formant transitions. Results indicated that a "crude quantitative evaluation of the relative

importance of the resonances and transitions reveals that the two are about equally important." Further, data showed that "of the three sets of resonances, /m/ would contain the greatest amount of place information, with /n/ following, and /ŋ/ so weak as to be neutral."

To further investigate the cues provided in the nasal resonances, Malecot paired each of the /b,d,g/ transitions with each of the /m,n,ŋ/ resonances. Results indicated that the information provided by the transitions dominated in all cases; i.e., stimuli containing /b/ transitions were heard as /m/, regardless of the set of resonances used. Those containing /d/ transitions were heard as /n/, and those containing /g/ transitions were heard as /ŋ/. However, in the initial position, and to a somewhat greater extent in the final position, inappropriate resonances did detract slightly from the results given above.

From his series of studies, Malecot concluded that the transitions of adjoining vowels serve as important place cues for the nasal sounds. This confirmed, for real speech, what was discovered in synthetic speech experiments (Cooper et al., 1952; Liberman et al., 1952; Liberman et al., 1954; Delattre et al., 1955). Further, the successful substitution of /b,d,g/ transitions for those of /m,n,ŋ/ support previous observations (Cooper et al., 1952), showing that the second formant transitions are the same for both nasals and plosives, that "they

constitute the acoustic similarities corresponding to the articulatory similarities of cognates" (Malecot, 1956).

Further, Malecot concluded that the nasal resonances serve primarily as cues to the manner of articulation but that they contain some information concerning the place of articulation, especially in the syllable final position. Given his results, Malecot felt that "nasal consonants should be classified as stops rather than as continuants."

Liberman, Delattre, and Cooper (1958) investigated cues for the discrimination of voiced and voiceless stops in the initial position. Specifically, their study concerned the elimination of portions of the first formant transition and its effect on the voiced-voiceless distinction. The stimulus materials were synthetic CV syllables consisting of /b,d,g/ followed by various vowels. The duration of the first formant transitions was varied in 10 msec. steps. Results indicated that by eliminating the beginning of the first formant transition initial voiced stops could be made to sound like their voiceless counterparts. It was found that the voiced alveolar stop /d/ required the removal of the greatest amount of acoustic energy to be heard as its voiceless counterpart /t/. The velar /g/ required slightly less, while the labial /b/ with relatively small amounts of the first formant removed was perceived as its voiceless counterpart /p/. In addition, it was found that neither

the removal of the fundamental frequency nor the introduction of noise in the second and third formant transitions of the voiced plosives was a sufficient cue for the perception of voicelessness. The authors concluded that the duration of the first formant was a sufficient cue and an important cue for distinguishing voiced from voiceless stops.

It has long been known that the fundamental frequency carries important cues for the prosodic features of speech. However, Haggard et al. (1970) found that the fundamental frequency may likewise be important for the discrimination between voiced and voiceless consonants. Using synthetic CV syllables, results indicated "that, for about 90 per cent of the subjects, the pitch change in the vowel can cue the voiced/voiceless distinction for a preceding stop consonant in English." Specifically, a low rising pitch curve at the onset of voicing led to the perception of the initial stop as voiced; whereas, a high falling pitch contour led to the perception of voicelessness. Control conditions indicated that this cue depended on the pitch sensation and not on the low frequency energy content. It appears that further investigation of the importance of the fundamental frequency for segmental discrimination may prove fruitful.

O'Connor et al. (1957) and Lisker (1957) independently investigated cues for the discrimination among /w,r,l,y/ in synthetic speech. Results indicated

that the first, second, and third formant transition were important cues for discriminating these sounds. The transition of the first formant was located at 240 Hz for /w,y/ and at approximately 360 Hz for /l/. The second formant transition was found to be important for discriminating between /w/ and /r,l/, between /w/ and /y/, and between /r,l/ and /y/; the discrimination between /r/ and /l/ depended upon the third formant transition, this transition being higher for /l/ than for /r/.

Hughes and Halle (1956) and Harris (1958) investigated cues for the discrimination among the fricative sounds in CV syllables. The studies paralleled each other and arrived at similar conclusions. Two variables were studied; namely, friction noise and formant structure. These variables were investigated by combining the friction noise of one spoken fricative vowel syllable with the vocalic portion of another syllable. Results indicated that the important cues for differentiating /s,ʃ/ from /f,θ/ are given by the friction noise. Discrimination between /s/ and /ʃ/ was likewise based on the friction noise. However, discrimination of /f/ from /θ/ was found to be based on the vocalic portion of the syllable. Similar results were also obtained for the voiced counterparts of these sounds.

This section has been concerned with the acoustic cues for consonant discrimination. The results of these studies contributed to the writers decision to employ the

Fairbanks Rhyme Test as the stimulus material in the present research. This decision was based, in part, on the evidence in this section about the importance of the consonant-vowel transition for consonant discrimination.

In addition, the research reviewed in this section indicated that much of the information for consonant discrimination is located in the higher frequency regions (second and third formant transitions and high frequency noise bursts). Thus, in the present research, the presentation of the low frequency speech bands should have a negligible effect on intelligibility. The present research was designed to test this hypothesis.

## CHAPTER III

### METHODS AND PROCEDURES

#### I. Preparation of Stimulus Materials

##### A. The Master Tape Recordings

The Fairbanks Rhyme Test (1958) was the stimulus material in the present research. Spoken by a male speaker in a normal voice, the five test lists were recorded (Magnecorder, model 1028, half-track) on magnetic tape in a sound treated room. The microphone (Sony Condensor, model C-37A) was approximately 12 inches from the speaker's lips. Each word was preceded by the carrier phrase "the next word is...." An interval of 5 seconds separated each utterance. Monitored on the VU meter of the recorder, the test words peaked between -3 and +2 VU. The recording speed was 7 1/2 inches per second.

##### B. The Test Tape Recordings

The master tape recordings were filtered and re-recorded to produce the test tapes. The filter bands used were 1100-2200, 50-100, 110-220, and 200-400 Hz.<sup>1</sup> These bandwidth values were selected because they are made up of

---

<sup>1</sup>All bandwidths are stated as 3 dB bandwidths, i.e., the cut-off frequencies were defined 3 dB below the peak values.

U.S. standard 1/3 octave filters. To produce the test tapes, the master tapes were played back on an Ampex AG 500 Tape Reproducer through an adjustable 1/3 octave band filter (General Radio, model 1925) and re-recorded (Ampex AG 500) such that the high frequency band (1100-2200 Hz) was recorded on track I and one of the low frequency bands (50-100, 110-220, and 200-400 Hz) was recorded on track II. See Figure 1.

The levels used in making the test tapes were determined by use of a calibration tape which will be described in the Calibration Section. During the re-recording, the gain was increased or decreased to insure that the peak levels of all test words were within the range -3 and +2 VU.

## II. Basic Instrumentation

### A. Apparatus Used in Preparing the Test Tape Recordings

Figure 1 is a block diagram showing the apparatus used in making the test tape recordings. The master recordings were played back on the Ampex AG 500 Tape Recorder through the General Radio, model 1925, Multi-filter and adding network (which had the effect of combining the 1/3 octave bandpass filters) and re-recorded on a second Ampex AG 500 Tape Recorder.

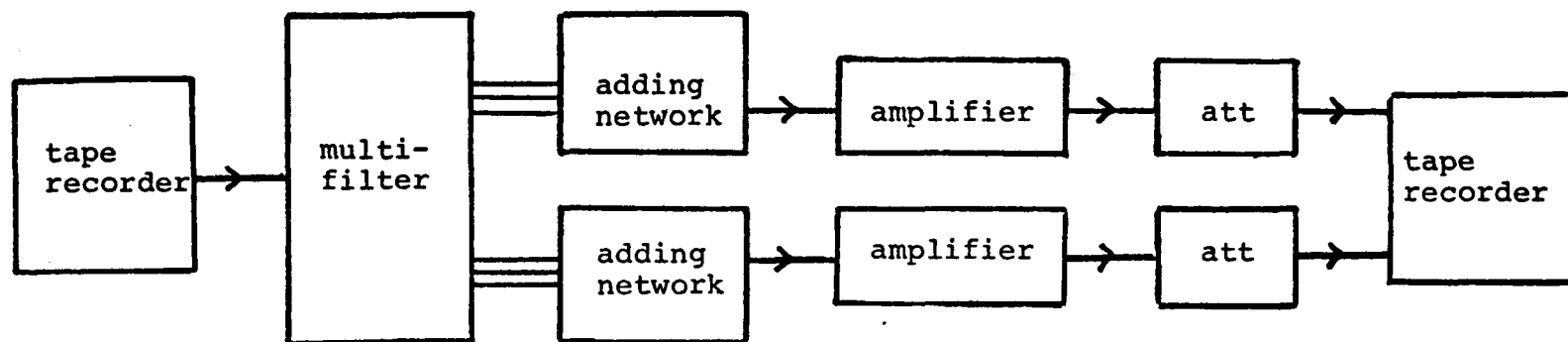


Figure 1.--Block diagram of apparatus used in making test tape recordings.

## B. Apparatus Used to Present the Stimulus Material

Figure 2 shows the apparatus used for presenting the stimulus material to the experimental subjects. A Magnecorder, model 1028, Tape Reproducer was used. Track I carried the high frequency band and track II the low frequency band. Track II was played through a Krohn-Hite, model 3202, Filter to eliminate unwanted noise energy outside the pass band that might be present on the tape. Because the low frequency bands were presented at such high levels, noise outside the pass band on the tape would have interfered with the high frequency speech band on Channel I. Hence, the bandpass filter was necessary. The skirts of the filter dropped off at a rate of 24 dB per octave. The 3 dB bandwidths of the Krohn-Hite Filter were 50-100, 100-200, and 200-400 Hz, respectively.

A two channel audiometer (Allison 22) and attenuators were employed to achieve level control of the stimulus material. The subjects heard the stimulus material via two channel headphones, TDH-39 with (MX-41/AR) cushions.

## III. Calibration

### A. Calibration Procedures for Recording the Master Tapes

The frequency response characteristics of the system used in recording the master tapes were determined at the frequencies of interest, as shown in Figure 3.

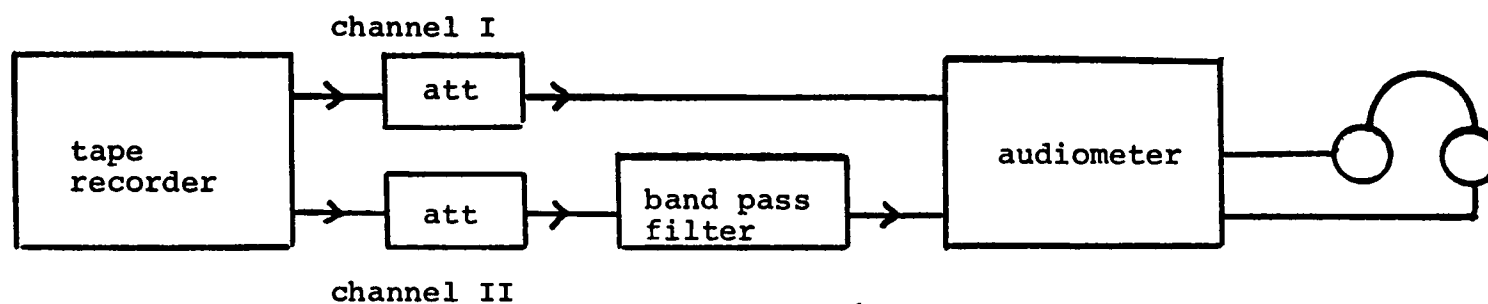


Figure 2.--Block diagram of the apparatus used for presenting the stimulus material.

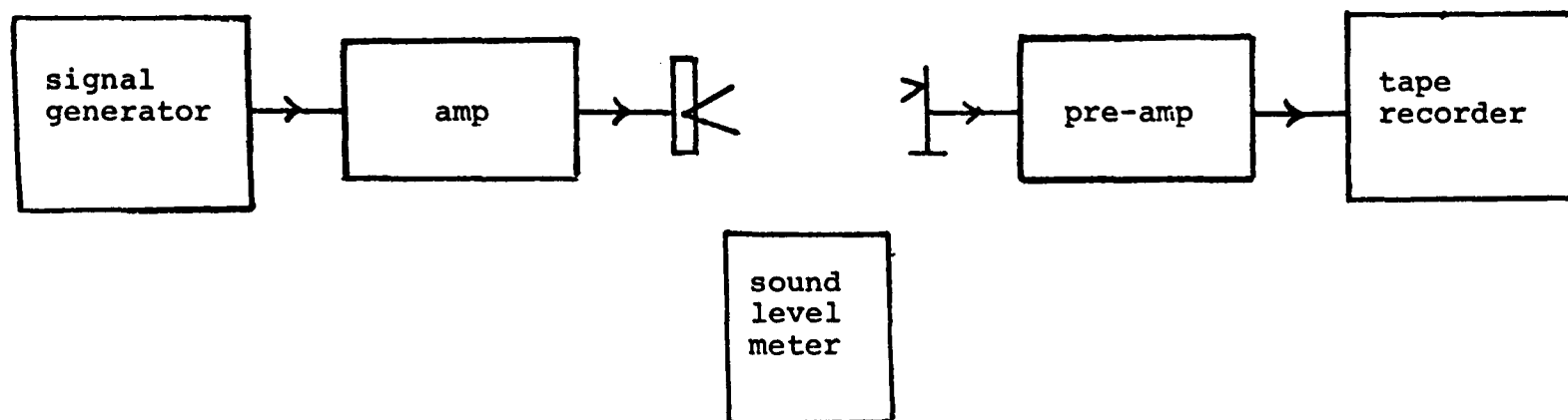


Figure 3.--Block diagram of apparatus used to determine the frequency response characteristics of the system used in recording the master tapes.

A loudspeaker was placed in the same location as the person who recorded the master tapes. The microphone (approximately 12 inches from the loudspeaker), preamplifier, and tape recorder were located as they had been for recording the master tapes. A Bruel and Kjaer Sine Wave Generator, type 1024, was used to produce energy at frequencies which represented the center frequencies of the filter bands used in the experiment; namely, 80, 160, 315, and 1600 Hz. The sound level produced by the loudspeaker was adjusted to equal 90 dB SPL at the position of the microphone. Acoustic levels were measured by a Bruel and Kjaer, model 2203, Sound Level Meter.

The gain of the recorder (Magne recorder, 1028) was adjusted to +2 VU when the sine wave generator produced 1600 Hz. The other three frequencies were then recorded without change in the equipment described above. The procedure was then repeated with the level again set at 90 dB SPL and the VU meter peaking this time at +3 VU for the 1600 Hz sine wave.

The tape recordings were played back and the output measured at the output terminals of the recorder. Average losses at 80, 160, and 315 Hz were 32, 23, and 14 dB, respectively. These losses were compensated for during the experiment. This was done to insure that the frequency response of the entire system was effectively flat within 3 dB.

## B. Calibration Procedures for Recording the Test Tapes

Figure 4 shows the apparatus employed to calibrate the system used for recording the test tapes. The Bruel and Kjaer Sine Wave Generator, type 1024, produced energy at the center frequencies of the filter bands. This sinusoidal energy was played through the system (see Figure 4) and the output plotted, using a Bruel and Kjaer Level Recorder, model 2305.

Figures 5, 6, 7, and 8 show the frequency response characteristics of the system which produced the test tapes. The General Radio 1/3 Octave Multi-filter, model 1925, was adjusted to produce octave bands, the center frequencies of which were 80, 160, 315, and 1600 Hz. The skirts of the filters fell at the rate of approximately 36 dB per octave. As a result of adding the 1/3 octave filter bands, slight ripples, on the order of  $\pm 1$  dB, occurred in the pass bands.

A similar procedure was used to produce a calibration tape containing 80, 160, 315, and 1600 Hz, each recorded for 30 seconds at -2 dB.<sup>2</sup>

## C. Calibration Procedures of the Playback System Used to Present the Stimulus Material

### 1. Calibration of the Gain of the Playback System

---

<sup>2</sup>This level was chosen because it produced the best signal-to-noise ratio.

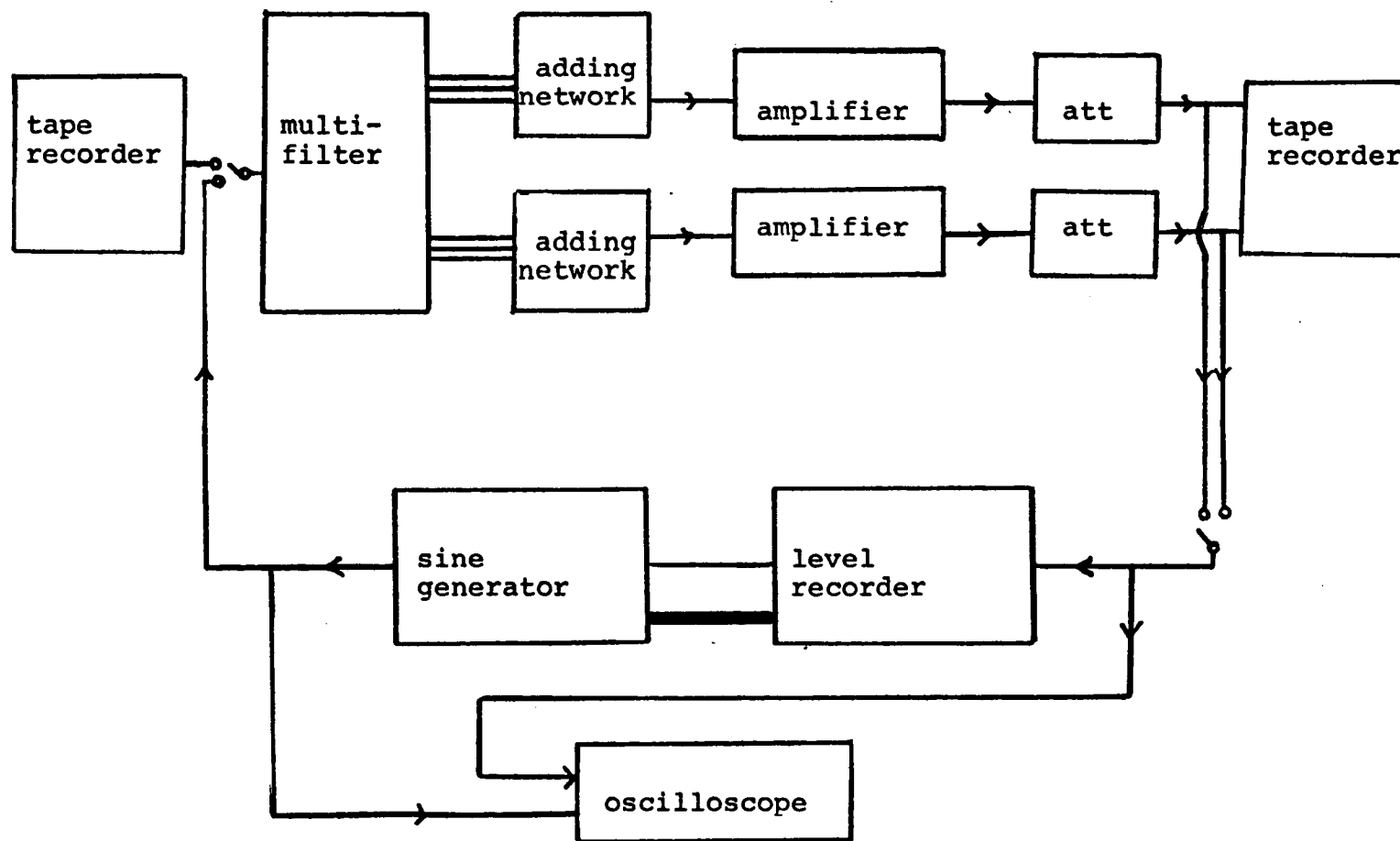


Figure 4.--Block diagram of apparatus used to calibrate system used in making test tape recordings.

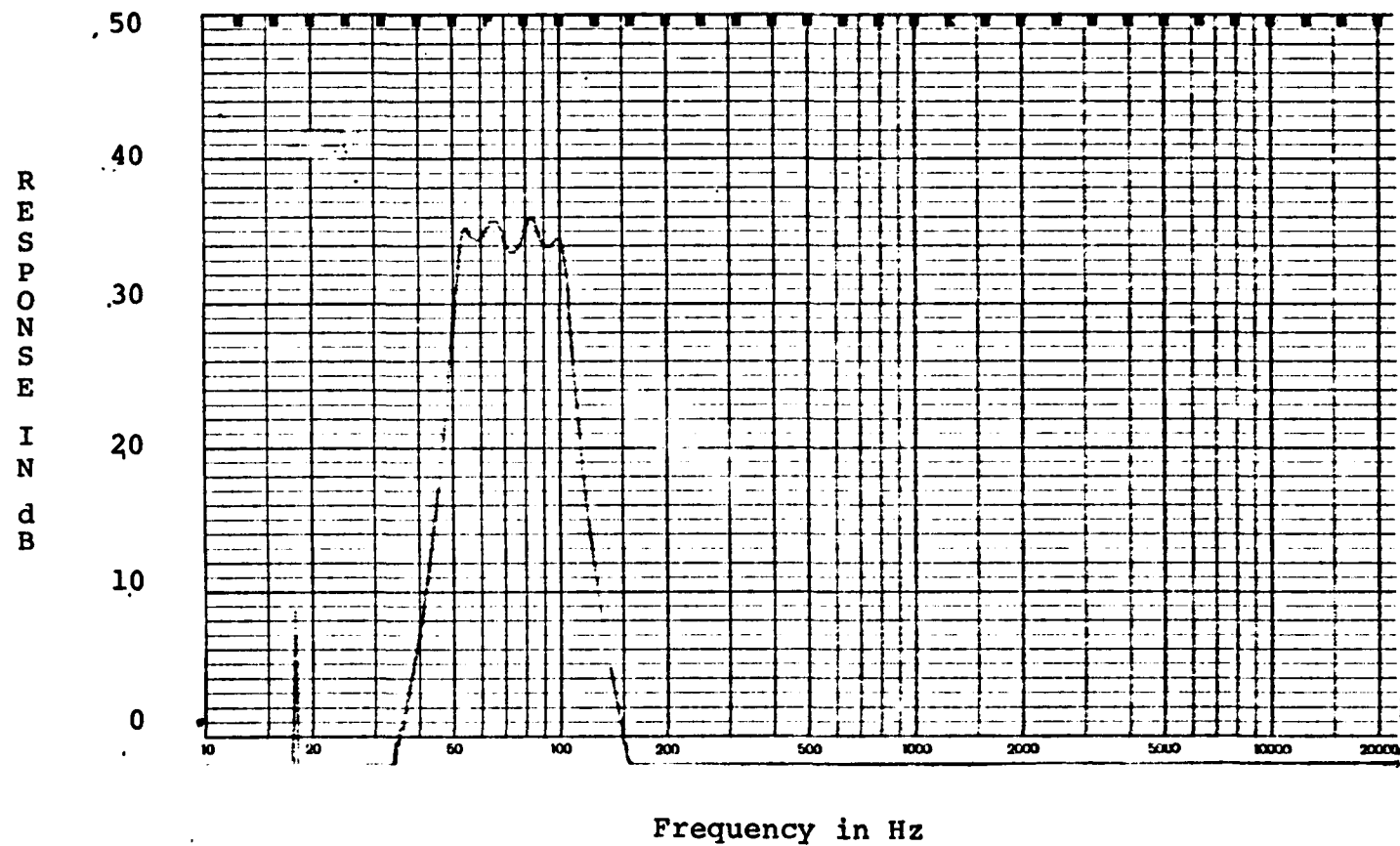


Figure 5.--Frequency response of the system used to produce the test tapes. The frequency response shown above was produced when the filter was set to produce an octave band centered at 80 Hz.

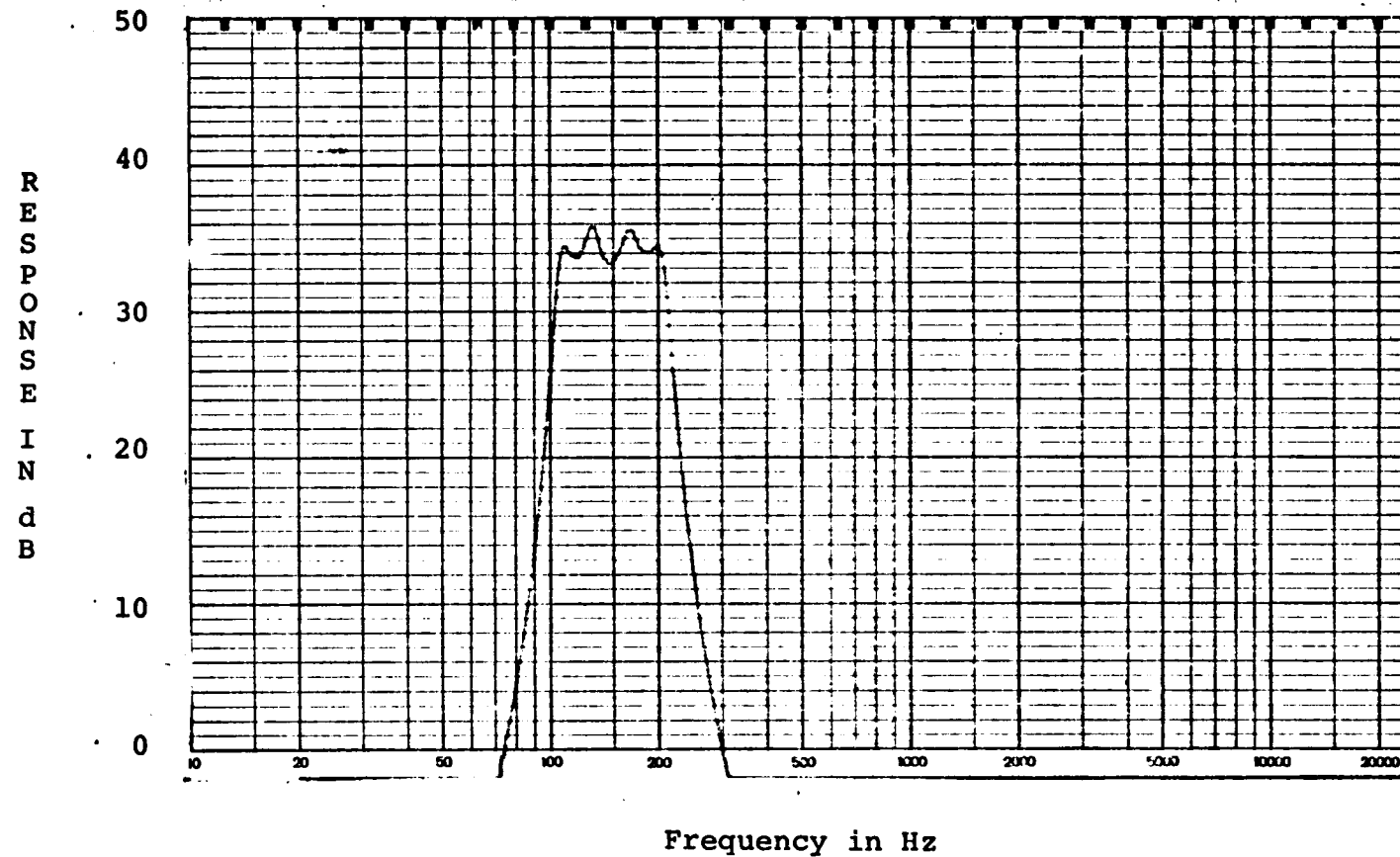


Figure 6.--Frequency response of the system used to produce the test tapes. The frequency response shown above was produced when the filter was set to produce an octave band centered at 160 Hz.

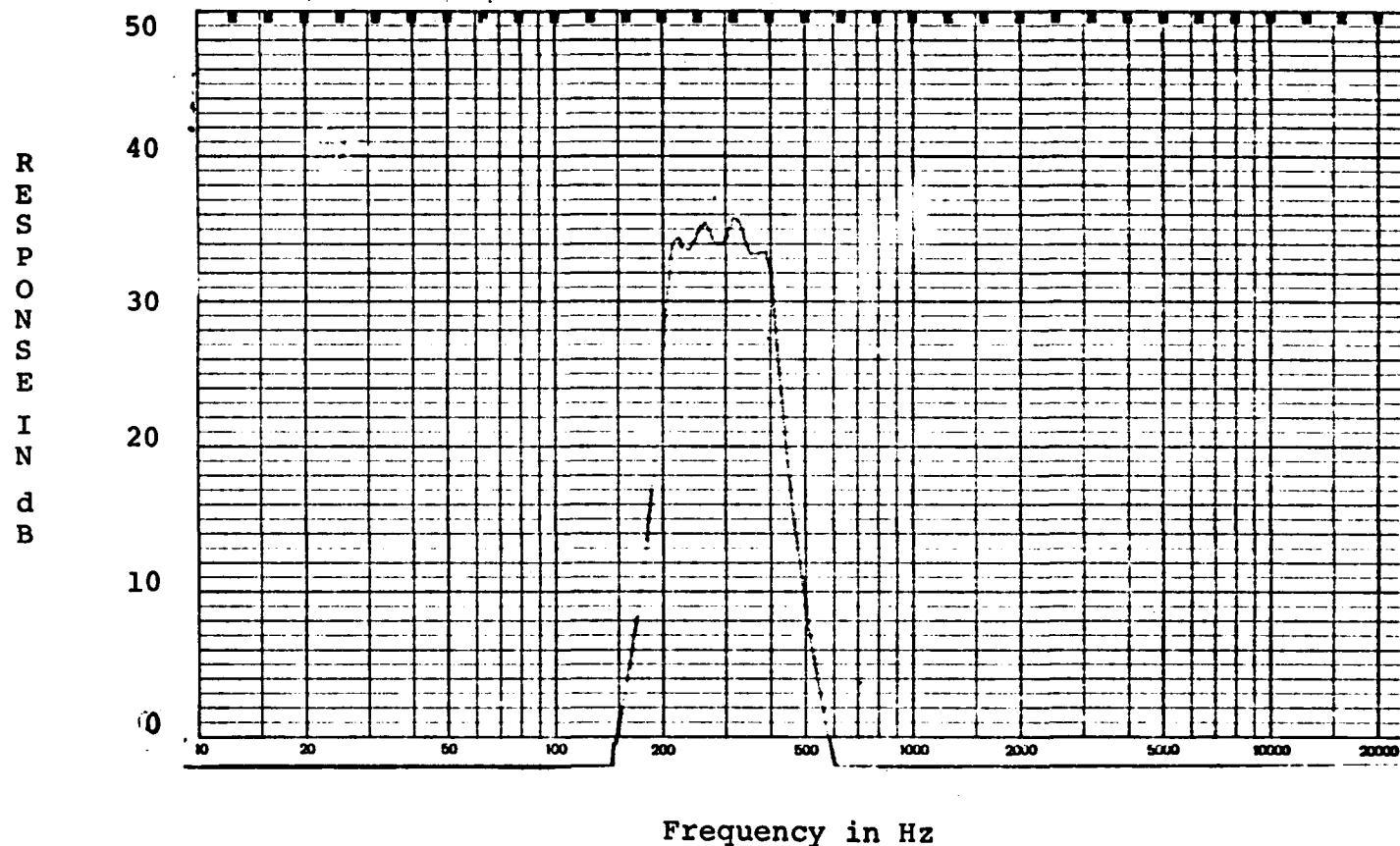
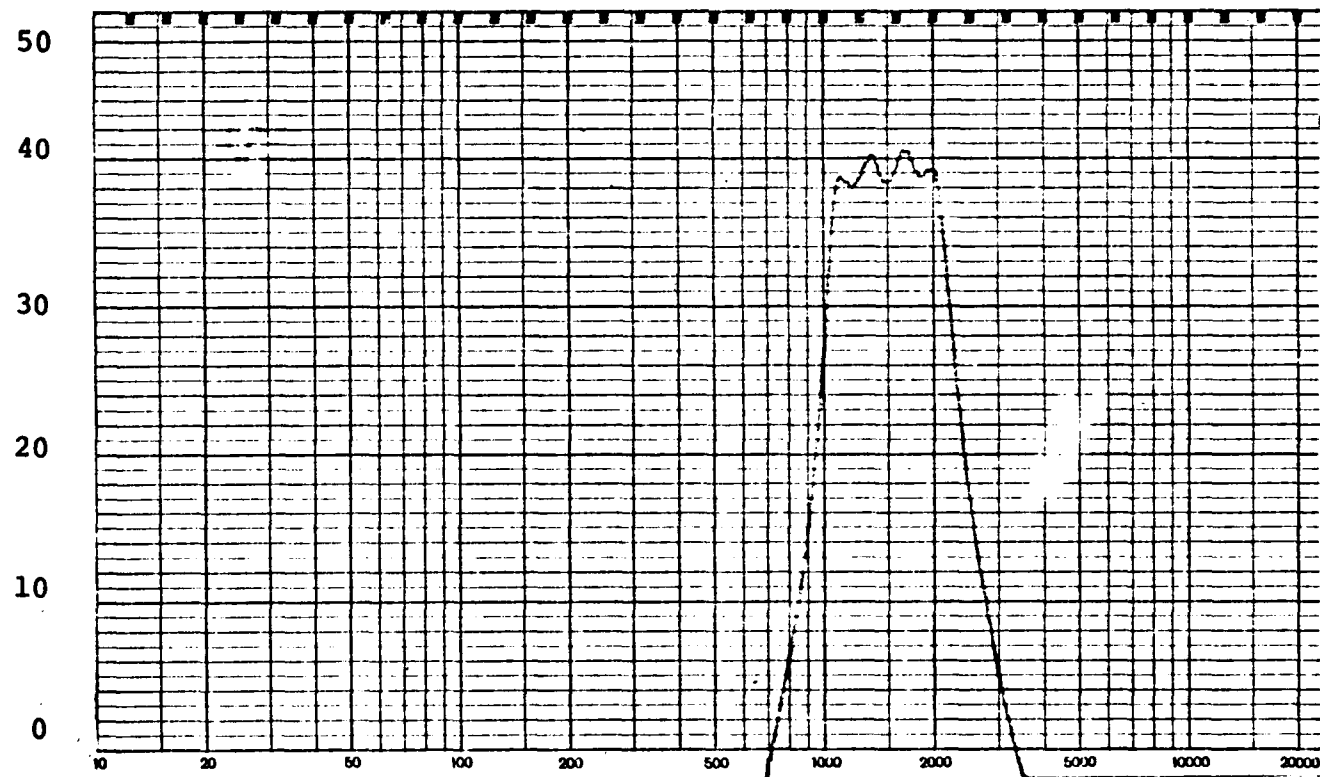


Figure 7.--Frequency response of the system used to produce the test tapes. The frequency response shown above was produced when the filter was set to produce an octave band centered at 315 Hz.

R  
E  
S  
P  
O  
N  
S  
E  
  
I  
N  
  
d  
B



Frequency in Hz

Figure 8.--Frequency response of the system used to produce the test tapes. The frequency response shown above was produced when the filter was set to produce an octave band centered at 1600 Hz.

Figure 9 shows the procedure used to measure the accuracy of the attenuator of the audiometer, these calibration measurements being replicated at 63, 125, 1000, and 2000 Hz. These frequencies were selected because they are the center frequencies of the octave band filters of the sound level meter which most closely correspond to the center frequencies of the passbands used in the experiment. The signal generator produced the sine waves of interest at .03 volt. The attenuator of the audiometer was adjusted to produce 70 dB SPL at the headphones when the VU meter of the audiometer peaked at 0 VU. All acoustic measurements were made by a Bruel and Kjaer Sound Level Meter, model 2203, in association with an artificial ear (Bruel and Kjaer, model 4152, with 6 cc cup). The same procedure was followed in 10 dB steps, from 70 dB SPL down through 40 dB SPL. This procedure revealed that the attenuator of the audiometer was linear within  $\pm 1$  dB for the frequencies of interest.

## 2. Determination of the Frequency Response of the Playback System

The frequency response characteristics of the playback system were measured at the frequencies of interest, as shown in Figure 10. The attenuator of the audiometer was set at a typical value and the output of the tape recorder adjusted to produce .03 volt. The calibration tape was played back on the recorder (Magne recorder, model

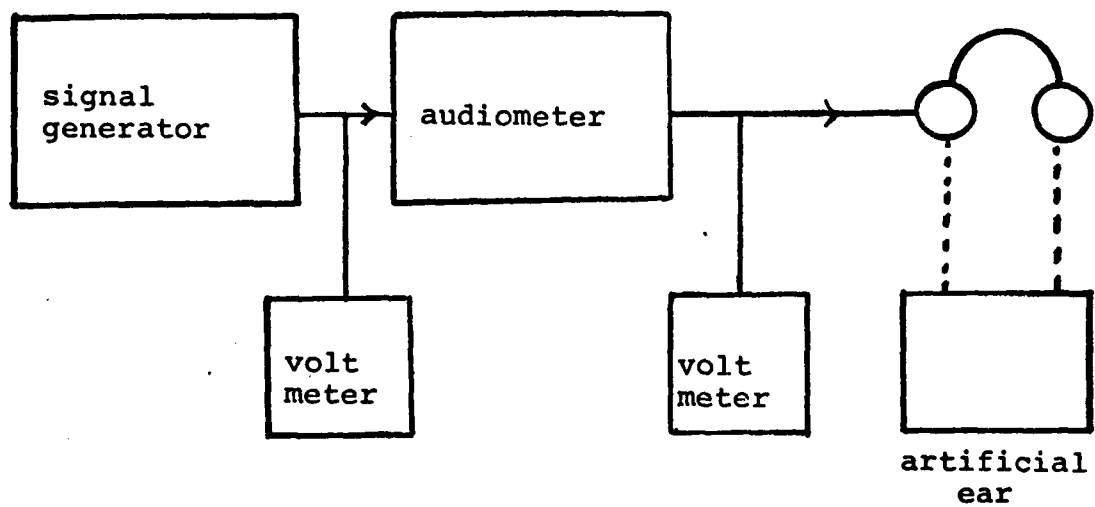


Figure 9.--Block diagram of apparatus used to calibrate the gain control of the playback system.

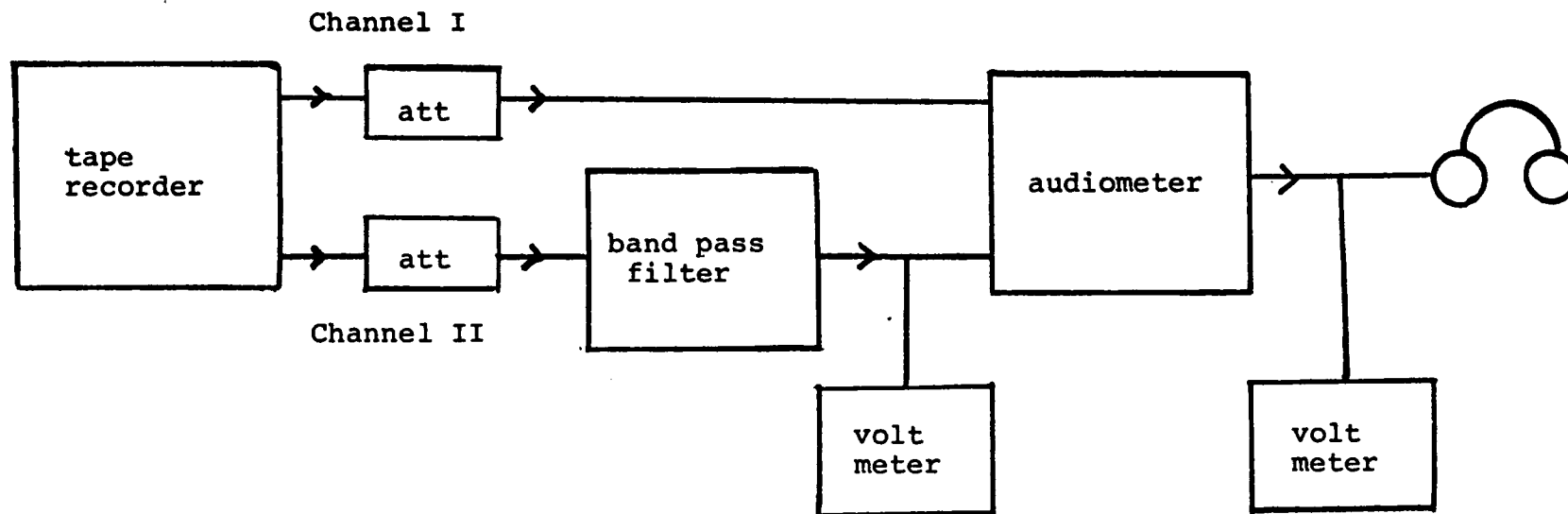


Figure 10.--Block diagram of the apparatus used to determine the frequency response characteristics of the playback system.

1028) and the voltage read across the headphones for 1600 Hz in channel I and for 80, 160, and 315 Hz in channel II. These levels were recorded so that they could be duplicated in the experimental sessions.

Prior to each experimental session, the procedure described above was followed, and the levels were set for the frequencies of interest. Likewise, at the conclusion of each session, the levels were again checked. In this way, any change in the output of the system during the experimental session could be detected.

The purpose of the calibration procedures described above was to make it possible to state acoustic levels in ways that could be reproduced by others. The fact that a speech stimulus is a transient signal, as opposed to a steady state signal, necessitated the involved calibration procedures. A simpler procedure would have been to use a VU meter for setting levels. However, the nature of a VU meter is such that it responds differentially to high and low frequencies, and hence, reliable readings for the separate bands are not possible.

#### IV. Subjects and Testing Procedure

Four college students with normal hearing served as subjects.<sup>3</sup> Each subject served a total of 8 hours, divided

---

<sup>3</sup>The data of a fifth subject had to be discarded because of change in the output of the equipment during the experimental session.

into several experimental sessions. The subjects were separated into two groups. Group I received the high frequency speech band in the right ear, and Group II, the high frequency speech band in the left ear. Subjects were provided with answer sheets like that shown in Table A in the appendix. All subjects received the following instructions:

The papers before you each contain 50 words. They are all real words. The first letter of each word has been omitted. You are going to hear sentences which have been distorted. Each sentence begins with the phrase "the next word is...". Following this phrase, you will hear the word that corresponds with the next word on your paper. After you hear the word, fill in the letter. You are not expected to be absolutely sure of all the words. If you are not sure, guess. Please leave no blanks; give a response to every item. Begin at the top of the first column and work down. When you finish the first column, go directly to the top of the second column. After completing the second column, continue in the same way to the third column. Are there any questions?

Three test trials containing only the high frequency speech band were run for two purposes, first to allow subjects to adjust to distorted speech, and second, to provide a reliable measure of the level for the high frequency speech band which would yield about a 30 per cent articulation score.<sup>4</sup> During these test trials, subjects

---

<sup>4</sup>The level that would yield about a 30 per cent articulation score was determined by dividing the word lists into groups of 10 words and manipulating the attenuator after each group to obtain the desired articulation score.

were further instructed to say the words aloud so that during the first test trial the experimenter could manipulate the attenuator to determine the level which would yield the desired articulation score. The other two test trials served as initial runs of the high frequency speech band only. Group I received the test trials in the right ear, and Group II received them in the left ear.

In all subsequent experimental conditions, the high frequency speech band was presented at the level which yielded about a 30 per cent articulation score, as determined in the test trials.

Under the experimental conditions, the high frequency speech band (1100-2200 Hz) was introduced into one ear and each of the three low frequency speech bands (50-100, 110-220, and 200-400 Hz) was introduced either to the same or to the opposite ear. Each low frequency speech band was presented monotically and dichotically at five intensity levels, i.e., at 0, 10, 20, 30, and 40 dB NSL. Losses in the output for the low frequency bands (see page 57) were compensated for by adding to the above mentioned level 32 dB for the 50-100 Hz band, 23 dB for the 110-220 Hz band, and 14 dB for the 200-400 Hz band. Table 1 summarizes the experimental design.

To avoid order effects, the experimental conditions were counterbalanced. Table B in the appendix provides a detailed description of the order of presentation of experimental conditions for each subject.

TABLE 1.--Experimental design. The blank cells of the matrix contained the articulation scores of the subjects.

| subjects | 1100-<br>2200<br>Hz | NSL              |   |                   |   |                   |   | NSL + 10 dB      |   |                   |   |                   |   | NSL + 20 dB      |   |                   |   |                   |   | NSL + 30 dB      |   |                   |   |                   |   | NSL + 40 dB      |   |                   |   |                   |   |
|----------|---------------------|------------------|---|-------------------|---|-------------------|---|------------------|---|-------------------|---|-------------------|---|------------------|---|-------------------|---|-------------------|---|------------------|---|-------------------|---|-------------------|---|------------------|---|-------------------|---|-------------------|---|
|          |                     | 50-<br>100<br>Hz |   | 110-<br>220<br>Hz |   | 200-<br>400<br>Hz |   | 50-<br>100<br>Hz |   | 110-<br>220<br>Hz |   | 200-<br>400<br>Hz |   | 50-<br>100<br>Hz |   | 110-<br>220<br>Hz |   | 200-<br>400<br>Hz |   | 50-<br>100<br>Hz |   | 110-<br>220<br>Hz |   | 200-<br>400<br>Hz |   | 50-<br>100<br>Hz |   | 110-<br>220<br>Hz |   | 200-<br>400<br>Hz |   |
|          |                     | R                | L | R                 | L | R                 | L | R                | L | R                 | L | R                 | L | R                | L | R                 | L | R                 | L | R                | L | R                 | L | R                 | L | R                | L | R                 | L | R                 | L |
| 1        | R                   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |
| 2        | R                   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |
| 3        | L                   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |
| 4        | L                   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |                  |   |                   |   |                   |   |

Key: R = right ear  
L = left ear

## V. Data Reduction

The number of correct responses for each condition was counted and the raw scores converted into proportions. To stabilize the error variance, these proportions were then subjected to the Arcsin Transformation (Kirk, 1968). An analysis of variance was performed on the transformed values of the experimental data. Experimental conditions, which the analysis of variance showed to be significantly different, were then compared with the test trials by use of the Duncan New Multiple Range Test (Edwards, 1960).

## CHAPTER IV

### RESULTS AND DISCUSSION

#### Results

This research was designed to investigate the effects of low frequency speech bands on intelligibility. As indicated in the Review of the Literature, the intelligibility of filtered speech is influenced by several factors. Hence, this investigation involved the manipulation of the low frequency speech bands as well as the manipulation of other variables; namely, the level of presentation of the low frequency speech bands, the mode of presentation (monotic or dichotic), and ear dominance. The results of these experimental manipulations are discussed in the following paragraphs.

Table 2 summarizes the analysis of variance<sup>1,2</sup> for main effects and first order interactions. The higher order interactions were not significant and therefore were not included in Table 2. Since Factor A, right-left ear

---

<sup>1</sup>The data were analyzed using a factorial analysis of variance. An alternative approach would have been a split-plot analysis (Kirk, 1968).

<sup>2</sup>In order to stabilize the variance, the data were transformed, using the Arcsin Transformation (Kirk, 1968). These transformed values were used in the analysis of variance.

TABLE 2.--Analysis of variance summary table of main effects and first order interactions. The higher order interactions were not significant and are not included in the table. The asterisk indicates a main effect that was significant at the .05 level of confidence.

| Source of Variance |                         | Sum of Squares | df | Mean Square | F      | F.05 |
|--------------------|-------------------------|----------------|----|-------------|--------|------|
| A                  | (ear effect)            | 0.037          | 1  | 0.037       | 1.42   | 5.32 |
| B                  | (mode of presentation)  | 0.028          | 1  | 0.028       | 1.08   | 5.32 |
| C                  | (bands)                 | 2.151          | 2  | 1.075       | 41.35* | 4.46 |
| D                  | (level of presentation) | 0.355          | 4  | 0.089       | 3.42   | 3.84 |
| PS                 | (paired subjects)       | 0.036          | 1  | 0.036       | 1.38   | 5.32 |
| AB                 |                         | 0.000          | 1  | 0.000       | -      |      |
| AC                 |                         | 0.037          | 2  | 0.018       | -      |      |
| AD                 |                         | 0.027          | 4  | 0.007       | -      |      |
| APS                |                         | 0.131          | 1  | 0.131       | 5.04   | 5.32 |
| BC                 |                         | 0.085          | 2  | 0.042       | 1.62   | 4.46 |
| BD                 |                         | 0.207          | 4  | 0.052       | 2.00   | 3.84 |
| BPS                |                         | 0.036          | 1  | 0.036       | 1.38   | 5.32 |
| CD                 |                         | 0.196          | 8  | 0.024       | -      |      |
| CPS                |                         | 0.116          | 2  | 0.058       | 2.23   | 4.46 |
| DPS                |                         | 0.060          | 4  | 0.015       | -      |      |
| ABCDPS             |                         | 0.206          | 8  | 0.026       |        |      |
| Total              |                         | 1.708          | 46 |             |        |      |

differences, was confounded by the difference between subjects (i.e., the two subjects of Group I received the high frequency band in the right ear and the two subjects of Group II received it in the left ear), both Factors A, PS, and the interaction term  $A \cdot PS$  may be inflated by individual subject differences. Thus, as shown in Table 3, a second analysis of variance was performed with the data collapsed<sup>3</sup> over Factor A.

Factor A, which represented right and left ear differences, was not significant at the .05 level of confidence. This indicated that the ear to which the high frequency speech band was presented did not significantly affect articulation scores.

The mode of presentation (monotic or dichotic), Factor B, was not significant at the .05 level of confidence. This indicated that mean articulation scores were not significantly altered when the high and low frequency speech bands were presented separately to the two ears (dichotic presentation) or together to one ear (monotic presentation); i.e., average differences between all monotic and dichotic conditions were not significant.

Figure 11 shows the mean articulation scores for the monotic and dichotic modes of presentation. The solid

---

<sup>3</sup>Since neither Factor A, nor any of the interactions involving A was significant, the technique of collapsing over Factor A appeared feasible. Further, this approach increased the number of degrees of freedom and thus provided a more powerful test of significance.

TABLE 3.--Analysis of variance summary table of main effects and first order interactions. For this analysis, the data were collapsed over Factor A (right-left ear differences). The asterisk indicates the main effects that were significant at the .05 level of confidence.

| Source of Variance        | Sum of Squares | df | Mean Square | F      | F.05 |
|---------------------------|----------------|----|-------------|--------|------|
| B (mode of presentation)  | 0.028          | 1  | 0.028       | 1.08   | 4.26 |
| C (bands)                 | 2.151          | 2  | 1.075       | 41.35* | 3.40 |
| D (level of presentation) | 0.355          | 4  | 0.089       | 3.42*  | 2.78 |
| S (subjects)              | 0.205          | 3  | 0.068       | 2.62   | 3.01 |
| BC                        | 0.085          | 2  | 0.042       | 1.62   | 3.40 |
| BD                        | 0.207          | 4  | 0.052       | 2.00   | 2.78 |
| BS                        | 0.071          | 3  | 0.024       | -      |      |
| CD                        | 0.196          | 8  | 0.024       | -      |      |
| CS                        | 0.237          | 6  | 0.039       | 1.50   | 2.51 |
| DS                        | 0.298          | 12 | 0.025       | -      |      |
| BCDS                      | 0.628          | 24 | 0.026       |        |      |
| Total                     | 4.461          | 69 |             |        |      |

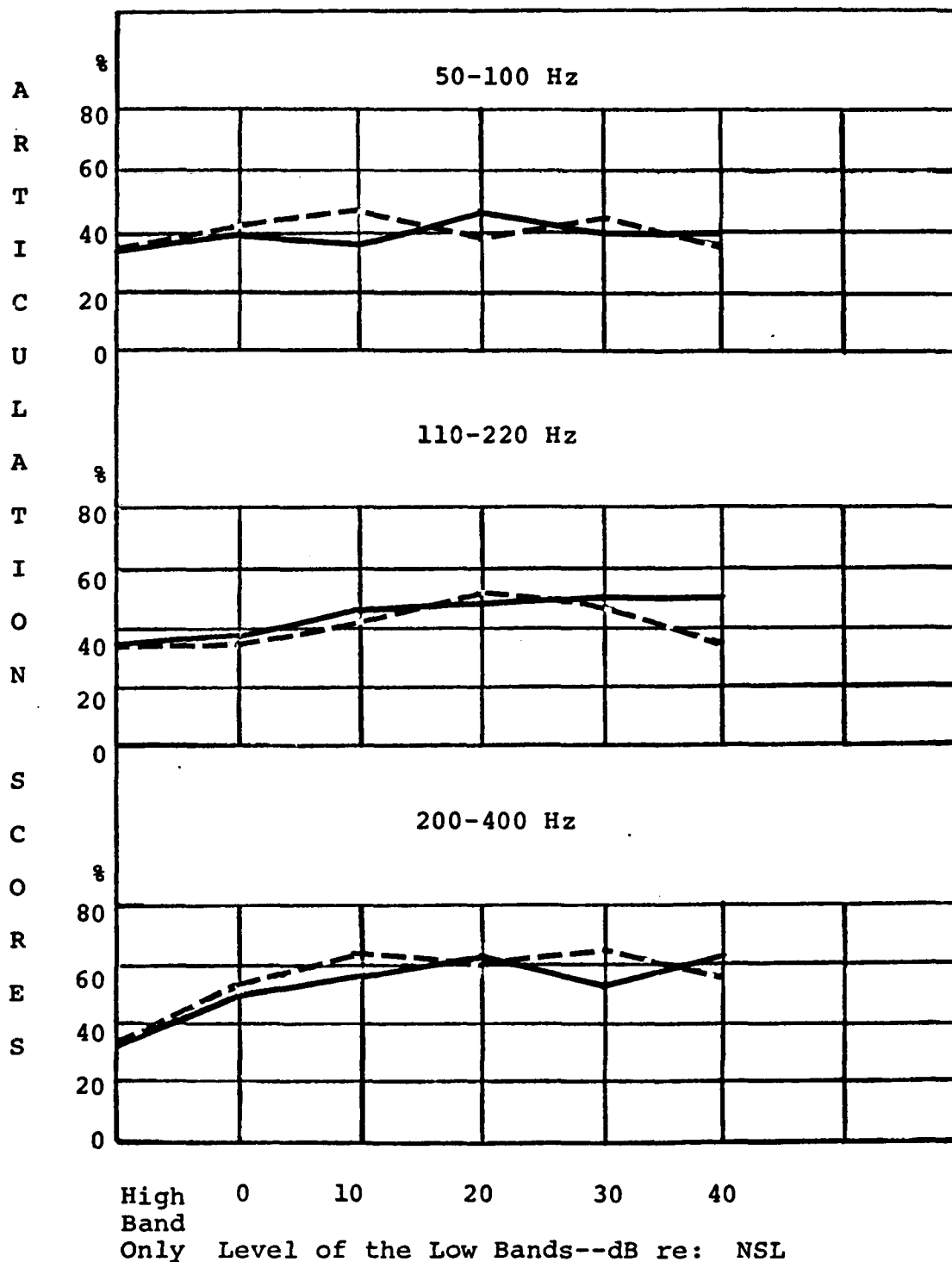


Figure 11.--Articulation scores for the monotic and dichotic listening conditions for each of the three low frequency speech bands. The solid line represents mean articulation scores for the dichotic condition and the dashed line represents mean scores for the monotic condition.

line represents mean articulation scores for the dichotic listening condition, and the dashed line represents mean scores for the monotonic condition. As can be seen in the figure, mean articulation scores for the monotonic condition decreased as the intensity level of the low frequency speech bands was increased from 30 to 40 dB NSL. In addition, it can be seen that articulation scores for the monotonic condition reached their first peak at a lower intensity level than scores for the dichotic listening condition. These trends, though not significant as determined by the analysis of variance,<sup>4</sup> were consistent for the three low frequency speech bands.

Factor D represented the effect of the level of presentation, averaged over the low frequency speech bands, on mean articulation scores. This factor was not significant at the predetermined level of confidence (.05) in the first analysis of variance, but in the second analysis of variance, which was a more powerful test, Factor D did reach significance at the .05 level of confidence. The second analysis was more powerful since there were more degrees of freedom in the residual error estimate; i.e., there were 24 degrees of freedom in the second analysis of variance as compared to 8 degrees of freedom in the first analysis of variance.

---

<sup>4</sup>Since the analysis of variance compares average over-all conditions, multiple range tests were done to compare specific conditions. These are discussed below.

Factor C, the effect of the different low frequency speech bands on mean articulation scores, was significant at the .05 level of confidence. This indicated that mean articulation scores were significantly altered by the presentation of the different low frequency speech bands.

To determine which of the low frequency speech bands resulted in significantly different mean articulation scores, the Duncan Multiple Range Test (Edwards, 1960) was performed. Results indicated that the mean articulation scores of the three low frequency speech bands were significantly different from each other at the .05 level of confidence; i.e., mean articulation scores averaged over Factors A, B, and D for the 200-400 Hz band (L3) were, on average, significantly higher than those for the 110-220 Hz band (L2), which were, on average, significantly higher than scores for the 50-100 Hz band (L1).

To determine whether mean articulation scores were significantly altered by the simultaneous presentation of the high and low frequency speech bands, as compared with the presentation of the high frequency speech band only, the Duncan Multiple Range Test was again performed. The test compared the condition in which none of the low bands were presented, i.e., when only the high frequency speech band was presented, with those conditions in which the high frequency band was presented in combination with a) the 50-100 Hz band (L1) at 0, 10, 20, 30, and 40 dB NSL,

respectively, b) the 110-220 Hz band (L2) at the above mentioned levels, and c) the 200-400 Hz band (L3) at the same levels. For L1 and L2, results indicated that, compared with the condition of the high band only, mean articulation scores were significantly higher at the .05 level of confidence when L1 and L2 were presented at 10, 20, and 30 dB NSL, respectively; however, the presentation of L1 and L2 at 0 and 40 dB NSL did not significantly alter articulation scores, as compared with the condition of the high band only. Results for L3 indicated that, compared with the condition of the high band only, the presentation of L3 in combination with the high band significantly increased mean articulation scores at all levels sampled.

Figure 12 shows the articulation scores for the three low frequency speech bands at 0, 10, 20, 30, and 40 dB NSL, respectively. For this figure, the Arcsin transformed values used in the analysis of variance were converted back to per cent correct. Since the ear effect (Factor A) and the mode of presentation (Factor B) were not significant factors, the data were collapsed over these conditions. The 95 per cent confidence intervals were determined in accordance with Edwards (1967). As may be seen in the figure, presentation of the low bands at 10, 20, and 30 dB NSL resulted in higher mean articulation

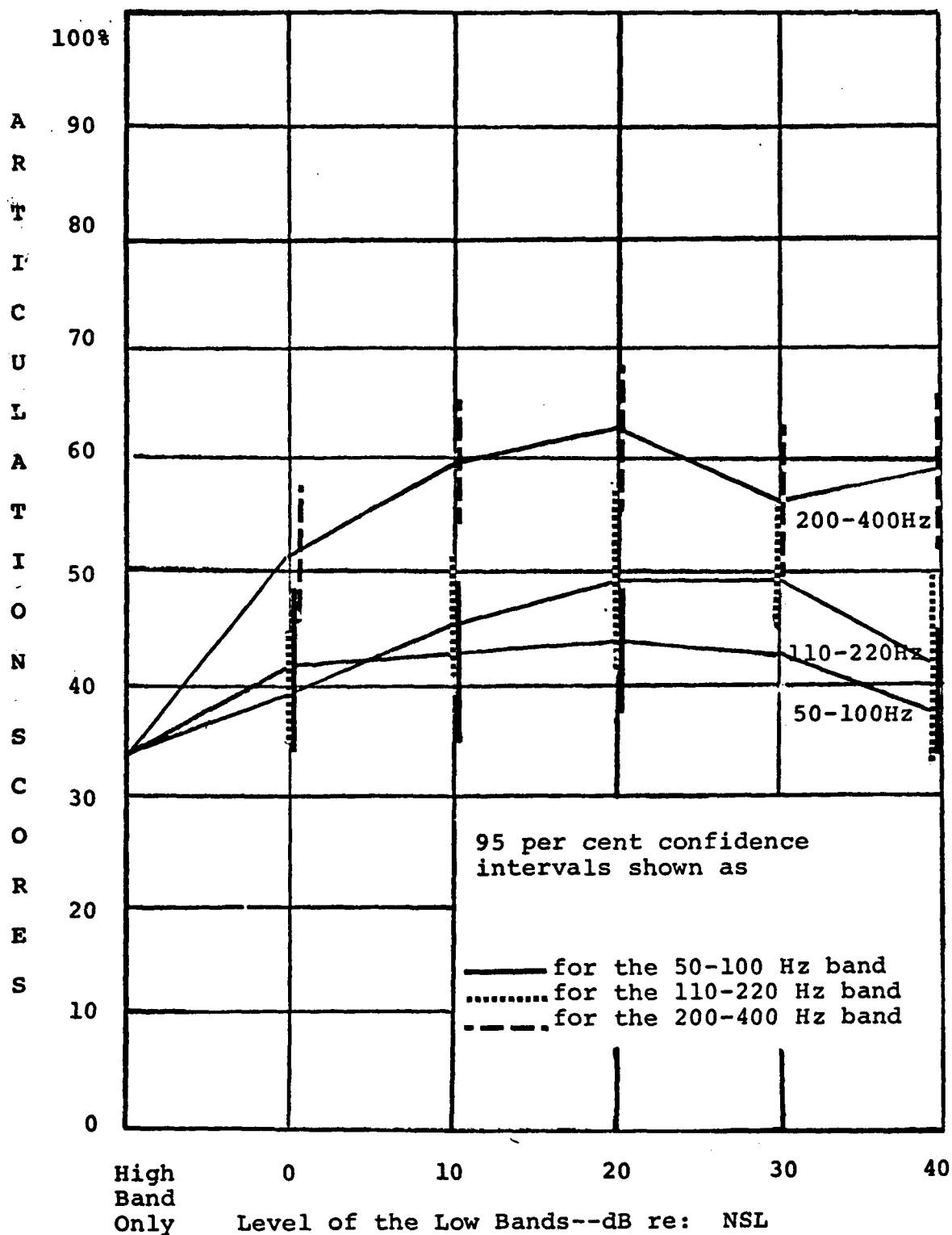


Figure 12.--Articulation Gain Functions for the 50-100, 110-220, and 200-400 Hz bands.

scores than presentation of the bands at 0 dB NSL.<sup>5</sup> The significance of this difference was confirmed at the .05 level of confidence by the Duncan Multiple Range Test (Edwards, 1960).

### Discussion

A significant result of the present research was that in combination with the high frequency speech band, the presentation of any of the low frequency speech bands substantially increased mean articulation scores. In addition, the data showed that as the center frequency of the low frequency speech bands increased, respectively, from 50-100, to 110-220, to 200-400 Hz, intelligibility likewise increased. In part, this progressive improvement in intelligibility may have been due to differences in concentration of cues for speech perception in the different frequency regions; i.e., the 200-400 Hz band contains more cues for intelligibility than the 110-220 Hz band, which, in turn, contains more cues than the 50-100 Hz band. Similarly, the improvement in intelligibility found with the increase in frequency may have been

---

<sup>5</sup>Mean articulation scores for the low bands presented at 40 dB NSL were not significantly different from those obtained at 0 dB NSL or at 10, 20, and 30 dB NSL.

due to the increase in amount of energy available. Since octave bands were used in the present research, and since bandwidth was therefore proportional to the center frequency, as the center frequency increased, progressively more energy was available. Likewise, this additional energy may have contributed to the improvement in intelligibility.

The increase in articulation scores obtained by the presentation of any one of the low frequency speech bands is especially revealing, considering that the Fairbanks Rhyme Test was the stimulus material employed in the present research. This test samples initial consonant and consonant-vowel transitions, and cues for consonant discrimination are generally believed to be located in the higher frequency regions, especially above 500 Hz. Yet, the data showed a significant increase in articulation scores as the result of presenting speech bands below 500 Hz.

To determine whether the improvement in articulation scores obtained as the result of adding frequencies below 500 Hz could be related to the articulatory features of place of articulation, manner of articulation, and voicing, responses for the high-band-only condition (HB) were compared with those obtained when the low band of 200-400 Hz (LB) was added at 10 or 20 dB NSL (HB + LB). The latter conditions were chosen for analysis because they provided the largest gain in articulation scores.

If any improvement in perception of these features occurred, it would be most evident in this comparison. It was further desired that within-subject analysis of responses for these two conditions be based on the same word list.<sup>6</sup> Therefore, it was necessary to compare responses for the HB condition with those obtained when the LB was added at 10 dB NSL for one subject and with those obtained for the LB at 20 dB NSL for the remaining three subjects. Further, for this analysis, ear effect and the mode of presentation were not considered, since neither of these variables was significant, as determined by the analysis of variance.

Tables 4 and 5 summarize the results of this analysis. Table 4 shows the per cent of error for the articulatory features under the HB and HB + LB conditions. Since this analysis was based on the results of four subjects, each responding to a list of 50 words, the total number of possible errors for each of the articulatory features of place of articulation, manner of articulation, and voicing was 200.

As indicated in Table 4, the per cent of error for each articulatory feature decreased substantially when the LB was introduced.

---

<sup>6</sup>Of necessity, between-subject analysis of responses was based on different lists. See Table B of the Appendix.

TABLE 4.--Errors according to articulatory feature

| Articulatory<br>Feature   | Number of<br>Possible<br>Errors | Number of<br>Errors |         | Per Cent<br>of Errors |         |
|---------------------------|---------------------------------|---------------------|---------|-----------------------|---------|
|                           |                                 | HB                  | HB + LB | HB                    | HB + LB |
| Place of<br>Articulation  | 200                             | 105                 | 69      | 52                    | 34      |
| Manner of<br>Articulation | 200                             | 95                  | 49      | 48                    | 24      |
| Voicing                   | 200                             | 56                  | 20      | 28                    | 10      |

TABLE 5.--Errors according to sound type

| Sound Type                      | Total Number in Lists | Number of Possible Errors | Number of Errors |         | Per Cent of Error |         |
|---------------------------------|-----------------------|---------------------------|------------------|---------|-------------------|---------|
|                                 |                       |                           | HB               | HB + LB | HB                | HB + LB |
| Voiceless Plosives<br>/p,t,k/   | 44                    | 132                       | 63               | 24      | 48                | 18      |
| Voiced Plosives<br>/b,d,g/      | 35                    | 105                       | 38               | 22      | 36                | 21      |
| Nasals<br>/m,n,/                | 25                    | 75                        | 41               | 3       | 55                | 4       |
| Voiceless Fricatives<br>/f,s,h/ | 47                    | 141                       | 69               | 55      | 49                | 39      |
| Lateral<br>/l/                  | 15                    | 45                        | 16               | 7       | 36                | 16      |
| Glides<br>/w,r,y/               | 28                    | 84                        | 17               | 18      | 20                | 21      |

The reduction in errors for place of articulation resulting from adding the LB can only be partially explained for certain sounds. Most of the improvement in place of articulation resulted from the improved perception of the labials /p,b,m,f,v/. Specifically, of the 35 fewer errors noted when the LB was presented, over half (i.e., 57 per cent) resulted from improvement in the perception of these sounds. Cues for the perception of place of articulation have been located already in the second formant transition, and this transition has been located at about 720 Hz for the labials (Delattre, et al., 1955). Considering that the roll-off of the filters used in the present experiment was 36 dB per octave, and that this analysis was based on the presentation of the LB at 10 or 20 dB NSL, it is possible that the second formant transition for these sounds was perceived.

The reduction in the number of errors for the manner of articulation when the LB was presented is consistent with the results of previous research. The first formant transition located below 500 Hz has been identified previously as an important cue for the perception of manner of articulation. Specifically, Liberman et al. (1952), Cooper et al. (1952), Liberman et al. (1955), Delattre et al. (1955), and Malecot (1956) found the first formant transition to be an important cue for plosives and nasals and located this transition between 0-240 Hz. O'Connor et al. (1957) and Lisker (1957) identified the first

formant transition for the glides /w,r,y/ and the lateral /l/ to be between 240-460 Hz. Thus, one would expect a substantial reduction in errors of manner of articulation when the LB of 200-400 Hz is presented.

The substantial decrease in the number of voicing errors occurring when the LB was presented is likewise consistent with the findings of previous research. It has been shown that cues for perception of voicing for continuants; i.e., all except plosives, are clearly conveyed by the fundamental frequency and the low order harmonics which contain energy in the region 200-400 Hz. In the case of plosives, voicing cues have been located in the first formant transition (Lieberman et al., 1958) and in the fundamental frequency of the following vowel (Haggard et al., 1970). As stated above, the roll-off of the filters used in the present experiment was 36 dB per octave. Further, the present analysis was based on the presentation of the LB at 10 or 20 dB NSL. Thus, the LB provided direct cues for voicing for the continuants and less obvious cues for voicing for the plosives.

Table 5 shows the per cent of error for each sound type. For this table, /v,z,dʒ/ were excluded from consideration due to insufficient sampling. Multiple errors often occurred on a given test item. For example, the substitution of /t/ for /m/ involves 3 errors, 1 for place of articulation, 1 for manner of articulation, and 1 for voicing. Therefore, the number of possible errors for a

given sound type is three times the total number of occurrences of that sound type. The table indicates that in comparison with the HB condition presentation of the LB decreased the per cent of error for all sound types analyzed, except the glides /w,r,y/. The presentation of the LB had the greatest effect on the nasals, the per cent error for nasals falling from 55 to 4 per cent. There was a substantial improvement in the voiceless plosives, voiced plosives, and the lateral. For the voiceless fricatives, a moderate improvement occurred, and no improvement was seen for the glides.

The greatest improvement in intelligibility resulting from the introduction of the LB occurred for nasals. This improvement would be expected because of the nasal resonance in the low frequencies (Malecot, 1956). Consistent with this expectation, the introduction of the LB reduced errors in manner of articulation from 17 for the HB condition to 1 for the LB condition. Since nasals are normally voiced, reduction in errors for manner of articulation should also reduce voicing errors. Voicing errors fell from 10 to 0 with the addition of the LB.

It is worth noting that of the 63 errors occurring on the voiceless plosives in the HB condition, 34 were errors in place of articulation. However, of the 24 errors occurring on these sounds when the LB was presented, 23 were errors in place of articulation. Thus, not only was place of articulation the most likely type of error with

voiceless plosives, but with the addition of the LB, all other error types were virtually eliminated.

In comparison with the voiceless plosives, the kinds of errors obtained for the voiced plosives were much more varied. However, the error types showed trends similar to those noted for the voiceless plosives. Of the 38 errors occurring on the voiced plosives in the HB condition, 14 were errors in place of articulation and 10 were errors in manner of articulation. When the LB was introduced, the total number of errors was reduced to 22, 10 being errors in place of articulation and 3 being errors in manner of articulation. Thus, for the voiced plosives, as in the case of the voiceless plosives, errors in place of articulation were most common, and presentation of the LB substantially reduced errors in manner of articulation.

Reduction of voicing errors was the most obvious difference between the voiceless and voiced plosives when the LB was introduced. For voiceless plosives, presentation of the LB totally eliminated voicing errors (from 11 to 0); for voiced plosives, voicing errors decreased moderately, from 14 for the HB condition to 9 when the LB was introduced. The number of voicing errors for the voiced plosives in the HB + LB condition is surprising because acoustic cues for the distinction between voiceless and voiced plosives have been located in the first formant transition and in the fundamental frequency of the following vowel. It was these frequency regions

(i.e., the frequency regions of the first formant and of the fundamental frequency) that were made available by the addition of the LB. It is possible, however, that subjective impressions of the filtering caused subjects in this experiment to perceive voiceless, as opposed to voiced plosives, in the HB + LB condition.

Since voiceless fricatives have most of their energy in the high frequency regions (Hughes and Halle, 1957; Harris, 1958), decrease in fricative errors was not expected when the LB was introduced. However, most of the improvement found can be attributed to the reduction in the number of voicing errors. For the HB condition, voicing errors totaled 12; when the LB was introduced, voicing errors decreased to 3. Thus, the data show that errors were made not only on the basis of what was heard, but also on the basis of what might have been heard, given that the LB was presented. That is, in the HB condition, subjects presumably guessed at voicing, while in the HB + LB condition, there was no need to guess since voicing cues were present.

The substantial reduction in voicing errors for the voiceless fricatives obtained when the LB was introduced is consistent with that noted above for the voiceless plosives. It is significant that for three of the four subjects all voicing errors for all the voiceless consonants were completely eliminated when the LB was presented.

For the fourth subject, a negligible number (3) of voicing errors remained.

Consistent with the high frequency content of voiceless fricatives, the number of errors in place of articulation and in manner of articulation did not change substantially when the LB was introduced. For the HB condition, errors in place of articulation and manner of articulation totaled 26 and 31, respectively. When the LB was introduced, these numbers were reduced only slightly, to 24 and 28, respectively.

It will be noted that a disproportionate number of errors in manner of articulation for the voiceless fricatives involved the substitution of plosives for voiceless fricatives. For the HB condition, 21 of the 31 errors in manner of articulation were plosive substitutions (e.g., /p/ for /f/, as in pat for fat); when the LB was introduced, 24 of the 28 errors involved plosive substitutions. Menon (1969) noted that the voiceless plosives have primary and secondary bursts in the high frequency regions. This may, in part, account for the substitution of /p/, /t/, or /k/ for the voiceless fricatives in the data of the present experiment.

In the case of the lateral, it will also be noted that for three of the four subjects the introduction of the LB eliminated all errors. In the case of the fourth subject, this trend was reversed; i.e., the presentation of the LB produced perceptual errors which were not present

in the HB condition. The fourth subject was frequently inconsistent with the other subjects.

Of the 11 errors in place of articulation which occurred in both the HB and HB + LB conditions, 10 were bilabial in nature. This is consistent with results reported by Franklin (1969, 1970), who likewise found a high incidence of bilabial substitutions. However, inconsistent with Franklin, who reported that most of these errors involved a /w/ substitution for /l/, the present writer found that all bilabials were substituted with equal frequency for /l/.

The presentation of the LB was found to have little effect on the perception of the glides. This could be expected because of the relatively few errors (17) occurring in the HB condition, a situation which was likewise noted by Franklin. Franklin further reported that maximum improvement in the /w/ and /r/ (8 and 22 percentage points, respectively) occurred when the LB was presented at threshold. The present analysis was based on the presentation of the LB at 10 or 20 dB NSL.

With the minor exception noted for the lateral, it is significant that the results reported above were consistent with those obtained by Franklin (1969, 1970). Specifically, in both the Franklin study and in the present research, the following were found:

a) The nasal consonants benefited greatly by the presen-

tation of the LB. b) The voiceless plosives most frequently involved errors in place of articulation. c) Voiced plosive errors varied more than those for their voiceless counterparts. d) Voiceless plosives were frequently substituted for the voiceless fricatives. e) Bilabials were frequently substituted for the lateral.

Another finding of the present research was that the presentation of the bands 50-100 and 110-220 Hz at an intensity level of 10, 20, or 30 dB NSL contributed to intelligibility, but that the presentation of these bands at 0 or 40 dB NSL resulted in no significant improvement in intelligibility.

When the low frequency speech bands were presented at 0 dB NSL, in combination with the high frequency band, the high frequency band may have masked the low frequency bands. This is consistent with Bilger and Hirsh (1956) who noted that at low intensities high frequencies are more effective maskers than low frequencies. However, at high intensity levels, the low frequencies are more effective maskers. Therefore, presentation of the low frequency bands at 40 dB NSL in the same ear as the high frequency band may have masked the high band. When the low band was presented to one ear (at 40 dB NSL) and the high band to the opposite ear, the spread of masking was less severe. This can be seen in Figure 11. Franklin (1969, 1970) found a similar occurrence when she presented a low frequency speech band at 40 dB SL to the same or

opposite ear as a high frequency speech band.

Since there was a significant difference between the monotic and dichotic modes of presentation only at 40 dB NSL, the mode of presentation did not play a major role in the present research. This is not entirely inconsistent with findings of previous research. Linden (1964) found that articulation scores were not significantly different when the speech bands 560-715 and 1800-2200 Hz were presented monotically or dichotically. Palva (1965) confirmed these findings for the speech bands 480-720 and 1800-2400 Hz. Franklin (1969, 1970) likewise found similar articulation scores when the speech band 240-480 Hz was presented at 0 or 20 dB SL in the same or opposite ear as the speech band 1020-2040 Hz. It was only when the low frequency speech band was presented at the high level of 40 dB SL that articulation scores were higher for the dichotic as opposed to the monotic modes of presentation. Whether the articulation scores were significantly higher for the dichotic listening condition cannot be determined from Franklin's data, as no statistics are provided. Huizing and Taselaar (1961) found that when the speech bands 140-280 and 1128-2256 Hz were presented in the monotic condition the high frequency speech band had to be presented at an intensity level of 26 dB to attain a 50 per cent articulation score; whereas, in the dichotic condition 12 dB was sufficient. The apparent superiority of the dichotic listening condition found in this experiment

probably resulted from the high level at which the low frequency speech band was presented, i.e., at an intensity level of 75 dB; in the monotic condition, the low frequency speech band (at 75 dB) would have masked the high frequency band.

The difference between the right and left ears as receptors for speech was another variable found not to be significant in the present research. This finding does not contradict the numerous studies available which support a right ear advantage for speech reception. Most of these studies involve binaural separation tasks, which, as defined by Ohta et al. (1960) is "the phenomenon in which two different signals delivered to both ears simultaneously are discriminated and separated from each other." In binaural separation experiments, a right ear advantage has been demonstrated fairly consistently. However, this task must be distinguished from binaural fusion, which Ohta et al. (1960) define as "the phenomenon whereby a single sound image is obtained when two signals are delivered to both ears, even if there is a difference of time, phase, or quality between them." Ohta et al. believe these two functions represent different binaural mechanisms located at different levels in the central auditory pathway.

In fusion experiments, the right ear advantage has not typically been found. To the writer's knowledge, the only research which involved a fusion task and which demonstrated a right ear advantage was done by Franklin (1969,

1970). The present research did not confirm her findings.

Although generalizations from studies employing normal subjects to studies dealing with hearing aids and with hearing impaired populations must be done cautiously, the results of the present research do tend to support the research of Ling (1965), who experimented with hearing aids which amplify frequencies as low as 80 Hz. Ling's preliminary observations indicated a significant improvement in the perception of rhythm, stress, and intonation patterns by hearing impaired subjects using his experimental aid, as opposed to a standard aid. Standard aids tend to favor high frequencies and attenuate low frequencies, providing usable amplification in the range 300-3000 Hz. The attenuation of frequencies below 300 Hz is consistent with results reported by Davis et al. (1947). In their study, Davis et al. amplified the low frequencies by as much as 40 dB, which could have resulted in masking of the high frequencies. The results of the present research, as well as the results reported by Franklin and Ling, tend to support the need for further research on low frequency amplification and to encourage further study of hearing impaired populations with low frequency residual hearing.

## CHAPTER V

### SUMMARY

#### Introduction and Statement of the Problem

The present research was designed to investigate the effects on intelligibility of three low frequency speech bands; namely, 50-100, 110-220, and 200-400 Hz. Previous research has already identified intensity, ear dominance, and mode of presentation (monotic or dichotic) as variables affecting the intelligibility of filtered speech. Hence, in this research, an attempt was made to measure the effects on intelligibility of low frequency speech bands with respect to a) the level at which the low frequency speech bands were presented, b) the mode of presentation (monotic or dichotic) of the high and low frequency speech bands, and c) the ear to which the high frequency speech band was presented.

#### Materials, Subjects, and Procedures

The Fairbanks Rhyme Test (1958) served as the stimulus material in the present research. The stimulus materials were filtered into four speech bands, a high band of 1100-2200 Hz and three low bands of 50-100, 110-220, and 200-400 Hz. These speech bands were presented to four sub-

jects with normal hearing. The subjects were divided into two groups, Group I receiving the high frequency speech band in the right ear and Group II receiving this band in the left ear. All subjects were given three test trials in which only the high frequency speech band was presented. During the first test trial, the experimenter manipulated the attenuator to determine the level that would yield a 30 per cent articulation score. The other two test trials served both as initial runs of the high frequency speech band only and as a base to which all subsequent experimental conditions were compared. The right ear was used for the test trials for Group I and the left ear was used for the test trials for Group II.

All subsequent experimental conditions consisted of the presentation of the high band at the level determined by the test trials and the simultaneous presentation of one of the three low frequency speech bands. The low bands were presented at 0, 10, 20, 30, and 40 dB NSL a) in the same ear as the high band (monotic condition), and b) in the opposite ear as the high band (dichotic condition).

### Results

Data were analyzed using the analysis of variance. Results showed that the mode of presentation (monotic or dichotic) and the ear to which the high frequency speech band was presented were not significant at the .05 level of confidence. The effect of the different low frequency

speech bands on mean articulation scores was found to be significant at the .05 level of confidence.

A series of Duncan Multiple Range Tests (Edwards, 1960) was performed to determine whether the addition of any one of the low frequency speech bands resulted in significantly different mean articulation scores and also to determine which of the three low frequency speech bands resulted in significantly different mean articulation scores. When compared with scores for the high band only, the data revealed significantly higher mean articulation scores when the 50-100 or 110-220 Hz band was presented at 10, 20, or 30 dB NSL or when the 200-400 Hz band was presented at any of the levels sampled. In addition, it was found that mean articulation scores were significantly higher for the 200-400 Hz band, as compared with scores for the 110-220 Hz band and that articulation scores for the 110-220 Hz band were higher than those for the 50-100 Hz band.

The level of presentation of the low frequency speech bands was not significant at the predetermined level of confidence (.05) but did reach significance at the .10 level of confidence. The Duncan Multiple Range Test was again performed, and this indicated that mean articulation scores at 10, 20, and 30 dB NSL were significantly higher than at 0 dB NSL.

#### Discussion

One of the most significant findings of the present

research was that articulation scores for the high frequency speech band increased as the result of adding any one of the low frequency speech bands. The findings of this study also showed that as the speech bands increased in frequency from 50-100, to 110-220, to 200-400 Hz articulation scores also increased. This improvement in intelligibility may have been caused, in part, by the number of cues for speech perception available in the different frequency regions. Similarly, since octave bands were used, and since bandwidths were therefore proportional to the center frequencies, the increase in articulation scores associated with the increase in frequency may be attributable to the greater amount of energy available in the higher speech bands. These two factors were obviously inter-related in the present research.

To determine whether the improvement in intelligibility resulting from the introduction of the 200-400 Hz band at 10 or 20 dB NSL could be explained in terms of the articulatory features of place of articulation, manner of articulation, and voicing, errors in the articulatory features for the high band only condition (HB) were compared with those obtained when the high and low bands (HB + LB) were presented. The data showed that the introduction of the LB reduced the per cent of errors for all the articulatory features. Specifically, in comparison

with the HB condition, presentation of the LB reduced voicing errors from 28 to 10 per cent, errors in manner of articulation from 48 to 24 per cent, and errors in place of articulation from 52 to 35 per cent. Thus, introduction of the LB decreased voicing errors most, substantially reduced errors in manner of articulation, and moderately decreased errors in place of articulation.

Analysis of the perception of the various sound types sampled showed that nasal consonants benefited most from the introduction of the LB, that there was likewise a substantial improvement in the perception of the voiceless and voiced plosives, a moderate improvement in the perception of voiceless fricatives and the lateral, and that there was no improvement in the perception of the glides.

Historically, speech frequencies below about 300 Hz have been considered unimportant for intelligibility. The present research demonstrated that speech bands as low as 50-100 Hz result in a small but statistically significant improvement in intelligibility. This finding is especially significant when one considers that the stimulus material used in the present research (the Fairbanks Rhyme Test) samples the initial consonant and consonant-vowel transition, and when one considers that it is generally reported that cues for consonant discrimination are located in frequencies above 500 Hz.

## APPENDIX

TABLE A.--Sample of answer sheet on which subjects  
recorded their responses

---



---

|      |      |       |
|------|------|-------|
| -ot  | -ip  | -old  |
| -ay  | -ore | -it   |
| -op  | -ang | -ed   |
| -eel | -en  | -end  |
| -ake | -ark | -id   |
| -aw  | -oil | -ack  |
| -ile | -ig  | -ail  |
| -eat | -age | -ight |
| -ook | -ast | -orn  |
| -ill | -ain | -od   |
| -ire | -est | -ock  |
| -ale | -un  | -ump  |
| -ent | -eal | -ate  |
| -oon | -in  | -ell  |
| -ick | -ust | -et   |
| -ame | -ine | -uck  |
| -ide | -ink |       |

---

TABLE B.--Summary of the order of presentation of experimental conditions for each subject

---

Key: A = Bands A1 = 1100-2200 Hz A2 = 50-100 Hz  
 A3 = 110-220 Hz A4 = 200-400 Hz  
 B = Mode of Presentation B1 = High Band (HB) - Left Ear (Lt E)  
 B2 = HB-Right Ear (Rt E)  
 B3 = HB-Lt E and LB-Rt E  
 B4 = HB-Rt E and LB-Lt E  
 B5 = HB-Lt E and LB-Lt E  
 B6 = HB-Rt E and LB-Rt E  
 C = Levels C1 = 30 per cent level C2 = Normal Speech Level (NSL)  
 C3 = NSL + 10 dB C4 = NSL + 20 dB  
 C5 = NSL + 30 dB C6 = NSL + 40 dB  
 \*the numbers within the boxes refer to the five  
 lists of the Fairbanks Rhyme Test

| Subjects |      | C1 |    | C2 |    | C3 |    | C4 |    | C5 |    | C6 |    |
|----------|------|----|----|----|----|----|----|----|----|----|----|----|----|
|          |      | B1 | B2 | B3 | B5 | B3 | B5 | B3 | B5 | B3 | B5 | B3 | B5 |
| I        | A1   | *5 |    |    |    |    |    |    |    |    |    |    |    |
|          | A1   | 1  |    |    |    |    |    |    |    |    |    |    |    |
|          | A1   | 3  |    |    |    |    |    |    |    |    |    |    |    |
|          | A1A2 |    |    | 4  | 3  | 2  | 5  | 1  | 4  | 3  | 2  | 5  | 1  |
|          | A1A3 |    |    | 1  | 4  | 3  | 1  | 5  | 2  | 5  | 4  | 2  | 3  |
|          | A1A4 |    |    | 2  | 5  | 4  | 2  | 3  | 1  | 1  | 3  | 4  | 5  |

TABLE.--Continued

|     |      | C1 |    | C2 |    | C4 |    | C3 |    | C6 |    | C5 |    |
|-----|------|----|----|----|----|----|----|----|----|----|----|----|----|
|     |      | B1 | B2 | B4 | B6 | B4 | B6 | B4 | B6 | B4 | B6 | B4 | B6 |
| II  | A1   |    | 4  |    |    |    |    |    |    |    |    |    |    |
|     | A1   |    | 5  |    |    |    |    |    |    |    |    |    |    |
|     | A1   |    | 2  |    |    |    |    |    |    |    |    |    |    |
|     | A1A2 |    |    | 1  | 2  | 3  | 5  | 4  | 1  | 3  | 5  | 2  | 4  |
|     | A1A3 |    |    | 3  | 4  | 2  | 1  | 5  | 2  | 5  | 4  | 3  | 1  |
|     | A1A4 |    |    | 4  | 5  | 1  | 3  | 2  | 5  | 4  | 2  | 1  | 3  |
|     |      | C1 |    | C2 |    | C6 |    | C5 |    | C4 |    | C3 |    |
|     |      | B1 | B2 | B5 | B3 | B5 | B3 | B5 | B3 | B5 | B3 | B5 | B3 |
| III | A1   | 2  |    |    |    |    |    |    |    |    |    |    |    |
|     | A1   | 1  |    |    |    |    |    |    |    |    |    |    |    |
|     | A1   | 4  |    |    |    |    |    |    |    |    |    |    |    |
|     | A1A4 |    |    | 4  | 3  | 1  | 5  | 3  | 2  | 4  | 5  | 1  | 2  |
|     | A1A3 |    |    | 3  | 1  | 5  | 2  | 4  | 4  | 5  | 3  | 2  | 1  |
|     | A1A2 |    |    | 1  | 4  | 2  | 3  | 2  | 5  | 3  | 1  | 4  | 5  |
|     |      | C1 |    | C2 |    | C4 |    | C6 |    | C3 |    | C4 |    |
|     |      | B1 | B2 | B6 | B4 | B6 | B4 | B6 | B4 | B6 | B4 | B6 | B4 |
| IV  | A1   |    | 1  |    |    |    |    |    |    |    |    |    |    |
|     | A1   |    | 3  |    |    |    |    |    |    |    |    |    |    |
|     | A1   |    | 4  |    |    |    |    |    |    |    |    |    |    |
|     | A1A4 |    |    | 5  | 1  | 1  | 5  | 2  | 4  | 3  | 4  | 2  | 3  |
|     | A1A3 |    |    | 2  | 5  | 2  | 5  | 1  | 3  | 4  | 1  | 4  | 3  |
|     | A1A2 |    |    | 2  | 3  | 2  | 4  | 1  | 5  | 3  | 1  | 5  | 4  |

## LIST OF REFERENCES

- Bilger, R.C., and Hirsh, I.J. "Masking of Tones by Bands of Noise." Journal of the Acoustical Society of America, 28 (1958), 623-630.
- Black, J.W. "Equally Contributing Frequency Bands in Intelligibility Testing." Journal of Speech and Hearing Research, 2 (1959), 81-83.
- Bocca, E., Calearo, C., Cassinari, V., and Migliavacca, F. "Testing 'Cortical' Hearing in Temporal Lobe Tumours." Acta Oto-Laryngologica, 45 (1955), 289-304.
- Calearo, C., and Antonelli, A.R. "'Cortical' Hearing Tests and Cerebral Dominance." Acta Oto-Laryngologica, 56 (1963), 17-25.
- Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M., and Gerstman, L.F. "Some Experiments on the Perception of Synthetic Speech Sounds." Journal of the Acoustical Society of America, 24 (1952), 597-606.
- Davis, H., and Stevens, R.S. Hearing Aids--An Experimental Study of Design Objective. Cambridge, Massachusetts: Harvard University Press, 1947.
- Delattre, P.C., Liberman, A.M., and Cooper, F.S. "Acoustic Loci and Transitional Cues for Consonants." Journal of the Acoustical Society of America, 27 (1955), 769-773.
- Edwards, A.L. Experimental Design is Psychological Research. New York: Holt, Rinehart, and Winston, Inc., 1960.
- \_\_\_\_\_. Statistical Methods. Second Edition. New York: Holt, Rinehart, and Winston, Inc., 1967.
- Egan, J.P., and Weiner, F.M. "On the Intelligibility of Bands of Speech in Noise." Journal of the Acoustical Society of America, 18 (1946), 435-441.
- \_\_\_\_\_. "Articulation Testing Methods." Laryngoscope, 58 (1948), 955-982.

- \_\_\_\_\_, Carterette, E.C., and Thwing, E.J. "Some Factors Affecting Multi-Channel Listening." Journal of the Acoustical Society of America, 26 (1954), 774-782.
- Fairbanks, G. "Test of Phonemic Differentiation: The Rhyme Test." Journal of the Acoustical Society of America, 30 (1958), 596-600.
- Flanagan, J., and Levitt, H. "Speech Interference from Community Noise." American Speech and Hearing Association Report Number 4, 1969, 167-174.
- Fletcher, H. Speech and Hearing. Princeton, New Jersey: D. Van Nostrand Company, 1929.
- Franklin, B. "The Effect of a Low-Frequency Band (240-480 Hz) of Speech on Consonant Discrimination." Unpublished Ph.D. dissertation, The City University of New York, 1969.
- \_\_\_\_\_. "The Effect on Consonant Discrimination of Combining a Low-Frequency Passband in One Ear and a High-Frequency Passband in the Other Ear." Journal of Auditory Research, 9 (1970), 365-378.
- French, N.R., and Steinberg, J.C. "Factors Governing the Intelligibility of Speech Sounds." Journal of the Acoustical Society of America, 19 (1947), 90-119.
- Haggard, M., Ambler, S., and Callow, M. "Pitch as a Voicing Cue." Journal of the Acoustical Society of America, 47 (1970), 613-617.
- Harris, K.S. "Cues for the Discrimination of American English Fricatives in Spoken Syllables." Language and Speech, 1 (1958), 1-7.
- Hayashi, R., Ohta, F., and Morimoto, M. "Binaural Fusion Test: A Diagnostic Approach to the Central Auditory Disorders." International Audiology, 5 (1966), 133-135.
- Hirsh, I.J., and Bowen, W.D. "Masking of Speech by Bands of Noise." Journal of the Acoustical Society of America, 25 (1953), 1175-1180.
- \_\_\_\_\_, Reynolds, E.G., and Joseph, M. "Intelligibility of Different Speech Materials." Journal of the Acoustical Society of America, 26 (1954), 530-538.

- House, A.S., Williams, C.E., Hecker, M.H., and Kryter, K.D. "Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set." Journal of the Acoustical Society of America, 37 (1965), 158-166.
- Hughes, G.W., and Halle, M. "Spectral Properties of Fricative Consonants." Journal of the Acoustical Society of America, 28 (1956), 303-310.
- Huizing, H.C., and Taselaar, M. "Experiments on Binaural Hearing." Acta Oto-laryngologica, 53 (1961), 151-154.
- Jerger, J.F. "Observations on Auditory Behavior in Lesions of the Central Auditory Pathways." Archives of Otolaryngology, 71 (1960), 65/797-74/806.
- Kimura, D. "Some Effects of Temporal-Lobe Damage on Auditory Perception." Canadian Journal of Psychology, 15 (1961a), 156-165.
- \_\_\_\_\_. "Cerebral Dominance and the Perception of Verbal Stimuli." Canadian Journal of Psychology, 15 (1961b), 166-171.
- \_\_\_\_\_. "Left-Right Differences in the Perception of Melodies." Quarterly Journal of Experimental Psychology, 15-16 (1963-1964), 355-358.
- Kirk, R.E. Experimental Design: Procedures for the Behavioral Sciences. Belmont, California: Brooks/Cole Publishing Company, 1968.
- Kryter, K.D. "On Predicting the Intelligibility of Speech from Acoustical Measures." Journal of Speech and Hearing Disorders, 21 (1956), 208-217.
- \_\_\_\_\_. "Methods for the Calculation and Use of the Articulation Index." Journal of the Acoustical Society of America, 34 (1962), 1689-1702.
- \_\_\_\_\_. "Speech Bandwidth Compression Through Spectrum Selection." Journal of the Acoustical Society of America, 32 (1960), 547-556.
- Leckie, D., and Ling, D. "Audibility with Hearing Aids having Low Frequency Characteristics." Volta Review, 1 (1968), 83-86.

- Lehiste, I., and Peterson, G. "Linguistic Considerations in the Study of Speech Intelligibility." Journal of the Acoustical Society of America, 13 (1959), 280-286.
- Liberman, A.M., Delattre, P.C., and Cooper, F.S. "The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants." American Journal of Psychology, 65 (1952), 497-516.
- \_\_\_\_\_, 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." Psychological Monographs: General and Applied, 68 (1954), 1-12.
- \_\_\_\_\_, Delattre, P.C., and Cooper, F.S. "Some Cues for the Distinction between Voiced and Voiceless Stops in the Initial Position." Language and Speech, 1 (1958), 153-167.
- Licklider, J.C.R. "The Influence of Interaural Phase Relations Upon the Masking of Speech by White Noise." Journal of the Acoustical Society of America, 20 (1948), 150-159.
- Linden, A. "Distorted Speech and Binaural Speech Resynthesis Tests." Acta Oto-laryngologica, 58 (1964), 32-48.
- Lisker, L. "Minimal Cues for Separating /w,r,l,y/ in Intervocalic Position." Word, 13 (1957), 256-267.
- Malecot, A. "Acoustic Cues for Nasal Consonants: An Experimental Study Involving a Tape-Splicing Technique." Language, 32 (1956), 274-284.
- Matzker, J. "Two New Methods for the Assessment of Central Auditory Functions in Cases of Brain Disease." Annals of Otology, Rhinology, and Laryngology, 68 (1959), 1185-1197.
- \_\_\_\_\_. "The 'Binaural Test'." International Audiology, 1 (1962), 209-211.
- Menon, K.M.N., Jensen, P.J., and Dew, D. "Acoustic Properties of Certain VCC Utterances." Journal of the Acoustical Society of America, 46 (1969), 449-457.
- Miller, G.A. "The Masking of Speech." Psychological Bulletin, 44 (1947), 105-129.

- \_\_\_\_\_, and Nicely, P.E. "An Analysis of Perceptual Confusions among Some English Consonants." Journal of the Acoustical Society of America, 27 (1955), 338-352.
- Milner, B. "Laterality Effects in Audition." Interhemispheric Relations and Cerebral Dominance. Mountcastle, Baltimore: John Hopkins, 1962, 177-195.
- 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 (1957), 24-43.
- Ohta, F., Hayashi, R., and Morimoto, M. "Differential Diagnosis of Retrocochlear Deafness: Binaural Fusion Test and Binaural Separation Test." International Audiology, 4 (1967), 58-62.
- Palva, A. "Filtered Speech Audiometry, I. Basic Studies with Finnish Speech Toward the Creation of a Method for the Diagnosis of Central Auditory Disorders." Acta Oto-laryngologica. Supplementum 210, 1965, 1-86.
- Pollack, I. "Effects of High Pass and Low Pass Filtering on the Intelligibility of Speech in Noise." Journal of the Acoustical Society of America, 20 (1948), 259-266.
- Shankweiler, D., and Studdert-Kennedy, M. "Identification of Consonants and Vowels Presented to Left and Right Ears." Quarterly Journal of Experimental Psychology, 19 (1967), 59-62.
- Speaks, C., Jerger, J., and Jerger, S. "Performance-Intensity Characteristics of Synthetic Sentences." Journal of Speech and Hearing Research, 9 (1966), 305-312.
- \_\_\_\_\_. "Intelligibility of Filtered Synthetic Sentences." Journal of Speech and Hearing Research, 10 (1967), 289-298.
- Steinberg, J.C. "Effects of Distortion upon the Recognition of Speech Sounds." Journal of the Acoustical Society of America, 1 (1929), 121-137.
- Stevens, S.S., ed. Handbook of Experimental Psychology. New York: John Wiley and Sons, Inc., 1951.

Studdert-Kennedy, M., and Shankweiler, D. "Hemispheric Specialization for Speech Perception." Journal of the Acoustical Society of America, 48 (1970), 579-594.

The Young Deaf Child: Identification and Management,  
Acta Oto-Laryngologica Supplementum 206, 1965.

Webster, J.C. "Important Frequencies in Noise-Masked Speech." Archives of Otolaryngology, 80 (1964), 494-504.