INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A

EFFECTS OF INTER-TALKER DIFFERENCES ON SPEECH PERCEPTION

by

BETHANY A. MULHEARN

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

1999

UMI Number: 9924831

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

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

4/16/99
Date

ARTHUR BOOTHROYD, Ph.D.
Chair of Examining Committee

4/22/99
Date

IRVING HOCHBERG, Ph.D.
Executive Officer

HARRY LEVITT, Ph.D.

LAWRENCE RAPHAEL, Ph.D.

Supervisory Committee

KATHERINE S. HARRIS, Ph.D.

External Examiner

The City University of New York

ABSTRACT

EFFECTS OF INTER-TALKER DIFFERENCES ON SPEECH PERCEPTION

by

Bethany A. Mulhearn

Advisor: Arthur Boothroyd, Ph.D.

Clinical measures of the speech perception of hearing-impaired listeners typically use one talker who may not be representative of the "average talker". While inter-talker differences in the acoustic properties of speech are known, the contribution of these differences to the confidence limits of speech perception measures has not been established. The goals of this study were: a) to measure inter-talker differences in terms of key parameters of the distribution of information across frequency, b) to establish the acoustic correlates of these differences, and c) to estimate their potential contribution to the confidence limits of test measures.

Consonant-vowel-consonant words were produced by three adult males and three adult females. Phoneme recognition was measured in normally hearing adults as a function of cut-off frequency for low- and high-pass filtering. For each talker, the following was derived: a) "cross-over frequency" for the two functions, i.e., the point where the frequency spectrum is divided into two equally intelligible parts, b) "half-band phoneme intelligibility, i.e., phoneme recognition at cross-over, and c) "full-band phoneme intelligibility", derived from half-band intelligibility by application of the articulation index theory.

Mean cross-over frequency for male talkers was 1599 Hz, with 95% confidence limits of +/-117 Hz. Mean cross-over frequency for female talkers was 1806 Hz, with 95% confidence limits of +/-115 Hz. The difference of 13% was statistically significant (p=0.027). There were also inter-subject differences of half-band and full-band intelligibility, with no evidence that these differences were gender-related. The acoustic data were consistent with the hypothesis that higher cross-over frequency for females results from higher formant frequencies. Using data obtained under low-pass filtering, it was found that inter-talker differences contribute significantly to the confidence limits of phoneme recognition. That contribution is considerably less, however, than that of inherent test-retest variability of speech recognition scores when they are based on the small number of test items typical of clinical practice. If inherent variability can be reduced, it may then be useful to seek ways of minimizing the effects of inter-talker differences.

ACKNOWLEDGMENTS

While this work bears the name of one author, it has been completed through the help and support of many people. These acknowledgments are a small token of very large gratitude. Easiest to thank is my family, Mike and Judie Mulhearn, Jeremy Milner, Chris Mulhearn, Christine Ahern, and Patricia Bennett. This group deserves special thanks for standing by and supporting me without asking when I would be finished.

I am indebted to Dr. Irving Hochberg for the support of the Graduate Program in Speech and Hearing Sciences, allowing me the opportunity to learn in a wonderful environment. Dr. Hochberg has been a steady source of encouragement and wisdom.

Many thanks to my committee, Dr. Harry Levitt, and Dr. Larry Raphael and my external examiner, Dr. Katherine Harris. It has truly been my pleasure to learn from such distinguished professors as well as to know three fine people. Their expertise and commitment will always serve as inspiration.

To my colleagues and friends at the Graduate Center - Brett Martin, Gary Chant, Loretta Walker, and particularly Eddy Yeung - your friendship and knowledge are qualities only matched by your willingness to share them. To Jodi Ostroff, who is crossing the finish line by my side as she has been from the start, my deepest thanks.

A very special thanks to my advisor and mentor of the past five years, Arthur Boothroyd. It is hard to describe my gratitude for the immense amount of time he has spent teaching and guiding me. His idea of what a mentor is should be an example to all. In my own career, I can only hope to follow his example of leadership, intelligence, and generosity.

TABLE OF CONTENTS

CHAPTER **PAGE**

APPENDICES

LIST OF TABLES

LIST OF FIGURES

FIGURE **PAGE**

CHAPTER ONE

INTRODUCTION AND PURPOSE

I. General Purpose:

The general goal of this study was to measure the effect of inter-talker differences on speech perception.

II. Definition of Terms:

In this document, the following definitions are to be assumed:

a) Speech perception is the process by which a listener infers language patterns from the sound patterns generated by a talker.

b) Phoneme intelligibility is the proportion of phonemes correctly identified by the listener in lists of consonant-vowel-consonant words.

III. Background and Rationale:

The motivation for this study arose from the need for valid tests of auditory speech perception for listeners with sensorineural hearing loss. These tests are necessary for diagnosis, quantification, and description of hearing disorders. Additionally, they are used to assess the need for sensory aids, to select appropriate characteristics for those aids, and to measure and compare their benefits.

Speech perception tests usually use one talker to produce the stimuli. The use of one talker, however, does not take differences among talkers into account. Acoustic characteristics of vowels differ on the basis of talker age and gender (Eguchi & Hirsh, 1969; Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). Also, there are gender and age differences in fricative spectra (Boothroyd, Erickson, and Medwetsky, 1994; Boothroyd & Medwetsky, 1992; McGowan and Nittrouer, 1988). Differences in these variables are also seen within gender and age groups.

Listeners with normal hearing have access to the full range of frequencies and intensities needed to perceive speech accurately. Persons with a sensorineural hearing loss must listen with limited access to the spectral and temporal information found in speech. The range of intensities available within a certain frequency range can be reduced - sometimes to zero. There may also be a loss of spectral and temporal resolution. In addition to the distortion caused by the hearing loss, there may also be distortion inherent in the sensory aid used for increased audibility of the speech signal.

It is possible that acoustic differences among talkers interact with an individual's hearing loss and/or sensory aid. That is, a listener may experience differences in speech perception performance for different talkers. These inter-talker differences may also affect individual listeners differently. Researchers have looked at the effect of inter-talker differences on word recognition in listeners with sensorineural hearing loss (Bess, 1983; Brandy, 1966; Gengel & Kupperman, 1980; Hood & Poole, 1980; Kirk, Pisoni, & Miyamoto, 1997; Kruel, Bell, & Nixon, 1969; Penrod, 1979). These studies found that varying talkers during word recognition testing influenced word recognition scores for listeners with hearing loss.

IV. Specific Aims:

This study assessed the effect of inter-talker differences on speech perception in normal hearing listeners under conditions of reduced spectral information.

Specific goals were:

1. To measure phoneme intelligibility for a sample of adult talkers under various conditions of high- and low-pass filtering.
2. To determine, for each talker:
    i) the "cross-over" frequency that divides the speech spectrum into two bands having equal contribution to phoneme intelligibility.
    ii) phoneme intelligibility for the two bands so established (half-band intelligibility).
    iii) full-band intelligibility, as estimated from half-band intelligibility.
3. To determine the contribution of inter-talker differences to cross-over frequency and to each measure of intelligibility obtained under 1 and 2.
4. To estimate the contribution of gender to the inter-talker differences so measured.
5. To test various objective measures of the acoustic speech signal as potential predictors of the inter-talker differences measured under 1-3.

CHAPTER TWO

LITERATURE REVIEW

The literature relevant to this study falls into two main categories: the effects of age and gender on the acoustic properties of speech and the effects of high and low pass filtering on speech perception in listeners with normal hearing.

I. Effect of Age and Gender on the Acoustic Properties of Speech

It is obvious that the speech and voices of men, women, and children sound different to a listener. Many studies have investigated the acoustic aspects of a talkers articulation and voice that lead to a listener identifying the speaker as adult or child. male or female. Another question that has been addressed is how listeners are able to perceive phonologically identical sounds across speakers in spite of their differing acoustic characteristics. The research reviewed here looks at which acoustic characteristics of speech are used for: identification of a talker's age and gender, identification of individual sounds of speech produced by different talkers, and identification of individual speakers.

Age and gender effects on the acoustic properties of speech have been investigated by looking at two main topics. The first topic is the source of speech: the glottal sound source including the fundamental frequency. The glottal sound source is the vibratory activity of the vocal folds. Variations in the glottal sound source have been studied by looking at the instantaneous source spectrum and the long term average spectrum of speech. The fundamental frequency is the lowest frequency of the glottal sound source and is the principal determinant of the perceived pitch of the voice. It is determined by several factors, including anatomy of the speaker, his/her characteristic subglottal pressure and vocal fold tension, and the moment-to-moment variations of these used to convey prosodic information. The second topic is the filter through which sound must travel to create speech sounds - namely, the various configurations of the vocal tract. Vocal tract resonances are manifested as formant frequency values, which are the principal acoustic characteristics of vowels and resonant consonants. These formant frequencies are altered with changes in vocal tract size and shape that result from articulation. Formant frequencies are also susceptible to anatomical variations that occur as a function of speaker age and gender and individual anatomy. Additionally, the difference in filter characteristics as a function of talker age and gender have been studied with regard to fricative production and perception.

Monsen and Engebretson (1977) investigated the laryngeal vibration patterns of five male and five female talkers who produced the vowel /ə/ into a long reflectionless tube. Productions were made using seven vocal registers: normal, loud, soft, creaky, falsetto, rising/falling fundamental frequency, and stressed/unstressed syllable. For both groups, the slope of the spectra was within the -12 to -18 dB/octave range. There was a difference, however, in the steepness of the slope of the spectrum for the two groups. Male talkers' spectra fell off approximately -12 dB/octave while the female talkers' spectra fell off approximately -15 dB/octave. Also, for all seven vocal registers, the fundamental frequency for female talkers was about 1 octave higher than that for male talkers. A methodological concern about this study involves the coupling of the metal tube to the speakers' mouth. Because of the teeth, the vocal tract is not uniform and will not be smoothly coupled to the tube. When the mouth is put to the tube, there is some discontinuity between the vocal tract and the glottal tube. Therefore, resulting waveforms will be somewhat affected by vocal tract shape.

Amplitude differences were found between glottal spectra of males and females in the Monsen and Engebretson study. The average rms intensity of the glottal wave of female talkers in this study was -6 dB relative to that of male talkers.

Eklund and Traunmuller (1997) found results similar to Monsen and Engebretson's (1977). Their study analyzed confusions in vowel quality and talker gender for whispered and phonated versions of the long vowels of Swedish. They found that above 2.15 kHz, the energy of vowels phonated by female talkers was slightly weaker than the energy of vowels phonated by male talkers.

Klatt and Klatt (1990) found that, on average, female voices showed more breathiness compared to male voices. This breathier quality in female talkers lead to a higher degree of noise excitation in the higher formants and thus less harmonic energy in those formants. The authors suggested that the energy in the higher frequency range may serve as a salient cue to identifying a speaker as female.

Boothroyd, et. al (1994), looked at gender differences in the long term average spectrum of speech as well as inter-talker spectral envelope differences. The long term average spectrum of speech is the average sound pressure level (SPL) in decibels (dB) as a function of frequency of the signal within a certain bandwidth. This study located frequencies and intensities of seven phonemes via 13 spectral points within 1/3 octave bands produced by male and female talkers. The long term average level found in the analysis was 68 dBSPL, which the authors noted concurred with the literature. The average intensity of the male talker exceeded that of the average female talker in all but one spectral point. Also, there were gender related inter-talker differences in the shape of the spectral envelope. These data are consistent with those reported by Pearsons, Bennett, and Fidell (1977) cited in Levitt and Webster (1991), where, for five different modes of vocal effort produced by male and female talkers, the average long term speech level of the males was higher than that of the female talkers. Pearsons et. al, also reported spectral shape differences; female speakers' spectra had substantially less power than the males' below 200 Hz.

Fundamental frequency, another source characteristic, has been demonstrated in the literature to vary according to age and gender. Two complementary studies (Fairbanks, Wiley, and Lassman, 1949 and Fairbanks, Herbert, and Hammond, 1949) examined fundamental frequency in 7 and 8 year olds. The fundamental frequency values were similar for the 7 and 8 year old boys and girls (average 294 and 297 Hz, respectively, for the boys and average 281 and 288 Hz, respectively, for the girls). The mean fundamental frequency value of each group was about 1 octave higher than that of a typical adult male. The gender similarities indicate that in this age range, it is possible to characterize fundamental frequency as a function of age alone, without distinguishing gender. Interestingly, the study of the 7 and 8 year old females by Fairbanks, Herbert, and

Hammond (1949) found that the female talkers demonstrated voice break[1] similar to that of the 7 and 8 year old boys. This is striking because it indicates that not only is voice break not exclusive to adolescent speakers; it is also not exclusive to male speakers.

Curry (1940) compared fundamental frequency in 10-,14-, and 18-year-old male talkers. Six talkers at each age level produced a 55-word passage from which the mean and median fundamental frequency values were calculated. Curry found that fundamental frequency fell as age increased; the difference between the values for the 14- and 18-year-olds was about 1 octave. Also, while voice breaks were noted at about the same rate for the 10- and 14-year-olds, none were noted for the 18-year-olds. Taken with the aforementioned data by Fairbanks, et al. (1949), it appears that voice breaks are produced by both pre-adolescent boys and girls, as well as adolescent boys, but not by adults.

Hollien and Paul (1969) and Michel, Hollien and Moore (1966) also measured the fundamental frequency in 15-, 16-, and 17-year-old female talkers. The talkers produced a portion of the "Rainbow Passage" (Fairbanks, 1960). The fundamental frequency of the passage was measured by an electronic counter which sampled the period of the signal every 33 milliseconds. The data from these two studies revealed mean fundamental frequency values in the range from 207.5-215.7 Hz with non-significant differences between the age groups. It was concluded, therefore, that female fundamental frequency reached adult-like values by age 15. This finding is different from that for male adolescents, who were still demonstrating voice break at a similar age (Curry, 1940.) A potential problem with the Michel, et al. (1966) data arises from the subject group, which

---

[1] In these studies, the term voice break refers to a brief change of vocal register resulting in a temporary fundamental frequency change of about 1 octave.

was a group of cheerleaders. These "trained speakers" may not be representative of the average teenage female speaker.

Studies looking at fundamental frequency in elderly males and females introduced interesting gender differences as a function of advanced age. Mysak (1959) looked at pitch and duration characteristics of males in two age groups: ages 65-79 and ages 80-92. The fundamental frequency values for these 2 groups were then compared to those of 15 middle-aged sons of subjects from the elderly groups. The fundamental frequency increased with age. One possible reason is changing physiology, such as reduced muscle mass and tension. The mean fundamental frequency value in the 80-92 year old group fundamental was 136.5 Hz. The mean fundamental frequency value in the 65-79 year-old group was 120.2 Hz. While the fundamental frequency values of the middle-aged sons were lower than their elderly fathers, no significant family relationships were found.

Hollien and Shipp (1972) investigated fundamental frequency changes and chronological age in male talkers. They used groups of 25 talkers in each of seven decades: 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80-89. Each talker produced the first paragraph of the "Rainbow Passage" (Fairbanks, 1960). Mean fundamental frequency values taken from the paragraph showed an interesting trend: fundamental frequency decreased with increasing age until 40-49 years old; then it began to rise with increasing age, creating a "saucer shaped" function. The total range for all fundamental frequency values was 107.1-146.3 Hz; the lowest fundamental frequency value was found for the 40-49 year olds and the highest value for the 80-89 year olds.

The results of this study are especially fascinating in light of a study by McGlone and Hollien (1963) which looked at fundamental frequency in females of advanced age.

Recall that an adult female fundamental frequency value was seen in 15 year old females (Hollien and Paul, 1969). McGlone and Hollien studied 20 female talkers in two groups of ten: ages 65-79 and 80-94 (note the similarity to the Mysak, 1959, data). The talkers produced the first paragraph of the "Rainbow Passage" (Fairbanks, 1960). Average fundamental period of the voice within1/26-second time intervals was measured. These two groups were not significantly different from each other nor were they from earlier data from Linke (1953, unpublished doctoral dissertation) data on young adult females.

The McGlone and Hollien (1963) data for female speakers, taken together with earlier data by Fairbanks, et al. (1949), Michel et al. (1966) and Hollien and Paul (1969) reveals the following trend: by adolescence, fundamental frequency has reached an adult level and remains at that level throughout a female's life. This trend is different from the saucer-shaped function seen for male speakers.

While many of the studies described above reported the fundamental frequency value as a mean fundamental frequency value from running speech (often a portion of the "Rainbow Passage"), there are data from other studies that used different stimuli.

A classic study by Peterson and Barney (1952) reported fundamental frequency values for males (n=33), females (n=28), and children (n=15) for ten vowels produced in an /hVd/ context. The child talkers consistently had the highest fundamental frequency for all vowels. The fundamental frequency values of the vowels spoken by adult female talkers were lower than those of the children, and adult male talkers had lowest fundamental frequency values of the three groups. The results for the child talkers in this study are somewhat difficult to interpret, however, because the ages and genders of the children were not specified. As seen in the studies described earlier in this review (Curry, 1940; Fairbanks, Herbert, & Hammond, 1949; Fairbanks, Wiley & Lassman, 1949; Hollien and Paul, 1969; and Hollien and Shipp, 1972), there is a clear age effect on fundamental frequency for talkers of both genders.

Eguchi and Hirsh (1969) synthesized and replicated some of the Peterson and Barney findings in their study looking at fundamental frequency for six isolated vowels in a group of children, categorized by either age alone or sex and age, and adults. From ages 3-10, male and female children were put into the same groups. From 11-13 years old, children were separated by age and gender. There were also two other groups: adult males and adult females. Fundamental frequency was found to decrease as age increased, with a marked decrease from ages 3-6. Not coincidentally, this is an age range during which rapid lengthening and thickening of the vocal folds takes place. Consistent with studies discussed earlier, 13-year-old girls had fundamental frequency values (mean = 239.8 Hz) similar to those of adult females. Thirteen year old boys' fundamental frequency values (mean = 221.1 Hz) were still above those of the adult male speakers (mean = 124.2 Hz), indicating that more fundamental frequency decrease was yet to occur before the boys' values reached adult values. Also, the fundamental frequency values of the 13-year-old boys had the greatest between- subject variability. This finding supports the idea of fundamental frequency as a "work in progress" for this particular age and gender group.

Hillenbrand, et. al, (1995) did a study that replicated the work of Peterson and Barney (1952). The talkers in this study, 45 males, 48 females, and 46 children in the 10-12 year old age group, produced the same ten vowels in the same vowel contexts as the talkers in the Peterson and Barney study. When compared to the fundamental frequency values found by Peterson and Barney, the adult fundamental frequency values reported by

Hillenbrand, et. al were similar, but the fundamental frequency values for the children in the Hillenbrand, et. al study were about 28 Hz lower than those found by Peterson and Barney. It is possible that the ages of the children influenced the fundamental frequencies of the child talkers; a comparison of the two studies cannot be made without knowing the ages of the children in the Peterson and Barney study. Coleman (1976) studied the manner in which fundamental frequency affects a listener's perception of a voice as male or female. Although the contributions of both fundamental frequency and vocal tract resonances were examined in this study, only fundamental frequency will be discussed here. Vocal tract resonances will be discussed later in this review. Twenty-five listeners decided if phrases and vowels were produced by a male or female talker. The fundamental frequency value was most closely correlated (r=0.94) with a listener's perception of male/female quality, indicating the importance of fundamental frequency in identifying a speaker as male or female.

In a study described above, Eklund and Traunmuller (1997) looked at fundamental frequency and gender differences in phonated and whispered Swedish vowels. Listeners were asked to identify both vowels and speaker sex in the whispered and phonated conditions. In the whispered condition, fundamental frequency is absent (because of vocal fold abduction) and therefore unavailable to cue speaker gender. Phonated vowels were correctly identified more often than whispered, and male-produced vowels were correctly identified more often than female-produced vowels. Vowel identification and speaker gender were only confused with each other in the whispered condition. The authors concluded that subjects were using fundamental frequency to cue both vowel identity and talker gender.

In concluding the discussion of age and gender differences in fundamental frequency, some general statements can be made. First, fundamental frequency decreases as a function of increasing age for both male and female talkers. The changes occur in different patterns, however, for the two groups. Adult female fundamental frequency is generally lower than the fundamental frequency of child talkers and higher than the fundamental frequency of adult male talkers. The male-female difference is approximately 1 octave. Lastly, fundamental frequency can act as an important cue to the correct perception of a talker's gender.

We now turn to age and gender differences of the voice filter, or vocal tract. The vocal tract assumes many different configurations as a result of the various articulator positions realized for sound production. Each of these configurations contributes to different vocal tract resonances, or formant frequency values. While the same relative articulatory positions are employed across speakers of a language when they produce identical sounds, there is some positional variability that manifests itself as differences in formant frequency values. Both age and gender are partly responsible for this variability; because head and neck size (which in turn influence the anatomy of the vocal tract) differ, formant frequency values differ.

The study by Peterson and Barney (1952) which investigated fundamental frequency values for adult males, adult females, and children, also identified the first, second, and third formant frequency values for these same three groups. The child talkers had the highest formant values across the ten vowels studied. Their first formant values were about 1/2 octave higher than those of the male talkers; second and third formant values were substantially higher as well. Female talkers' formant frequencies were in between those of males and children. Again, as stated earlier, lack of specific age and gender data for the child talkers makes it impossible to ascertain up to what point a child's formant frequency values deviate from adult values.

Eguchi and Hirsh (1969), in the study described earlier, also looked at first, second, and third formant frequency values for adult male talkers, adult female talkers, and child talkers. Formant frequency values were examined for six vowels for these three groups. Formant frequencies of the children were about 25% higher than those of adult males and 20% higher than those of adult females. While both first and second formant frequencies decreased as age increased, the second formant tended to have a greater drop than the first formant. The patterns of formant frequency values are similar to those for fundamental frequency: there was rapid change from ages 3-5 (a time when there is rapid anatomical vocal tract development), and the values of 13-year-old girls were similar to those of adult females while those of 13-year-old boys were still higher than adult males.

Formant frequency differences can also be used to determine the gender of the talker. The study by Coleman (1976) (described earlier) contained a second experiment where a laryngeal vibrator at 120 and 240 Hz (representing the average male and female fundamental frequency, respectively) was substituted for speakers' fundamental frequencies in order to examine the role of vocal tract resonances on perception of gender of a speaker. First, second, and third formant frequency values were averaged for each of four vowels and presented to the listener in one of four combinations: 1.)low vocal tract resonances/low fundamental frequency (2 male characteristics), 2.)high vocal tract resonances/high fundamental frequency (2 female characteristics), 3.)high vocal tract resonances/low fundamental frequency, and 4.)low vocal tract resonances/high fundamental frequency. Twenty-five listeners judged the voice as male or female. For the two clear male and two clear female characteristic conditions, the listeners almost always identified the speaker gender correctly (100% for male characteristics, 98% for female characteristics). In both of the mixed or ambiguous conditions, however, the voices were more likely to be judged as male even in the presence of a female characteristic, indicating perceptual prominence of a male vocal tract characteristic (i.e., lower fundamental and formant frequency values). Therefore, in this study the female fundamental frequency was perceptually weaker than the male vocal tract resonances while the male fundamental frequency retained its perceptual prominence.

Another study by Coleman (1971) eliminated between-gender differences of fundamental frequency by substituting an electrolarynx at 85 Hz for the fundamental frequency. The purpose of the study was to measure the contribution of the vocal tract resonances alone to the perception of the voice as male or female. Values for first, second, and third formants were averaged together and treated as "vocal tract resonance characteristics". Consistent with other data (Peterson and Barney, 1952; Eguchi and Hirsh, 1969; Hillenbrand, et. al, 1995), all of the female talkers had higher formant frequencies than the male talkers. In the absence of fundamental frequency variation, all male talkers and all but two female talkers were judged to have voice quality appropriate to their gender. Classification of a talker as male occurred more often and with more precision. However, while the author intended to remove the influence of fundamental frequency from the study, it still may have been a factor. Although the 85 Hz tone of the electrolarynx was lower than the average fundamental frequency of either male or female talkers, it was closer to the fundamental frequency of male talkers. Therefore, while male vocal tract resonances may indeed be a stronger cue to voice quality perception than female vocal tract resonances, the frequency of the electrolarynx must be kept in mind. It can be concluded from both of Coleman's studies that a combination of fundamental frequency and vocal tract resonances contribute to perception of the voice as male or female. One question that arises from both of the Coleman studies involves the validity of averaging the first, second, and third formants and calling this average "vocal tract resonances". This procedure uses the assumption that each formant is perceptually weighted equally by a listener and can thus be averaged. Further study of this question is required.

While it is often suggested that male/female vocal tract resonance differences are due to anatomical differences, a study by Diehl, Lindblom, Hoemeke, and Fahey (1996), looked for an explanation beyond anatomy. The authors' hypothesis was that behavioral characteristics of a particular gender were the cause of certain articulatory adjustments in the vocal tract. In particular, the authors looked at fundamental frequency as a function of gender in order to investigate the tendency of female talkers, with a higher fundamental frequency, to exhibit greater between category vowel dispersion in the F1 x F2 plane than male talkers. The "sufficient contrast hypothesis" asserts that this greater dispersion is a compensatory effect needed before higher fundamental frequency causes reduced vowel identifiability due to sparser harmonic sampling of spectral envelopes. Subjects identified two synthesized vowels, /I/ and /U/. It was found that once fundamental frequency reached about 150 Hz, identification accuracy of vowels decreased as fundamental frequency increased. It was also found that accuracy of vowel identification decreased as fundamental frequency increased more for the female-like formant patterns than the male-like patterns. According to these findings, the higher fundamental frequency of the average female talker (about 200-250 Hz) should create difficulty identifying vowel sounds. The authors concluded that since a listener is usually able to perceive correctly the vowels produced by a female talker (with a higher fundamental frequency), there must be some type of compensation for the reduced identification accuracy demonstrated in this study. The authors discuss the fact that higher fundamental frequency is often part of a large fundamental frequency excursion which, within a syllable, can greatly reduce the deleterious effects of the high fundamental frequency. Thus, they conclude that the greater first formant x second formant dispersion seen in female talkers is a compensatory strategy that enables female speakers to offset the negative effects of the high fundamental frequency.

Effects of age and gender differences on vocal tract configuration have also been studied by looking at differences in fricative productions. Boothroyd and Medwetsky (1992) investigated the spectral distribution of the fricative energy of /s/ and examined the gender, subject, and vowel context effects. They analyzed intervocalic samples of /s/ in nonsense strings and found marked gender differences. For male talkers the average value of the lowest frequency prominent spectral peak was about 1 octave lower than that of female talkers (4.3 kHz to 7.2 kHz, respectively). For male talkers, this value of 4.3 kHz was consistent with previous literature where the value was given has typically been repeated as 4 kHz, with no reference to gender. For female talkers, however, 4 kHz is obviously quite low in comparison. Large within-gender variations were also found in this study: the average within-gender frequency range of the fricative spectrum was about 2 kHz.

McGowan and Nittrouer (1988) compared fricative production of children and adults using the phonemes /ʃ/ and /s/. Gender differences and vowel contexts were analyzed as well. Adults and children of preschool to early school ages produced consonant-vowel-consonant-vowel nonsense strings using /s/ and /ʃ/ in combination with /i/ and /u/. Second formant frequency values were compared and the results were as follows: for all identical utterances, second formant values of female adults were higher than second formant values of male adults; for all identical utterances, second formant values of children were higher than second formant values of adults. Therefore, the spectral differences between these two fricatives followed the same age and gender difference patterns as those seen for vowels. As a potential explanation for the age and genderrelated second formant differences seen in this study, the authors point out that second formant will change according to shape and size of the cavities behind and in front of the constriction. Thus, when one speaker has a longer cavity, second formant values will be lower.

Nittrouer, Studdert-Kennedy, and McGowan (1989) did a study similar to the one just described where the acoustic structure of fricative-vowel syllables produced by children and adults was analyzed. The subjects produced fricative-vowel-fricative-vowel nonsense syllables using /ʃ/ and /s/ with the vowels /i/ and /u/. Centroids and second formant frequencies were computed for each. The analyses revealed that for all talkers, the /s/ spectra had a gradual rise across the whole frequency range and were weighted more heavily in the high frequencies with centroid values near 8 kHz. The /ʃ/ spectra were different, with a sharp rise followed by a flat or falling period with centroid values close to 6 kHz. For children, the /ʃ/ spectra rose slowly to a greater amplitude in the high frequencies than the adult with higher centroid values than the adult production. The children's /s/ and /ʃ/ spectra are less different from each other than in the adults productions; the difference between the spectra of the fricatives increased as a function of increasing age. All ages showed a vowel context effect (lower centroid values before /u/ than /i/), for /s/ but not /ʃ/. Older children showed a stronger differentiation for fricatives than younger children and adults showed a stronger differentiation than the children. In fact, as age decreased, each age group of children consistently differentiated less between fricatives. Second formant values decreased as a function of age with a significant drop occurring from age 7 to adulthood. Centroid and second formant differences are explained by the authors as anatomical differences. The centroid value is determined by an index of front cavity size and second formant is determined by length of cavity behind constriction; both of these will change as a child grows and thus, cavity size and length changes. The fricative differentiation differences in age may be accounted for by less control over mouth constriction and shape by younger children. Therefore, while children coarticulated the components of the fricative-syllables more precisely than adults, there was less evidence of precise syllabic detail as seen in adults. The authors note that the children's fricative judgments (unconstrained by motoric limitations) were more influenced by fricative-vowel transitions and less influenced by the steady-state fricative spectrum than were the judgments of the adults.

Age-related amplitude differences for fricatives have also been studied as part of some of the papers discussed here. The study by McGowan and Nittrouer (1988) looked at age-related amplitude differences in fricatives. The second formant amplitude of /ʃ/ was found to be higher than that for /s/ in both adults and children. The authors attribute this difference to physiological differences between children and adults, where children have a more prominent back cavity resonance in their fricative spectra than adults.

The literature described here indicates that age and gender affect the acoustics of speech in terms of the characteristics of the source and the filter of speech. Both fundamental frequency and formant frequencies decrease as a function of increasing age; although at different rates depending on speaker sex. Gender differences are also seen at the glottal sound source spectrum and long term average speech spectrum. Age and gender differences are also visible in production of fricative sounds. Therefore, there are numerous possibilities for cues to speaker age and gender.

II. Phonetic Level Perception by Normal Listeners of Filtered Speech

One approach to determining which frequencies of speech are most important for accurate speech perception has been to study individuals with normal hearing listening to low-pass and high-pass filtered speech. Results have often been reported in terms of an importance function, which describes the contribution of various frequency regions to speech intelligibility. One parameter of the importance function is the cross-over frequency, which divides the spectrum into regions that contribute equally to speech intelligibility.

The following review is divided into three sections. The first section contains studies in which the test materials used have affected the outcome of a test, in terms of both percent correct and crossover frequency. The next section contains studies where the talker gender has influenced the crossover frequency. The third section consists of studies detailing the performance of listeners perceiving filtered speech and consonant confusion patterns of those listeners. The end of this section will also include a discussion of methods using low- and high-pass filtering that have been developed to predict intelligibility of a speech signal.

Test materials can influence the outcome of speech perception studies. In the original study that described importance functions for intelligibility of speech, French and Steinberg (1947), used nonsense consonant-vowel-consonant syllables in several high- and low-pass filter conditions. The most important frequency region for accurate speech perception was found to be between 1500-2500 Hz. The cross-over frequency was 1930 Hz. This is among the highest reported in the body of literature reviewed here.

Hirsh, Reynolds, and Joseph (1954), looked at intelligibility of speech using several different speech materials. Subjects listened to nonsense consonant-vowel-consonant syllables, meaningful consonant-vowel-consonant syllables, spondaic words, disyllabic words of trochaic and iambic stress, polysyllabic (>2 syllables) words, and sentences consisting of five key words. The materials, produced by two male and two female talkers, were high-pass and low-pass filtered at six different settings each with four signal-to-noise ratios. For all filtering conditions, the highest and lowest articulation scores were found for polysyllabic words and nonsense syllables, respectively. Additionally, nonsense syllables and monosyllabic words were the first materials to be affected by filtering and were overall most susceptible to degradation by filtering. For all materials, however, the articulation functions relating recognition to cut-off frequency had similar slopes. For the high-pass conditions, the articulation function began to decrease with a cut-off frequency of 3200 Hz. For the low-pass conditions, the function began decreasing with a cut-off frequency of 800 Hz. Interestingly, at 6400 Hz cut-off for high-pass filtering, the articulation score was still about 50% for polysyllabic words. The authors attributed this finding to the areas enclosed by the filter skirt, which still included low frequencies. The cross-over frequency was nearly 1700 Hz, and was about the same for all materials. From this study, it appears that a speech material with more redundancy, such as polysyllabic words, leads to greater intelligibility.

Speaks (1967) looked at the intelligibility of filtered synthetic sentences (but natural speech) and then compared his results to previous work using monosyllabic stimuli. Three normal-hearing listeners heard third-order approximations to English sentences with several high-pass and low-pass filter settings and at several intensity levels. Performance, measured as the percent of correct sentences identified, improved as a function of both increased frequency range and increased intensity level. For this study, the cross-over frequency was found to be 725 Hz, a striking difference when compared to the 1700 Hz found by Hirsh, et al., (1954) and the 1900 Hz found by French and Steinberg, (1947). For sentences, a lower frequency range is needed for intelligibility than for single words. The authors concluded that, when compared to monosyllabic words, less of the spectrum was required to make sentences intelligible; that the critical region of the spectrum was lower than that for monosyllables; and that sentences were more resistant to degradation by filtering than monosyllables. It should be noted, however, that the response format was a ten-alternative closed-set response task which may have yielded a higher percent correct score than an open set recognition task. Another caveat regarding the findings of this study arises from the use of synthetic sentences as test material. Generalization of the results of this study to conversational speech should be done cautiously.

Studebaker and Pavlovic (1987) determined the frequency importance function for continuous discourse produced by one female and two male talkers. Listeners heard a thirty-second speech sample (from Cox and McDaniel, 1984) recorded under seven high-pass and low-pass conditions, a broadband condition, and nine speech-to-noise ratios, and then estimated the percentage of words understood. Transfer functions were derived to determine the relationship between Articulation Index and percent intelligibility. Transfer functions and frequency importance functions were then compared to previous data for different speech materials. Transfer functions indicated higher scores for the continuous discourse at the same Articulation Index values than for those used in studies with consonant-vowel-consonant nonsense syllables (French and Steinberg, 1947) and monosyllabic words (Black, 1959). Also, frequency importance functions were different for the continuous discourse results; the function was bimodal with 400-500 Hz as the frequency region of maximum importance. The cross-over frequency was notably lower than other studies as well: 1189 Hz as compared to 1900 Hz for consonant-vowel-consonant nonsense syllables (French and Steinberg, 1947), and 1700 Hz for polysyllabic words (Hirsh et al., 1954). When compared to the 725 Hz cross-over frequency for synthetic sentences (Speaks, 1967) discussed above, the result found by Studebaker, et al., (1987) is quite different. Both studies, however, support the conclusion that the cross-over frequency decreases as the redundancy of the test material increases.

Another factor that may affect intelligibility and cross-over frequency is the gender of the talker. Studebaker, et al. (1987), described above, questioned the contribution of talker gender to the importance functions they found in their study using one female and two male talkers. They analyzed the cross-over frequencies from several previous studies using male talkers only (Speaks, 1967; Wang, Reed, and Bilger, 1978), both male and female talkers (their study; French and Steinberg, 1947; Hirsh, et al., 1954) and female talkers only (Miller and Nicely, 1955 and Pollack and Pickett, 1964). They noted that studies using male talkers alone had lower cross-over frequencies than studies using male and female talkers or only female talkers. It must be noted, however, that the importance functions from these various studies were derived using different materials.

Another aspect of the literature deals with the specific consonant confusions made by normal hearing listeners under low-pass and high-pass filtering. The consonant confusion analyses provided information on how filtering disrupted perception of particular speech sounds by normal listeners and how their performance related to that of listeners with impaired hearing.

A classic study by Miller and Nicely (1955) provided an analysis of specific consonant confusions made by normal-hearing listeners. The authors summarized patterns of confusions according to voicing, nasality, duration, affrication and place of articulation and calculated relative information transmitted by each feature under several filtering conditions and in masking noise. Place of articulation was most affected while voicing and nasality were least affected by low-pass filtering. All five features studied were disrupted in much the same way by high-pass filtering. Much of the acoustic power of some of the consonants was removed when the low frequencies were removed; parts of the acoustic signal was inaudible, making errors much less predictable than by low-pass filtering. The authors hypothesized that the presence of random noise also contributed to the difficulties seen in the high-pass filter condition. While the noise added to the high-frequency reduction imposed by low-pass filtering, for the high-pass condition it produced a narrow band-pass system.

Dubno, Dirks, and Ellison (1989), looked at the contribution of particular frequency regions to the perception of stop consonants differing from each other in terms of place of articulation. Normal hearing and hearing-impaired listeners heard both synthetic and naturally produced (by a male talker) consonant-vowel syllables where /b,d,g/ were paired with /ɑ,i,u/. The consonant-vowel syllables were presented in several low-pass and high-pass filter conditions as well as fullband conditions at several intensities. For the low-pass conditions the filtering affected the perception of these three consonants differently. Once the low-pass filter cut-off frequency was 1400 Hz, identification scores for decreased for /d/ but not for /g/ and /b/ until additional low frequency energy was removed. For the high-pass conditions, /b/ and /g/ were most affected by the removal of high frequency information, i.e., as the cut-off frequency increased. These results indicate that although consonants have the same place of articulation, they may not be affected equally by both types of filtering. A point to keep in mind is the fact that a portion of the stimuli used in this study were synthetic, which may yield different intelligibility than natural speech. While there were stimuli produced naturally, only a male talker was used, which may affect test results (see discussion above).

Wang, et. al (1978), compared the effects of filtering and sensorineural hearing loss on consonant confusion patterns. Normal-hearing subjects listened to vowel-consonant and consonant-vowel nonsense syllables under several low-pass and high-pass filter conditions. They compared their results to listeners with sensorineural hearing loss from a previous study (Bilger and Wang, 1976). The cross-over frequency of the importance function was between 1750-2000 Hz. Also, low-pass filtering caused a greater degradation in intelligibility than high-pass filtering. Sibilance was disrupted in the low-pass filter condition when the cut-off frequency dropped from 2800 to 1400 Hz. Frication and duration were also less accurately perceived under the low-pass filter conditions. In general, high-pass filtering produced less consistent changes in feature recognition than low-pass filtering, a result also seen in Miller and Nicely (1955). The authors concluded that the scores for the low-pass filter condition with a 1400 Hz cut-off were similar to those seen for listeners with a high-frequency sensorineural hearing loss and the scores seen for the high-pass filter condition with a 2800 Hz cut-off, while less predictable than those for the low-pass condition, were similar to those for listeners with a flat sensorineural hearing loss.

Sher and Owens (1974) looked at consonant confusions for meaningful consonant-vowel-consonant words naturally produced by a male talker. Eight voiceless consonants and six voiced consonants were used in either the initial or final position of the words, which were presented in a four-alternative closed-set response task. One group of subjects had normal hearing and the other had a high-frequency sensorineural hearing loss. The normal hearing subjects heard the words through a low-pass filter with a 2040 Hz cut-off frequency. Percent correct score was computed for each listener. Additionally, the performance of this group was compared to a group of listeners with sensorineural hearing loss. The results were that the number of phoneme errors was similar for the two groups and that the error and substitution probabilities were also similar. The authors concluded that the results obtained using normal-hearing listeners' responses to filtered speech could be generalized to the hearing-impaired population.

A similar study by Owens, Talbott, and Schubert (1972), also compared consonant errors of listeners with sensorineural hearing loss to those of normal-hearing subjects listening to low-pass filtered (cut-off Hz 780 Hz) consonant-vowel-consonant words. The results were comparable to those found by Sher and Owens (1974). Performance of normal-hearing listeners under low-pass filters with cut-off frequencies below 1000 Hz was similar to that of those with high frequency sensorineural hearing loss. The authors expressed some surprise at this result. Their expectation was that the performance of the sensorineural hearing loss subjects, who were "accustomed to" listening to a degraded auditory signal, would be better than listeners with normal hearing listening to filtered speech.

Finally, it should be noted that there have been methods specifically designed to describe speech intelligibility data. In the 1920s, Bell Laboratories developed a method to predict the intelligibility of a speech signal. Described by French and Steinberg (1947), the Articulation Index predicts the intelligibility of a speech signal by dividing up the signal into a series of bands that contribute equally to the overall intelligibility. The use of the Articulation Index provides a method of mathematically treating speech intelligibility data to determine the additive contribution of each of the bands in the signal by using the log of error probability.

French and Steinberg (1947) based their derivation of the Articulation Index on results of articulation tests that determined the effects of high-pass and low-pass filtering on speech perception in normal listeners. The speech signal was divided into 20 frequency bands, each contributing a maximum of 0.05 to the Articulation Index. Subjects in this study heard nonsense consonant-vowel-consonant syllables through both low-pass and high-pass filters and at various intensity levels. The cross-over frequency for the importance function was 1900 Hz. It is important to note that the Articulation Index for this study is valid only for these materials with the same talkers used in this study. It should also be noted also that the 20-bands were derived using both male and female talkers.

Later investigations of the work originally done by French and Steinberg (1947) have both validated it and added to it. Kryter (1962a) proposed two methods for calculating Articulation Index. The first method is the 20-band method as reported in French and Steinberg (1947). Kryter's method, however, assumed that the dynamic range of speech was 36 dB (as opposed to French and Steinberg's 30 dB) and used linear intensity weighting to describe the contribution of an intensity unit within the dynamic range. The second method is the 1/3 octave band method and the full-octave band method derived from the 20-band method. Using hypothetical sample data, the author deduced that the 20-band, 1/3 octave-band and full-octave-band methods yielded results that were in agreement with each other. Consequently, the 1/3 octave-band and full-octave band methods described by Kryter were adopted as national standards.

Kryter (1962b) also conducted a study validating the Articulation Index as a predictor of speech intelligibility using his modified calculation methods described above. He used nonsense consonant-vowel-consonants (similar to French and Steinberg 1947), under conditions of low-pass, high-pass, multiple passband, and single passband filters at several cut-off frequencies in quiet and noise. He then calculated Articulation Index in accordance with the modified 20-band method. A comparison of the intelligibility predicted by Articulation Index and the actual performance of the listeners under all conditions indicated that the performance scores fell very close to those predicted by Articulation Index. Thus, the validity of Articulation Index as a predictor of speech intelligibility under low-pass, high-pass, and single passband filtering and in noise was demonstrated. Additionally, a comparison of Articulation Index calculated by Kryter's modified 20 band method and the original French and Steinberg (1947) 20-band method showed almost identical results for the two data sets. Interestingly, the results of this study did not generalize to the multiple passband system; performance scores did not fall close to the values predicted by Articulation Index. This finding questions the assumption that each of the bands in the speech signal always contributes equally to intelligibility. The author also cautions that the relationship between Articulation Index and intelligibility must be interpreted carefully, keeping in mind that the relationship only holds when test materials and abilities of talkers and listeners are held constant.

Pavlovic and Studebaker (1984) also looked at Articulation Index predictions by testing some of the assumptions underlying Articulation Index; namely: a 30 dB dynamic range, linear intensity weighting, exclusion of pauses in speech in calculating its level, and use of actual peak values. These assumptions lead to development of different

Articulation Index calculation methods (see Kryter, 1962a, above). In order to test these assumptions, Articulation Index predictions were tested using nonsense syllables, produced by 1 male and 1 female talker, in thirteen different listening conditions of filtered speech and noise. When the importance function found by French and Steinberg (1947) was used to compare the predicted and observed scores, it was seen again that the actual scores agreed fairly well with the predicted scores. There were however, differences in the function produced as a function of talker, which lead to the use of results from only one talker. Therefore, these results support importance functions derived in earlier studies as well as the idea that these functions are applicable only to a particular talker, listener, and stimulus.

While the studies described in this section found importance functions for normally hearing listeners, a study by Boothroyd (1967) looked at importance functions in hearing-impaired children. Boothroyd found that the importance function shifted toward the lower frequencies with increasing hearing loss and increasing audiogram slope.

III. Summary

In summary, it is clear that there are acoustic differences among talkers - both within and between gender and age groups. It is reasonable to assume that these acoustic differences will affect the distribution of useful information across the frequency spectrum, but few data are available to indicate the magnitude of such effects. The present study was designed primarily to provide data on individual differences of cross-over frequency among a small group of adult male and female talkers.

CHAPTER THREE

METHODS

I. Subjects

There were two groups of subjects in this study, talkers and listeners.

a. Talkers

Talkers were 3 adult males and 3 adult female native talkers of American English with no known disorders of hearing, speech, or language. Ages ranged from 27 to 63 years old with a mean of 38.5 years. Individual talker characteristics are shown in Table 3.1, below.

Table 3.1 Talker characteristics

TALKER	AGE	REGION OF ORIGIN	FEATURES
Female 1	27	Rhode Island	None
Female 2	35	New Jersey	Large neck musculature
Female 3	32	Hawaii	5' tall; slight lateral lisp
Male 1	27	Massachusetts	None
Male 2	46	New Jersey	None
Male 3	63	New York	None

b. Listeners

Listeners were 8 native talkers of English. Six listeners were from the New York area and two were natives of Massachusetts. Ages ranged from 18-40 years with a mean of 29.8 years. Hearing sensitivity was determined to be within normal limits in both ears - defined by pure tone audiometric thresholds no greater than 20 dBHL at octave intervals from 250-4000 Hz. The hearing criterion was established by screening immediately before testing. The listeners were informed of the results of the screening.

II. Stimuli

1. Test Material

Test material was twenty AB isophonemic word lists, each consisting of ten consonant-vowel-consonant words. Lists 1-15 were originally developed by Arthur Boothroyd (1968), who also modified the lists for American English pronunciation (1984). Lists 16-20 were developed by Balaji Oruganti in 1996. Each list contains the same 10 vowels and 20 consonants that occur most frequently in English. Each consonant-vowel-consonant word appears only once in the set of 200 words. See Appendix A for a complete listing of the words.

2. Recording of Test Material

Each talker recorded twenty AB isophonemic word lists. Talkers were seated inside the sound isolated booth at a distance of 18" from the microphone of a sound level meter. A foam ball fitted over the microphone of the sound level meter served as a windscreen. The words were recorded using a Brüel & Kjær sound level meter model 2235 with the 1/2" sound field condenser microphone, set to the "C" scale in the 20-90 dB range. The signal-to-noise ratio of the recording environment was approximately +52 dB. AC output of the sound level meter was recorded onto a Panasonic RT-R124MP Digital Audio Tape via the Panasonic SV255 Portable Digital Audio Tape recorder, located outside the sound isolated booth. Recording sessions were monitored under headphones by the examiner. During the recording session, the talker monitored recording levels using a Leader LMV-181A AC Millivoltmeter.

3. Editing of Recorded Materials

Each recorded word was digitized using 16 bit resolution and a sampling rate of 22050 Hz. Each word file was truncated at onset and offset (determined visually from the waveform and confirmed by listening). One hundred milliseconds of silence was then added at the beginning and end. The 100 millisecond buffers served to avoid onset and offset artifacts when processing.

Two additional editing tasks were necessary. The first involved removal of structure-borne infrasound that is present in the CUNY Graduate Center. The second involved an adjustment to the overall level of each list so that the long-term rms level of a list was constant across lists and talkers. Note that the second adjustment was intended to preserve natural variations of level across segments and words but to correct for any tendency of talkers to change effort or microphone distance during the recording session.

These changes were accomplished using DaDisp (version 4.0) software from DSP corporation. Using a worksheet and program designed for the purpose (see Appendix B), the ten word files from each list were imported. Each file was then high-pass filtered using an elliptic filter with a cut-off frequency of 150 Hz and a slope of 50 dB per octave. The 100 millisecond buffer was removed from each filtered word file and the ten words were concatenated to create a single file. The long-term rms level, rms amplitude distribution, and long-term average spectrum were determined from this file. The measured long-term rms level (in dB re 1 digital unit) was subtracted from 73. The difference, in dB, was then added to each of the 10 word files so as to bring the long-term rms level of each list (excluding silences) to 73 dB re 1 digital unit. The edited word files were then saved with new names. The edited, filtered, and equalized word files were used as input to the high- and low-pass filtering algorithms described next.

4. High- and Low-pass Filtering

Filtering was also accomplished in DaDisp. High- and low-pass elliptic filters were created with slopes of 100 dB per octave and cut-off frequencies at half-octave intervals ranging from 707 to 5656 Hz. The responses of these filters are shown in Figure 3.1. The complete DaDisp worksheet is included in Appendix C. Based on the results of a pilot study, only 3 cut-off frequencies were used in the present study, namely, 707, 1414, and 2828 Hz.

III. Stimulus Presentation

Stimuli were delivered via a Gateway 2000 P5-166 XL computer using custom written software. The output from the computer went directly into a custom-made Stimulus Presentation System box to control the presentation level in 2 dB steps. The output from the box was sent to TDH-50 headphones. The listeners heard the words binaurally under headphones in a sound-attenuating single wall AIC test booth. The listeners heard each word once, with the possibility of one replay in cases of inattention or distraction. The listeners repeated the word into a microphone located inside the test booth and wrote their responses in the appropriate space on a score sheet.

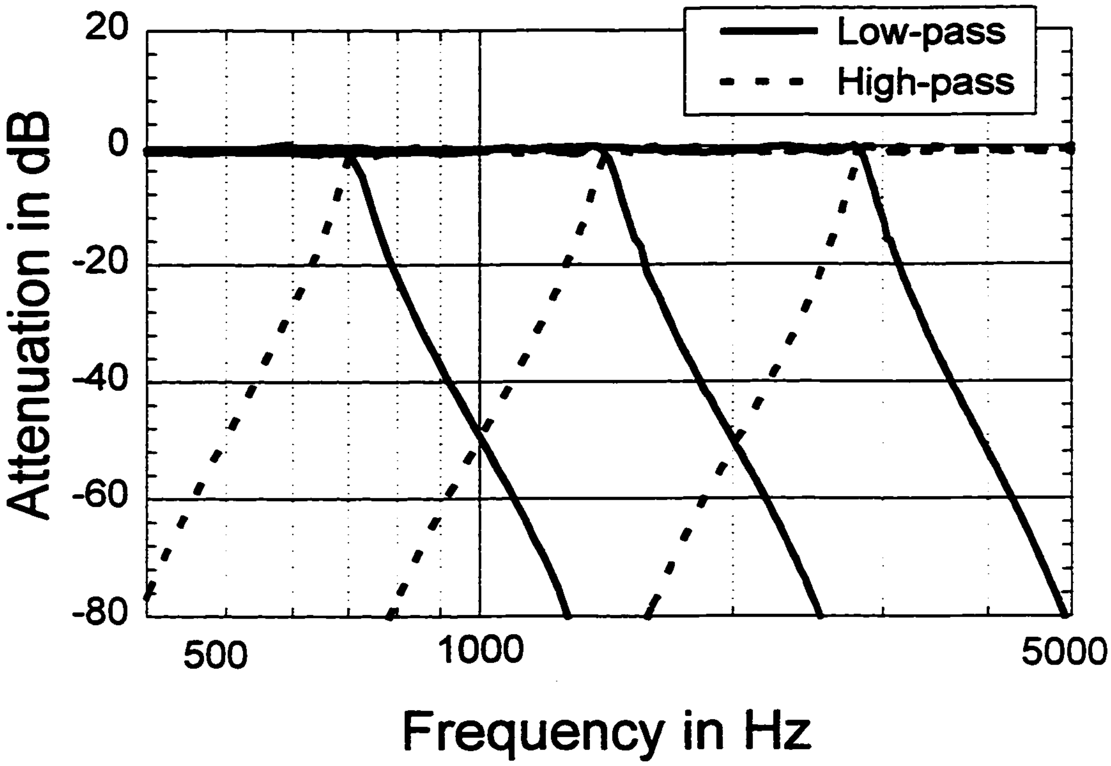


Figure 3.1. Filter transfer functions for high- and low-pass filters with cut-off frequencies of 707, 1414, and 2828 Hz.

IV. Calibration

The presentation level for testing was adjusted so that the peak rms levels of the vowels in the unfiltered stimuli averaged around 70 dBSPL in a 6 cc coupler. No further adjustments were made when presenting filtered stimuli, even though the peak vowel levels were often lower than for the unfiltered speech. The goal was to present the filtered stimuli at the same level as they would have been presented as part of the full-band stimulus.

V. Test Conditions

The listeners heard two lists for each of the six talkers under six filter conditions, for a total of 72 word lists. The following were randomized: the order of the filtering conditions, the order of the talkers, the order of presentation of the AB lists (see Appendix D for the test randomizations), and the order of the words within the lists (randomized by the software program during presentation).

VI. Scoring

Oral responses were scored on-line by the examiner. Written responses were scored off-line after conclusion of testing. Due to the possibility of misunderstanding during the on-line scoring, written responses took priority over oral responses when there was a discrepancy between the responses. Each of the three phonemes in the AB words was scored individually. Omissions, additions, or substitutions of phonemes were scored as errors.

For on-line scoring, a correct response for a phoneme was recorded by entering the number corresponding to the position of the phoneme, i.e., 1, 2, or 3. An "All Correct" option was also available in the program in the case of a whole word correct response. The responses were recorded and saved automatically into a Microsoft Excel, version 7.0 worksheet which calculated a percent-phonemes-recognized score. This worksheet was corrected, if necessary, based on the listeners' written responses[2].

VII. Test Procedure

Each listener passed a hearing screening before testing. The listener was then seated in the sound-attenuating test booth and instructed how to respond during the testing. Supra-aural headphones were placed on the listener and testing began.

The listener heard, repeated, and wrote the filtered words from the AB word lists. A replay of a word was provided upon request in cases of inattention or distraction. The listener heard one word list in each of the six filtering conditions for each of the six talkers. At the completion of testing of six filtering conditions for three talkers (the halfway point), the listener was given a break. A break was also provided upon request to a listener at any point in the testing. Testing was completed in two sessions. One session consisted of testing all six filtering conditions for all six talkers.

Each listener heard 72 word lists (6 talkers x 6 filter conditions x 2 replications). There are, however, only 20 lists in the AB materials. During the course of the study, a listener heard a given word list approximately three times - albeit in randomized

---

[2] The written response took priority over the verbal response only when it was phonetically unambiguous and in disagreement with the on-line score.

order. There was, therefore, the possibility of a learning effect resulting from increasing exposure to the test words. This effect might be expected to be greater in instances where a list heard under difficult conditions had previously been heard under an easier condition. Randomization of filter conditions and talkers within a session was intended to avoid confounding learning effects with the effects of talker and filtering condition. The two replications provided an opportunity to test for the magnitude of any learning effect.

VIII. Acoustical Analyses

Details of the acoustic analyses will be given in Chapter 5.

IX. Summary

Stimuli were high- and low-pass filtered with cut-off frequencies of 707, 1414, and 2828 Hz. Eight normally hearing listeners heard one word list (ten words) spoken by each of the six talkers under six filtering conditions. For each word they both wrote and said what they heard. The process was repeated a few days later. The percentage of phonemes recognized was calculated for each listener, each talker, each filtering condition, and each replication. Acoustic analyses were also performed on a sample of words for each talker.

CHAPTER FOUR

RESULTS OF PERCEPTUAL STUDY

I. Raw data

Appendix E contains, for each listener, the phoneme recognition scores for each talker x filter condition x replication. Appendix E1 shows the percent phonemes recognized for six talkers under three high-pass and three low-pass filter conditions for eight listeners for the first of two replications. Appendix E2 shows the percent phonemes recognized for six talkers under three high-pass and three low-pass filter conditions for eight listeners for the second replication. In each of these appendices, the data for the female talkers for the six filter conditions are shown on the top table and the data for the three male talkers for the six filter conditions are shown on the bottom table.

II. Repeated-measures analysis of variance

The percent phoneme recognition scores were arcsine-transformed to increase homogeneity of variance and subjected to a single-group three-way repeated-measures analysis of variance. The three factors were: replication (two levels), talker (six levels), and filter condition (six levels). Note that in this analysis, listener is a random grouping variable and conclusions of significance apply to the means of the listener population represented by these eight listeners. At this stage, talker is treated as a fixed effect. Any conclusions of significance, therefore, apply only to these six talkers.

The analysis is shown in Table 4.1. It will be seen that there were highly significant effects of replication, talker, and filter condition and a significant talker x filter condition interaction. It is important to note, that although replication was significant, it did not interact with the other two effects. In other words, in the arcsine domain, any learning effect between replication 1 and 2 can be assumed to apply uniformly across talkers and filter conditions.

Table 4.1. Repeated-measures analysis of variance in arcsine-transformed phoneme recognition scores. Shading indicates significance at at least the 5% level.

Source of variance	Sum of Squares	Degrees of freedom	Estimated mean square	Error term	F ratio	p level
Listener	3060.7	7	437.2	--		
Replication	**1220.7**	**1**	**1220.7**	**RxL**	**22.7**	**0.002**
Talker	**1827.6**	**5**	**365.5**	**TxL**	**4.6**	**0.003**
Filter	**166781.8**	**5**	**33356.4**	**FxL**	**276.9**	**0.000**
RxT	211.0	5	42.2	RxTxL	0.8	0.536
RxF	307.5	5	61.5	RxFxL	1.4	0.250
TxF	**4608.3**	**25**	**184.3**	**TxFxL**	**3.9**	**0.000**
RxTxF	1172.1	25	46.9	RxTxFxL	0.9	0.593
RxL	376.6	7	53.8			
TxL	2810.1	35	80.3			
FxL	4215.8	35	120.5			
RxTxL	1776.2	35	50.7			
RxFxL	1542.7	35	44.1			
TxFxL	8226.5	175	47.0			
RxTxFxL	9028.4	175	51.6			
Total	207166	575				

When the analysis was repeated with percent correct scores, the only difference was a significant ($p< 0.05$) interaction between replication and filter condition. This can be explained by the fact that the scores for the filter conditions high-pass 707 Hertz and low-pass 2828 Hertz were close to 100% and provided little room for increase in the percent domain.

The magnitude of the learning effect, collapsed across talker and filter condition, was quite small. Mean score rose from 77.7% at replication 1 to 80.4% at replication 2, a difference of only 2.7 percentage points, or 1 phoneme in a list of 10 consonant-vowel-consonant words.

Differences among talkers were somewhat larger, ranging from 76.7% to 82.3%, a range of 5.6 percentage points. However, in post-hoc testing using the least-significant difference test, female talkers TF2 and TF3 and all three male talkers were not found to differ significantly. Female talker TF1 gave a significantly higher score than the other five talkers. Even the difference between least and most intelligible of these six talkers was not very great, amounting to about two phonemes in a list of ten consonant-vowel-consonant words.

There were, of course, large differences in scores across filtering conditions, which will be described later. At present, the main interest is in the interaction between talker and filter condition. Figure 4.1 shows phoneme intelligibility as a function of filter cut-off frequency for all six talkers. Under certain filter conditions the range of scores is 12-15 percentage points with a tendency for higher scores for the female talkers under the high-pass filtering conditions and higher scores for the male talkers under the low-pass filtering conditions - this issue will be explored further below.

III. Mean phoneme intelligibility

Table 4.2 shows phoneme intelligibility for each talker, collapsed across listeners, as a function of filtering condition and replication. Note that the phoneme intelligibility is obtained by combining intelligibility scores for initial consonant, vowel, and final consonant and collapsing across the eight listeners. Means are also shown collapsed across replication and gender groups.

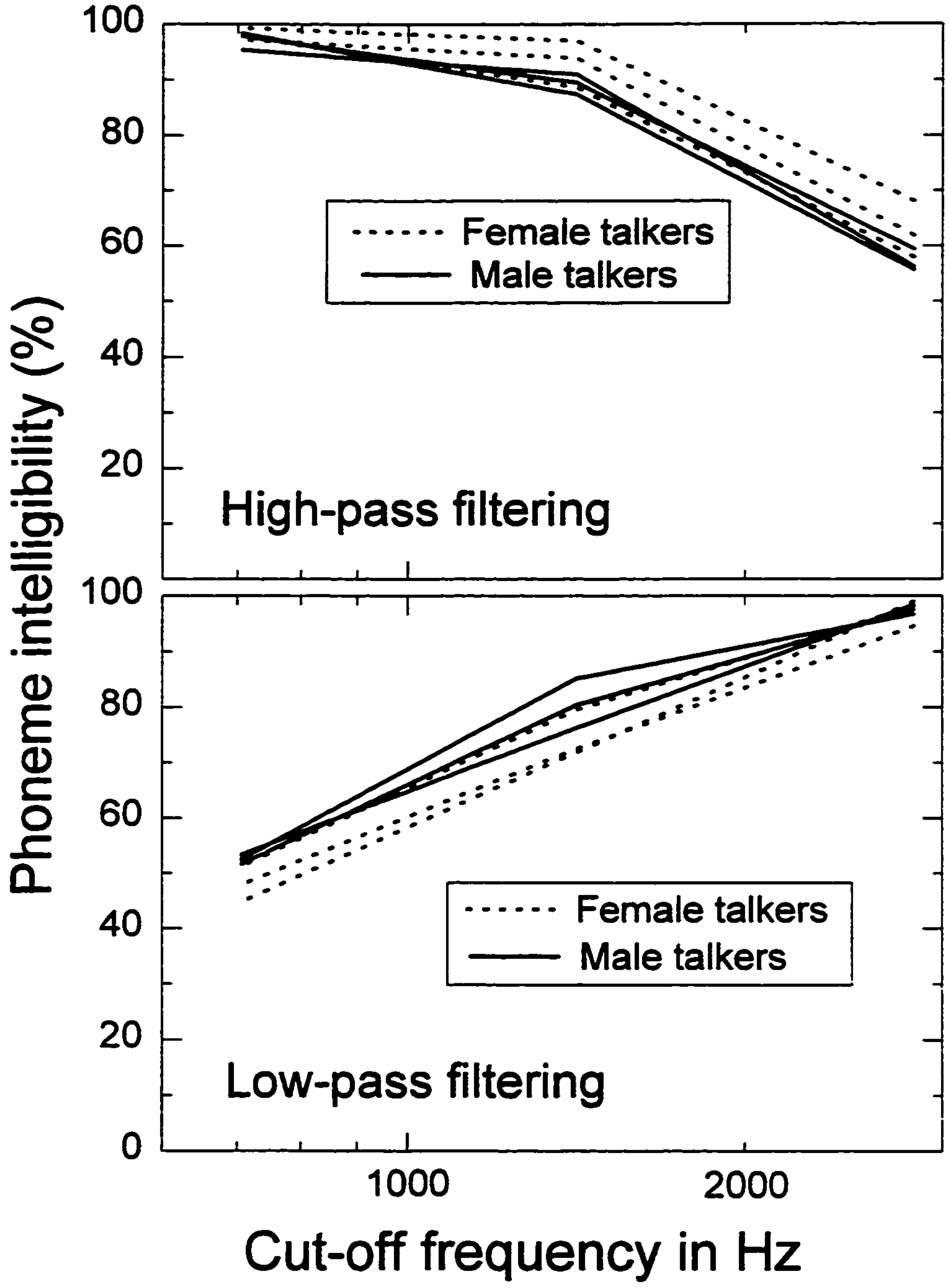


Figure 4.1. Phoneme intelligibility as a function of filter cut-off frequency in Hertz for 6 talkers. The solid lines denote the 3 male talkers and the broken lines denote the 3 female talkers.

Table 4.2 % correct scores, logarithmic cross-over frequencies, and % correct at cross-over for 3 high-pass and 3 low-pass filter conditions for 6 talkers.

Replication 1							xover in kHz	% correct at xover
Talker	HP2828	HP1414	HP707	LP2828	LP1414	LP707		
TF1	63.33	96.67	98.75	97.08	78.33	49.59	1.805	84.93
TF2	55.00	84.58	97.92	99.58	72.08	42.92	1.646	78.11
TF3	58.75	93.33	97.50	96.25	67.08	51.25	1.881	79.09
mean	59.03	91.53	98.06	97.64	72.50	47.92	1.777	80.80
SD	4.17	6.24	0.64	1.74	5.64	4.41		
SE	2.41	3.60	0.37	1.00	3.25	2.55		
TM1	52.92	87.50	97.92	97.50	77.92	49.58	1.599	81.38
TM2	52.92	88.34	96.67	97.92	75.83	56.25	1.644	80.63
TM3	55.83	89.17	93.75	96.25	81.26	48.33	1.584	83.71
mean	53.89	88.33	96.11	97.22	78.33	51.39	1.609	81.88
SD	1.68	0.83	2.14	0.87	2.74	4.26		
SE	0.97	0.48	1.23	0.50	1.58	2.46		

Replication 2							xover in kHz	% correct at xover
Talker	HP2828	HP1414	HP707	LP2828	LP1414	LP707		
TF1	72.92	97.50	100.00	99.17	80.83	52.91	1.851	87.95
TF2	60.83	92.92	97.92	98.33	71.67	46.66	1.817	81.31
TF3	65.00	94.58	97.08	92.91	77.92	44.58	1.832	83.52
mean	66.25	95.00	98.33	96.80	76.81	48.05	1.833	84.27
SD	6.14	2.32	1.50	3.40	4.68	4.33		
SE	3.54	1.34	0.87	1.96	2.70	2.50		
TM1	58.33	87.50	98.75	97.50	82.92	53.75	1.520	84.44
TM2	65.83	90.83	99.17	98.75	76.67	50.42	1.742	83.31
TM3	56.67	92.92	97.08	97.08	89.17	56.67	1.500	89.84
mean	60.28	90.42	98.33	97.78	82.92	53.61	1.587	85.39
SD	4.88	2.73	1.10	0.87	6.25	3.13		
SE	2.82	1.58	0.64	0.50	3.61	1.81		

Replication 1 and 2							xover in kHz	% correct at xover
Talker	HP2828	HP1414	HP707	LP2828	LP1414	LP707		
TF1	68.1	97.1	99.4	98.1	79.6	51.2	1.825	86.41
TF2	57.9	88.7	97.9	99.0	71.9	44.8	1.730	79.77
TF3	61.9	94.0	97.3	94.6	72.5	47.9	1.861	81.25
mean	62.6	93.3	98.2	97.2	74.7	48.0	1.806	82.55
SD	5.15	4.21	1.07	2.33	4.28	3.23		
SE	2.97	2.43	0.62	1.34	2.47	1.87		
TM1	55.6	87.5	98.3	97.5	80.4	51.7	1.563	82.89
TM2	59.4	89.6	97.9	98.3	76.1	53.3	1.690	81.80
TM3	56.2	91.0	95.4	96.7	85.2	52.5	1.543	86.65
mean	57.1	89.4	97.2	97.5	80.6	52.5	1.599	83.59
SD	2.0	1.8	1.6	0.8	4.6	0.8		
SE	1.2	1.0	0.9	0.5	2.6	0.5		

Also shown in Table 4.2 are cross-over frequency and full-band intelligibility. These last two quantities are derived from the filtered phoneme intelligibility data. Their derivation will be explained later.

IV. High- and low-pass filter functions

Figure 4.2 shows phoneme intelligibility, collapsed across two replications, as a function of filter cut-off frequency for each of the six talkers. Similar graphs were prepared for replications 1 and 2 separately, but there was no evidence of a significant effect of replication on cross-over frequency. For that reason, only collapsed data are shown here. It will be seen from Figure 4.2 that all cross-over frequencies fell between 1414 and 2828 Hz. Results for these 4 filter conditions were, therefore, used for estimation of cross-over frequency by interpolation.

V. Cross-over Frequencies

It will be noted that a logarithmic scale is used for frequency in Figure 4.2. The values of cross-over frequency shown in this figure and in Table 4.2 are, therefore, logarithmic cross-over frequencies. The formula used for calculating cross-over frequency is derived in Appendix F.

The mean value of cross-over frequency, collapsed across 6 talkers, was 1702 Hz with a standard deviation of 130 Hz and a standard error of 60 Hz. In a simple analysis of variance, however, there was strong evidence of a main effect of gender. The results of this analysis are shown in Table 4.3.

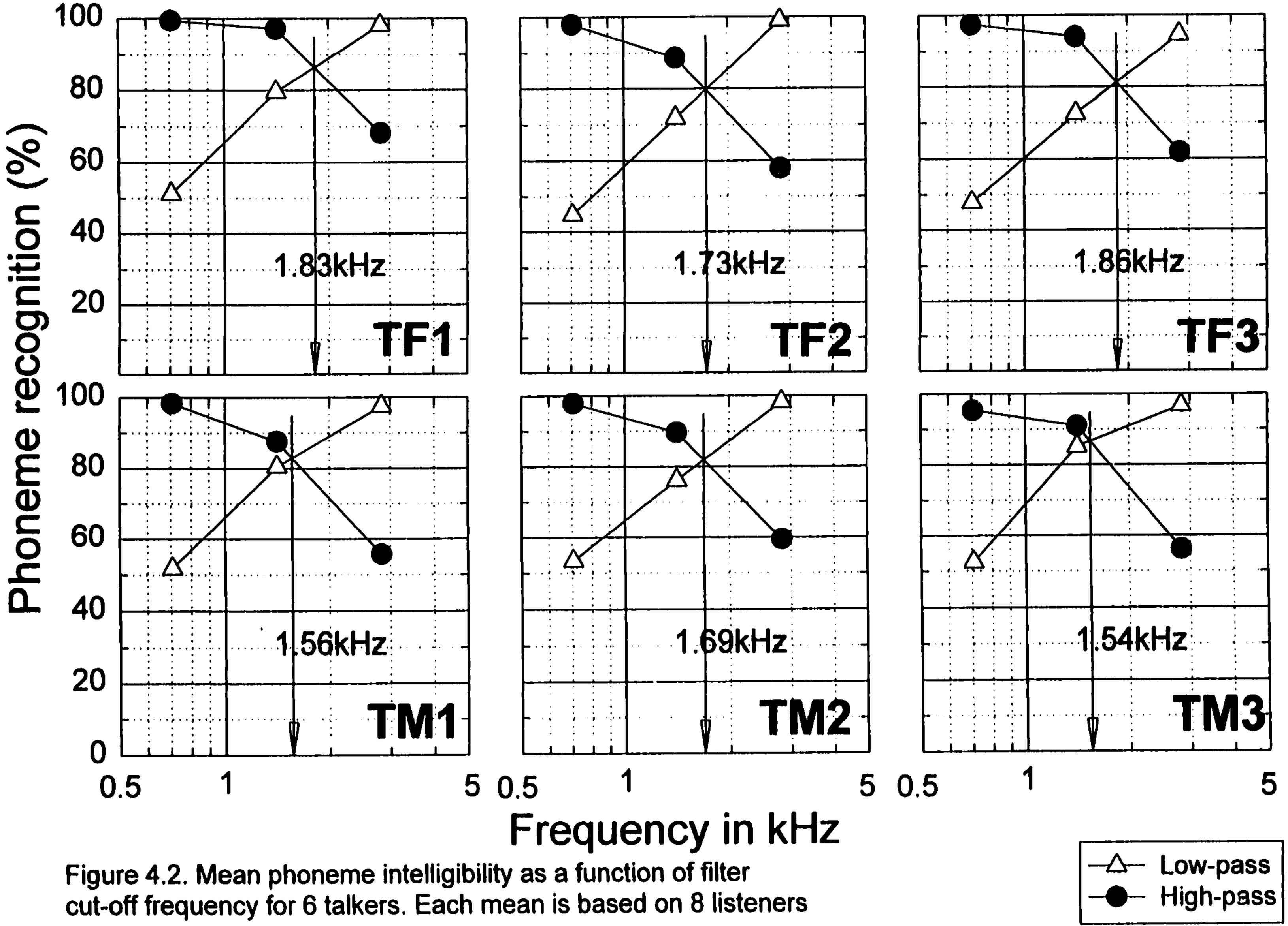


Figure 4.2. Mean phoneme intelligibility as a function of filter cut-off frequency for 6 talkers. Each mean is based on 8 listeners and 2 replications.

Table 4.3 shows that the within-gender variance is 5385 Hz^2, giving a within-gender standard deviation of 73 Hz (sqrt(5385)) and a standard error for each gender mean of 42 Hz (73/sqrt(3)). The mean cross-over frequencies were 1806 and 1599 Hz, for females and males, respectively. Thus, the mean cross-over frequency for the women was some 200 Hz higher than for the men. The difference amounts to 13% or 1/6 octave.

Table 4.3. Analysis of variance of the cross-over frequencies in kHz for 6 talkers

Source of Variance	Sum of Squares	Degrees of Freedom	Estimated Mean Square	F ratio	p level
Gender	62628	1	62628	11.6	0.027
Talker Within	21541	4	5385		
Total	84170	5			

Assuming that cross-over frequency is normally distributed within the two gender populations represented by these talkers, the following conclusions can be drawn with 95% confidence:

i) Mean cross-over frequency averaged across all women lies between 1689 and 1923 Hertz.

ii) Mean cross-over frequency averaged across all men lies between 1482 and 1716 Hertz.

Note that these confidence limits were obtained by multiplying the appropriate standard deviation or standard error by 2.78 which is the value of student's 't' for 4 degrees of freedom.

Figure 4.3 provides an illustration of individual cross-over frequencies together with the 95% confidence limits for the gender means.

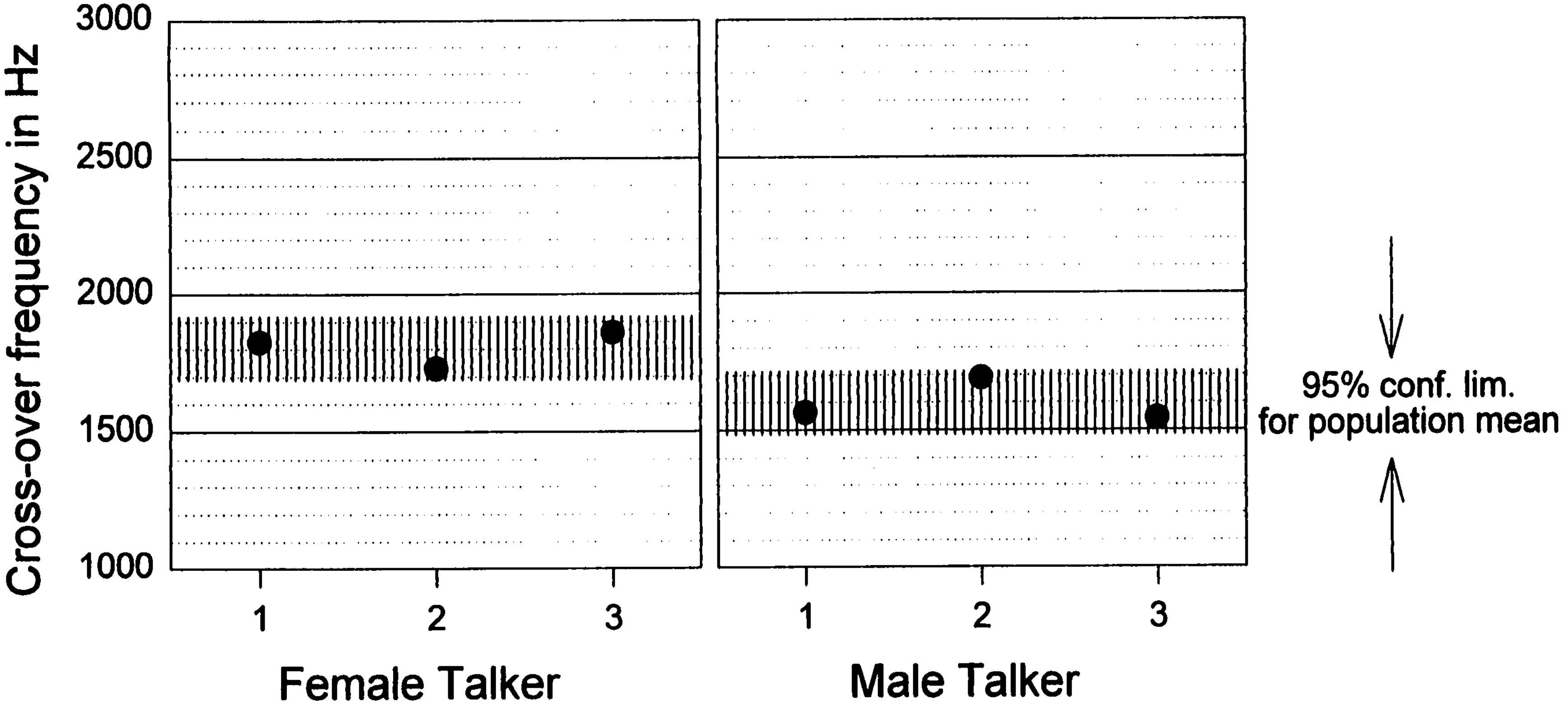


Figure 4.3 Cross-over frequency in Hz as a function of talker collapsed across 8 listeners and 2 replications. The shading around the data points indicates the 95% confidence limits for the population mean.

VI. Half-band phoneme intelligibility

Figures 4.4 and 4.5 show phoneme intelligibility as a function of cut-off frequency for replications 1 and 2, respectively. Also shown are the values of phoneme intelligibility obtained, by interpolation, at the cross-over frequency. In a repeated-measures analysis of variance, shown in Table 4.4, there was strong evidence of a replication effect - hence the presentation of both sets of data. There was, however, no main effect of gender, nor was there an interaction between gender and replication - as will be seen from Table 4.4. The group means (6 talkers) for replications 1 and 2 were 81.3% and 84.8%, respectively. Thus, measured half-band phoneme intelligibility increased by 3.5 percentage points between two replications.

It will be seen from Table 4.4 that the estimated within-gender and within-replication variance was 1.04%^{2} giving a standard deviation of 1.02 percentage points. The corresponding standard error for the mean of a single talker (based on 2 replications) is 1.02/((sqrt(2)) (= 0.72 percentage points). The standard error for the mean of 6 talkers (based on 2 replications) is 1.02/((sqrt(12)) (= 0.29 percentage points).

Table 4.4 Analysis of variance of the half-band intelligibility data

Source of Variance	Sum of Squares	Degrees of Freedom	Estimated Mean Square	Error Term	F ratio	p level
Gender	5.8832	1	5.8832	SW	0.31	0.606
Replication	**42.2903**	**1**	**42.2903**	**RxSW**	**40.64**	**0.003**
GxR	0.1228	1	0.1228	RxSW	0.12	0.748
Subject Within	75.4325	4	18.8581	—		
RxSW	4.1621	4	1.0405	—		
Total	127.8910	11				

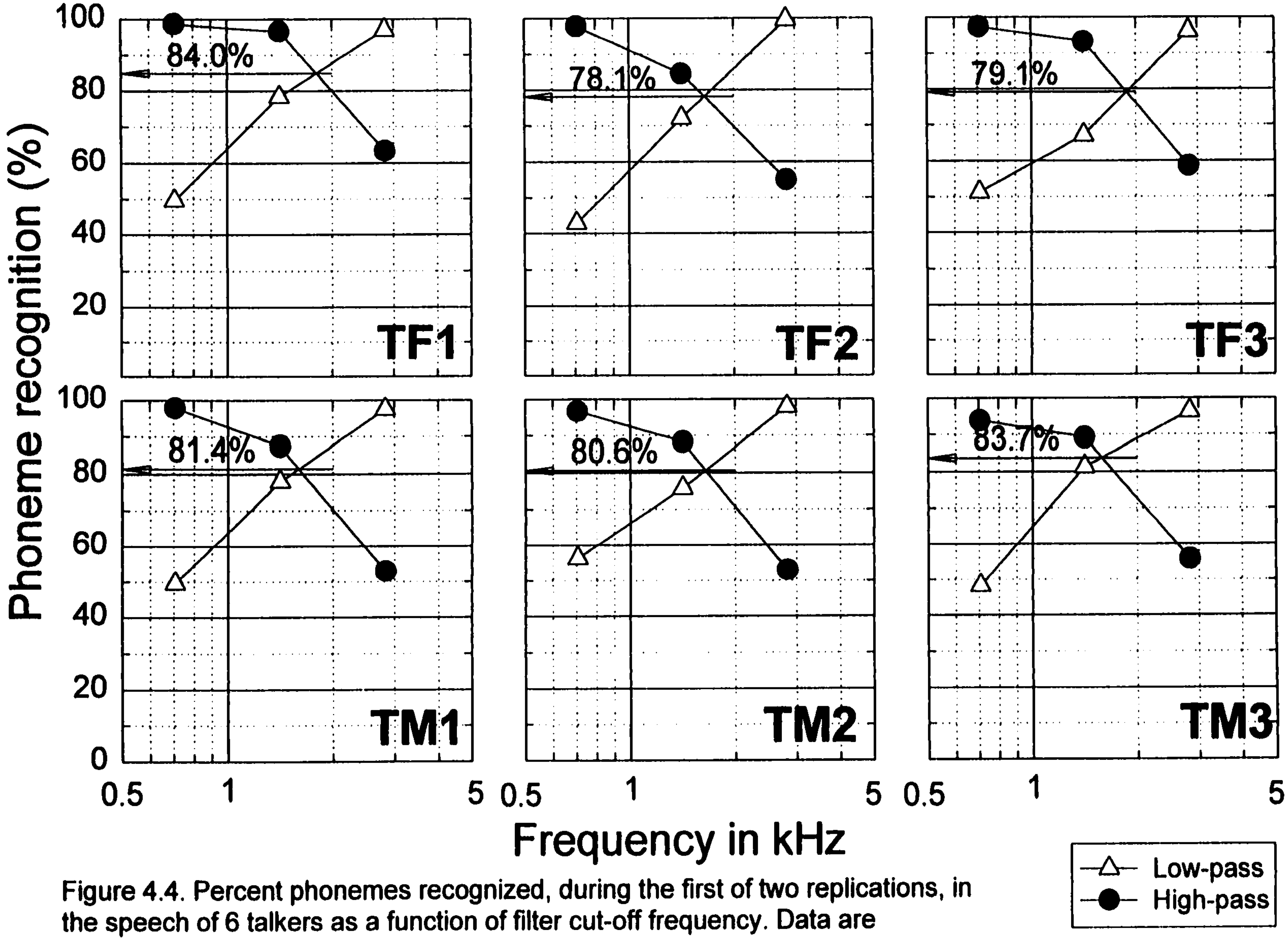


Figure 4.4. Percent phonemes recognized, during the first of two replications, in the speech of 6 talkers as a function of filter cut-off frequency. Data are collapsed across 8 listeners. Arrows show interpolated "cross-over" scores.

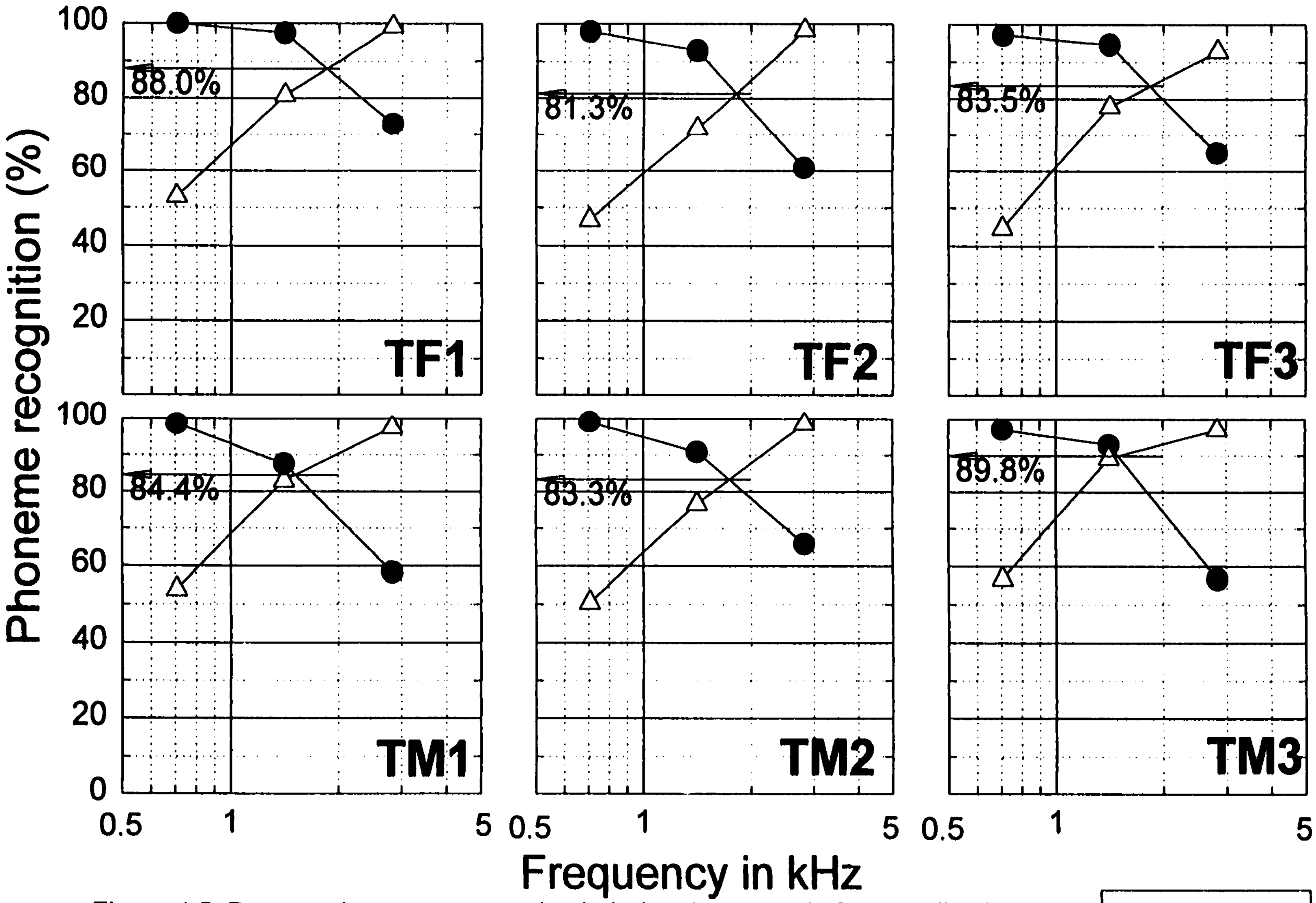


Figure 4.5. Percent phonemes recognized, during the second of two replications, in the speech of 6 talkers as a function of filter cut-off frequency. Data are collapsed across 8 listeners. Arrows show interpolated "cross-over" scores.

When the data are collapsed across replication, the mean half-band intelligibility of the six talkers is 83.1 percent with a standard deviation of 2.8 percentage points and a standard error of 1.2 percentage points. Using these values and 2.57, which is the value of student's 't' for 5 degrees of freedom, it can be concluded that:

i) Population mean half-band intelligibility averaged across all adult talkers lies between 80.2% and 86.2%.

Table 4.5 (below) shows the population and individual means for the half-band and full-band phoneme intelligibility scores.

Table 4.5. Population and individual means for half-band and full-band intelligibility scores:

	Phoneme Intelligibility (%)		
		1/2-band	full-band
95% confidence limits	mean	83.1	97.1
for population mean	upper	86.1	98.1
	lower	80.2	96.1
for individual mean	upper	90.4	99.1
	lower	75.9	94.2

VII. Full-band intelligibility

Full-band intelligibility can be predicted from half-band intelligibility using probability theory (Boothroyd and Nittrouer, 1988), which is also the basis of Articulation Index theory (French and Steinberg, 1947).

The equation is:

$$f = 100*(1-(1-h/100)^2) \quad \text{..} \quad (1)$$

where f = full-band intelligibility in %, and

h = half-band intelligibility in %

The underlying assumption is that the two half-bands are independent - that is, failure to recognize via one band has no effect on probability of recognition via the other.

Applying this equation to the half-band data just reported it can be concluded that:

i) Population mean full-band intelligibility averaged across all adult talkers lies between 96.1% and 98.1%.

Note that data from the two replications are collapsed to calculate the full-band intelligibility score. These data are summarized in Table 4.5 above.

VIII. Gender effects for filtered speech

The phoneme intelligibility measures under the six filter conditions were collapsed across the two replications, arcsine transformed to increase homogeneity of variance, and subjected to an analysis of variance with gender as a grouping variable and filter condition as repeated measures. The results are shown in Table 4.6. Of particular interest is the interaction between filter condition and gender. Basically, scores for the male talkers were higher under the low-pass conditions (by 4.5, 5.9, and 0.3 percentage points for cut-off frequencies of 707, 1414, and 2828 Hz, respectively) and scores for the female talkers were higher under the high-pass conditions (by 5.5, 3.9, and 1.0 percentage points for cut-off frequencies of 2828, 1414, and 707 Hz, respectively). These findings are illustrated in Figure 4.6. It is this interaction that accounts for the significant effect of gender on cross-over frequency reported earlier. Note that the magnitude of the gender effect is around 5 percentage points for scores under the two extreme filtering conditions (which are both around 50%).

Table 4.6. Analysis of variance in the arcsine transforms of phoneme intelligibility under six filtering conditions, collapsed across replications

	Sum of Squares	Degrees of freedom	Estimated Mean Square	F ratio	p level
Talker	2.57	1	2.57	0.11	0.76
Filter	8922.76	5	1784.55	265.80	0.00
TxF	104.44	5	20.89	3.11	0.03
Total	11118.59	11			

IX. Summary of Findings

1. The average cross-over frequency for the female talkers (mean = 1806 Hz) was about thirteen percent higher than the male talkers (mean = 1599 Hz). This was a significant difference.
2. Female talkers tended to be more intelligible (by approximately 5 percentage points) under high-pass filter conditions and male talkers tended to be more intelligible (by approximately 5 percentage points) under low-pass filter conditions.
3. Due to the small number of talkers, there was overlap in the 95% confidence limits for male and female talkers' cross-over frequencies.
4. There was no gender effect seen for half-band phoneme intelligibility.
5. There was a 10-15 percentage point spread in the phoneme intelligibility scores in the most severe filtering conditions (high-pass cut-off 1414 Hz and 2828 Hz and low-pass 707 Hz and 1414 Hz).
6. The average half-band phoneme intelligibility score was 83%.
7. Predicted full-band phoneme intelligibility score for the mean of all talkers was between 96% and 98%.

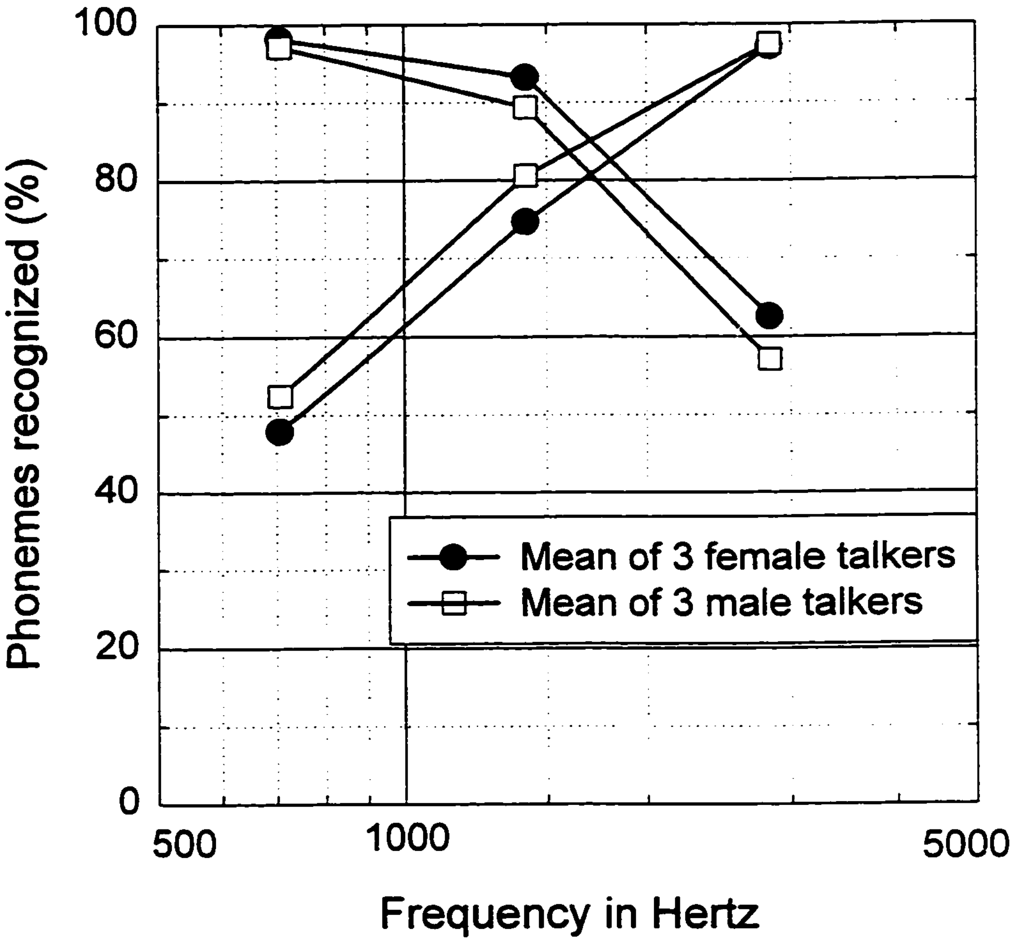


Figure 4.6. Phoneme intelligibility as a function of filter cut-off frequency for female and male talkers. With low-pass filtering, the male talkers are more intelligible. With high-pass filtering, the female talkers are more intelligible.

CHAPTER FIVE

ACOUSTIC ANALYSES

I. Purpose

The purpose of the acoustic analyses was to document acoustical differences among the six talkers with the goal of identifying potential correlates of behavioral findings.

II. Methods

The consonant-vowel-consonant words in each of the 20 word lists containing the point vowels /ɑ,i,u/ were imported into a DaDisp for Windows worksheet (see Appendix G). Therefore, a total of sixty words per talker was analyzed. The steady-state portion of each vowel was identified using a spectrogram created in the worksheet. From the steady portions of the vowels, fundamental frequency and first, second, and third formant frequency values were obtained. Due to the effects of coarticulation, there was difficulty locating the steady state portion of the vowel in cases where the vowel /u/ abutted a nasal consonant (either in the initial or final position). As a result, these tokens were excluded from this analysis. Another DaDisp worksheet (see Appendix H) calculated the long-term average speech spectrum of each of the 20 word lists produced by each talker. The values for the 20 lists were then averaged to obtain one long-term average speech spectrum value for each of the six talkers.

III. Results

1. Fundamental Frequency and Formant Frequency Values

Table 5.1 shows the mean fundamental and formant frequency values collapsed across male talkers, female talkers, and across all six talkers. Also shown are the differences between male and female fundamental and formant frequency values in octaves and percent. Tables 5.2-5.7 show the fundamental and formant frequencies in Hz for the three vowels /ɑ,i,u/ spoken by the three male talkers (talkers TM1, TM2, and TM3) and three female talkers (talkers TF1, TF2, and TF3), respectively. The mean and standard deviation of each fundamental and formant frequency value are also shown in these tables. The difference between male talkers' and female talkers' fundamental frequencies is approximately 0.9 octave (mean male fundamental frequency = 118 Hertz; mean female fundamental frequency = 220 Hertz). While the female formant frequency values of F1 and F3 were similarly around 20% higher than those for the male talkers (18% and 16%, respectively), there was only a 12% difference between the male and female F2 value.

Figure 5.1 shows the scatter plots of F2 vs. F1 for the same three vowels for all six talkers. The formants of the three male talkers are very similar, as are the formants for female talkers TF1 and TF3. For female talker TF2 for the vowel /i/, there was a lower mean F2 value for than the other two female talkers. Also, talker TF2 had a higher F1 and F2 value for the vowel /ɑ/ than the other two female talkers. This difference may reflect a smaller vocal tract size and length than the other two female talkers. Figures 5.2-5.7 illustrate the formant values for each individual talker.

Table 5.1. Mean frequencies (in Hz) of Fo and the first three formants for 3 male and 3 female talkers. Vowels were produced in consonant-vowel-consonant context.

		Talker									Female-male difference	
		Male				Female				All		
Vowel	Formant	TM1	TM2	TM3	Mean	TF1	TF2	TF3	Mean	Mean	Octaves	Percent
/i/	Fo	107	142	112	120	215	217	230	221	171	0.88	84
	F1	272	281	269	274	326	328	348	334	304	0.28	22
	F2	2290	2197	2298	2262	2895	2654	2903	2817	2540	0.32	25
	F3	3077	3156	2729	2987	3567	3254	3526	3449	3218	0.21	15
/a/	Fo	108	129	102	113	195	190	210	198	156	0.81	75
	F1	690	727	657	692	776	917	756	816	754	0.24	18
	F2	1213	1207	1214	1212	1238	1434	1222	1298	1255	0.10	7
	F3	2535	2351	2604	2496	2877	2710	2976	2854	2675	0.19	14
/u/	Fo	108	140	97	115	204	207	225	212	163	0.88	84
	F1	284	306	303	298	339	339	343	340	319	0.19	14
	F2	1231	1266	1200	1232	1368	1259	1241	1289	1261	0.07	5
	F3	2234	2278	2496	2336	2774	2768	2748	2763	2550	0.24	18
									Average of differences over 3 vowels	Fo	0.86	81
										F1	0.24	18
										F2	0.16	12
										F3	0.21	16

Table 5.2. Fundamental (Fo) and Formant frequencies in Hz for tokens of three vowels produced in consonant-vowel-consonant context by talker M1. Tokens of /u/ abutting nasals were excluded.

	Vowel /i/																					
	cheek	cheese	teak	wreath	heel	heap	reap	reach	weave	beep	siege	teeth	thieve	heath	chief	beach	thief	theme	teach	tease	mean	stdev
Fo	108	108	108	108	108	108	108	108	108	108	108	108	108	108	86	108	108	108	108	108	107	4.8
F1	258	237	258	280	258	258	323	280	301	258	258	280	258	280	280	258	280	323	258	258	272	22.4
F2	2239	2304	2476	2110	2476	2304	2067	2110	2239	2283	2304	2304	2347	2369	2283	2261	2239	2369	2412	2304	2290	108.5
F3	3230	2885	3316	3144	3316	3359	3058	3058	2993	2929	2972	2864	3230	2950	2756	3230	2972	3036	3101	3144	3077	166.8
	Vowel /a/																					
	jot	log	bomb	job	shop	got	not	pot	rob	rod	cough	shock	dodge	fog	moth	hop	chop	dog	lodge	watch	mean	stdev
Fo	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	0.0
F1	689	646	689	689	689	711	732	732	668	689	689	646	668	668	732	754	732	624	668	646	690	35.3
F2	1270	1120	1120	1163	1120	1270	1249	1227	1227	1184	1249	1314	1184	1141	1378	1227	1227	1163	1184	x	1213	69.2
F3	2412	2778	2864	2390	2864	2304	2412	2649	2606	2046	2455	2326	2498	2476	2713	2584	2369	2476	2799	2799	2535	217.3
	Vowel /u/																					
	x	x	food	choose	x	juice	goose	noose	lose	shoes	shoot	x	x	x	loose	suit	x	x	x	booth	mean	stdev
Fo	x	x	108	108	x	108	108	108	108	108	108	x	x	x	108	108	x	x	x	108	108	0.0
F1	x	x	280	280	x	280	301	301	280	280	258	x	x	x	301	258	x	x	x	301	284	16.2
F2	x	x	904	1249	x	1378	1120	1314	1270	1292	1270	x	x	x	1357	1292	x	x	x	1098	1231	138.5
F3	x	x	2175	2196	x	2218	2283	2261	2196	2153	2239	x	x	x	2304	2261	x	x	x	2283	2234	49.2

Table 5.3. Fundamental (Fo) and Formant frequencies in Hz for tokens of three vowels produced in consonant-vowel-consonant context by talker M2. Tokens of /u/ abutting nasals were excluded.

	Vowel /i/																					
	cheek	cheese	teak	wreath	heel	heap	reap	reach	weave	beep	siege	teeth	thieve	heath	chief	beach	thief	theme	teach	tease	mean	stdev
Fo	151	151	151	151	151	151	129	129	151	129	151	129	129	172	129	129	129	151	151	129	142	12.9
F1	258	280	323	280	301	301	280	323	258	280	280	258	258	258	323	258	280	301	258	258	281	23.7
F2	2110	2196	2196	2024	2196	2175	2175	1895	2239	2218	2218	2218	2175	2347	2110	2261	2175	2541	2218	2261	2197	123.7
F3	3618	3467	3079	2735	2993	3252	2606	3230	3144	3252	3165	3079	3144	3295	3101	3230	3122	3295	3165	3144	3156	217.8
	Vowel /a/																					
	jot	log	bomb	job	shop	got	not	pot	rob	rod	cough	shock	dodge	fog	moth	hop	chop	dog	lodge	watch	mean	stdev
Fo	129	129	129	108	129	129	129	129	129	129	129	129	129	151	129	129	129	129	129	129	129	7.0
F1	711	732	732	711	732	754	797	754	732	711	732	754	711	689	732	754	689	624	732	754	727	34.8
F2	1292	1120	1227	1270	1292	1249	1227	1249	1206	1206	1141	1249	1206	1184	1141	1206	1270	1098	1184	1120	1207	58.6
F3	2239	2756	2153	2283	2196	2455	2110	2369	2584	2455	2476	2283	2369	2390	2239	2476	2196	2390	2476	2412	2351	157.2
	Vowel /u/																					
	x	x	food	choose	x	juice	goose	noose	lose	shoes	shoot	x	x	x	loose	suit	x	x	x	booth	mean	stdev
Fo	x	x	129	129	x	129	x	151	129	150	151	x	x	x	151	151	x	x	x	129	140	11.3
F1	x	x	323	323	x	301	x	280	280	301	323	x	x	x	323	301	x	x	x	301	306	17.0
F2	x	x	1077	1421	x	1378	x	1335	1507	1163	1184	x	x	x	1163	1335	x	x	x	1098	1266	147.7
F3	x	x	2239	2239	x	2283	x	2283	2390	2261	2218	x	x	x	2283	2218	x	x	x	2369	2278	59.0

Table 5.4. Fundamental (Fo) and Formant frequencies in Hz for tokens of three vowels produced in consonant-vowel-consonant context by talker M3. Tokens of /u/ abutting nasals were excluded.

	Vowel /i/																					
	cheek	cheese	teak	wreath	heel	heap	reap	reach	weave	beep	siege	teeth	thieve	heath	chief	beach	thief	theme	teach	tease	mean	stdev
Fo	97	97	118	97	118	118	118	97	118	118	97	118	118	118	97	118	118	118	118	118	112	10.1
F1	269	237	269	280	280	258	291	301	248	269	269	269	248	269	269	269	280	301	269	237	269	17.8
F2	2207	2229	2336	2304	2390	2390	2003	2304	2347	2229	2347	2326	2401	2379	2143	2283	2293	2390	2315	2336	2298	97.6
F3	2584	2584	2724	2713	2864	2778	2745	2778	2842	2745	2756	2789	2789	2778	2552	2584	2745	2799	2606	2832	2729	95.2
	Vowel /a/																					
	jot	log	bomb	job	shop	got	not	pot	rob	rod	cough	shock	dodge	fog	moth	hop	chop	dog	lodge	watch	mean	stdev
Fo	118	118	97	118	118	97	97	97	86	97	97	97	97	97	97	97	129	97	97	97	102	11.3
F1	711	624	624	657	678	678	689	711	646	646	668	700	592	646	624	732	711	560	624	646	657	43.8
F2	1152	1593	1324	1066	1281	1324	1152	1109	1098	1034	1055	1174	1249	1163	1400	1152	1195	1443	1055	1583	1214	168.1
F3	2369	2821	2961	2724	2207	2466	2745	2638	2713	2606	2713	2207	2444	2562	3015	2638	2207	2681	2670	2670	2604	228.5
	Vowel /u/																					
	x	x	food	choose	x	juice	goose	noose	lose	shoes	shoot	x	x	x	loose	suit	x	x	x	booth	mean	stdev
Fo	x	x	129	118	x	97	97	118	118	97	129	x	x	x	97	97	x	x	x	97	109	14.0
F1	x	x	258	280	x	280	301	323	258	301	301	x	x	x	366	345	x	x	x	323	303	34.0
F2	x	x	818	1227	x	1443	1012	1507	1055	1163	1400	x	x	x	1206	1421	x	x	x	947	1200	225.9
F3	x	x	2692	2326	x	2390	2606	2455	2541	2390	2218	x	x	x	2670	2541	x	x	x	2627	2496	152.4

Table 5.5. Fundamental (Fo) and Formant frequencies in Hz for tokens of three vowels produced in consonant-vowel-consonant context by talker F1. Tokens of /u/ abutting nasals were excluded.

	Vowel /i/																					
	cheek	cheese	teak	wreath	heel	heap	reap	reach	weave	beep	siege	teeth	thieve	heath	chief	beach	thief	theme	teach	tease	mean	stdev
Fo	194	194	237	215	194	237	215	215	215	215	215	215	194	215	258	237	215	194	237	194	215	18.5
F1	345	323	345	301	345	345	323	323	301	323	323	323	301	301	366	345	323	301	366	301	326	21.3
F2	2993	2110	2993	2649	2713	3101	2821	2756	3015	2972	2950	2972	3101	3058	2885	2950	2993	3058	2950	2864	2895	222.1
F3	3295	2972	3295	3058	3036	3531	3252	3187	3553	3402	4070	3768	3919	3984	3898	3898	3316	3467	4242	4199	3567	400.9

	Vowel /a/																					
	jot	log	bomb	job	shop	got	not	pot	rob	rod	cough	shock	dodge	fog	moth	hop	chop	dog	lodge	watch	mean	stdev
Fo	194	194	194	194	194	194	194	194	194	194	172	194	194	194	194	215	215	194	194	194	195	8.5
F1	775	818	840	732	797	754	732	754	754	732	775	797	754	797	732	840	818	711	861	754	776	42.8
F2	1292	1357	1249	1120	1206	1357	1292	1163	1120	1292	1270	1292	1314	1184	1120	1249	1227	1270	1249	1141	1238	75.7
F3	2821	2864	3230	2799	2519	3036	3144	2799	2821	2885	2799	2649	2821	2950	3122	2972	2649	2907	3079	2670	2877	183.6

	Vowel /u/																					
	x	x	food	choose	x	juice	goose	noose	lose	shoes	shoot	x	x	x	loose	suit	x	x	x	booth	mean	stdev
Fo	x	x	194	194	x	194	215	194	194	194	194	x	x	x	237	237	x	x	x	194	204	17.7
F1	x	x	280	345	x	323	345	345	345	345	345	x	x	x	345	366	x	x	x	345	339	21.7
F2	x	x	991	1593	x	1572	1206	1529	1206	1378	1486	x	x	x	1270	1615	x	x	x	1206	1368	205.9
F3	x	x	2390	2821	x	2864	2756	2842	2842	2756	2842	x	x	x	2734	2756	x	x	x	2907	2774	138.5

Table 5.6. Fundamental (Fo) and Formant frequencies in Hz for tokens of three vowels produced in consonant-vowel-consonant context by talker F2. Tokens of /u/ abutting nasals were excluded.

	Vowel /i/																					
	cheek	cheese	teak	wreath	heel	heap	reap	reach	weave	beep	siege	teeth	thieve	heath	chief	beach	thief	theme	teach	tease	mean	stdev
Fo	215	215	237	215	215	237	194	194	215	215	215	237	194	237	215	215	215	215	237	215	217	13.8
F1	345	323	355	323	323	345	301	323	323	323	301	345	323	345	345	323	323	301	345	323	328	15.8
F2	2627	2584	2692	2369	2713	2649	2390	2735	2713	2735	2670	2692	2972	2713	2562	2799	2326	2735	2799	2606	2654	154.2
F3	3208	3467	3176	2606	3359	3230	3165	3144	3596	3618	2993	3402	3359	3445	3252	3467	2950	3187	3402	3058	3254	240.1
	Vowel /a/																					
	jot	log	bomb	job	shop	got	not	pot	rob	rod	cough	shock	dodge	fog	moth	hop	chop	dog	lodge	watch	mean	stdev
Fo	172	172	205	194	194	194	194	194	172	172	215	172	194	194	194	215	194	194	172	194	190	13.7
F1	947	1357	851	926	883	947	969	969	926	904	797	926	991	947	818	904	926	969	904	947	917	110.0
F2	1486	1895	1174	1486	1464	1529	1572	1378	1443	1443	1400	1464	1550	1335	1400	1464	1507	1357	1486	1357	1434	136.2
F3	3661	2842	2724	2606	2541	2649	3165	2735	2778	2670	2627	2735	2735	2670	2778	2670	2627	2649	3144	2283	2710	282.6
	Vowel /u/																					
	x	x	food	choose	x	juice	goose	noose	lose	shoes	shoot	x	x	x	loose	suit	x	x	x	booth	mean	stdev
Fo	x	x	183	215	x	194	215	194	215	215	258	x	x	x	194	194	x	x	x	194	207	20.9
F1	x	x	345	323	x	345	345	345	345	345	345	x	x	x	345	323	x	x	x	323	339	10.1
F2	x	x	1055	1206	x	1357	1357	1292	1335	1335	1270	x	x	x	1184	1249	x	x	x	1206	1259	92.0
F3	x	x	2756	2627	x	2821	2756	2778	2799	2692	2692	x	x	x	2821	2778	x	x	x	2929	2768	80.2

Table 5.7. Fundamental (Fo) and Formant frequencies in Hz for tokens of three vowels produced in consonant-vowel-consonant context by talker F3. Tokens of /u/ abutting nasals were excluded.

	Vowel /i/																					
	cheek	cheese	teak	wreath	heel	heap	reap	reach	weave	beep	siege	teeth	thieve	heath	chief	beach	thief	theme	teach	tease	mean	stdev
Fo	237	215	258	237	237	237	237	237	205	258	205	215	215	258	215	258	237	215	215	215	230	18.3
F1	345	323	366	366	345	388	366	366	388	345	301	323	323	388	323	388	366	323	323	301	348	29.0
F2	2885	2885	2993	2756	2756	2972	2821	2821	2864	2950	2918	3036	2907	3036	2756	2993	2972	2929	2929	2885	2903	87.5
F3	3445	3359	3531	3510	3273	3510	3768	3575	3402	3467	3381	3488	3488	3553	3854	3424	3661	3531	3747	3553	3526	143.2
	Vowel /a/																					
	jot	log	bomb	job	shop	got	not	pot	rob	rod	cough	shock	dodge	fog	moth	hop	chop	dog	lodge	watch	mean	stdev
Fo	205	205	205	215	215	194	194	215	194	215	215		215	205	205	237	237	183	205	237	210	15.0
F1	883	818	797	668	754	711	754	711	754	818	861	764	732	775	700	732	797	732	775	775	756	53.5
F2	1260	1141	1195	1227	1184	1206	1314	1206	1227	1292	1249	1195	1174	1174	1120	1292	1260	1206	1238	1238	1222	50.0
F3	3079	3155	3262	3015	2347	3015	3187	2929	2950	3112	2907	2961	2993	2982	3122	3047	2993	2972	2950	2821	2976	183.8
	Vowel /u/																					
	x	x	food	choose	x	juice	goose	noose	lose	shoes	shoot	x	x	x	loose	suit	x	x	x	booth	mean	stdev
Fo	x	x	237	194	x	237	237	237	194	194	237	x	x	x	280	237	x	x	x	194	225	27.9
F1	x	x	323	323	x	345	345	345	345	345	345	x	x	x	323	388	x	x	x	345	343	17.9
F2	x	x	1163	1206	x	1357	1357	1292	1292	1335	1270	x	x	x	1163	1270	x	x	x	947	1241	119.2
F3	x	x	2649	2627	x	2821	2756	2778	2778	2799	2692	x	x	x	2756	2842	x	x	x	2735	2748	68.3

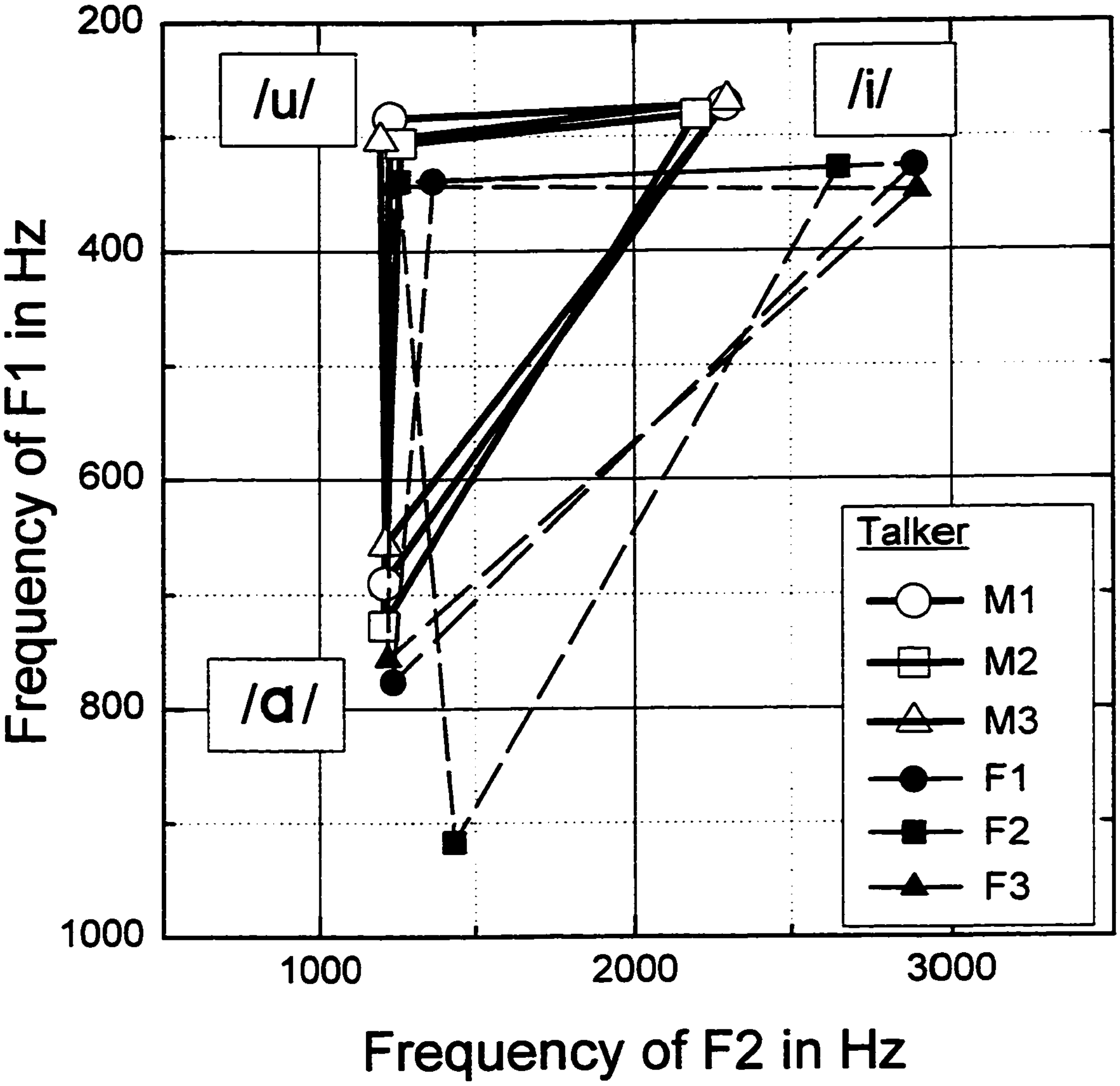


Figure 5.1. F1 versus F2 scatter plots for 3 vowels produced in consonant-vowel-consonant words by six talkers. Data points show individual means collapsed over several tokens.

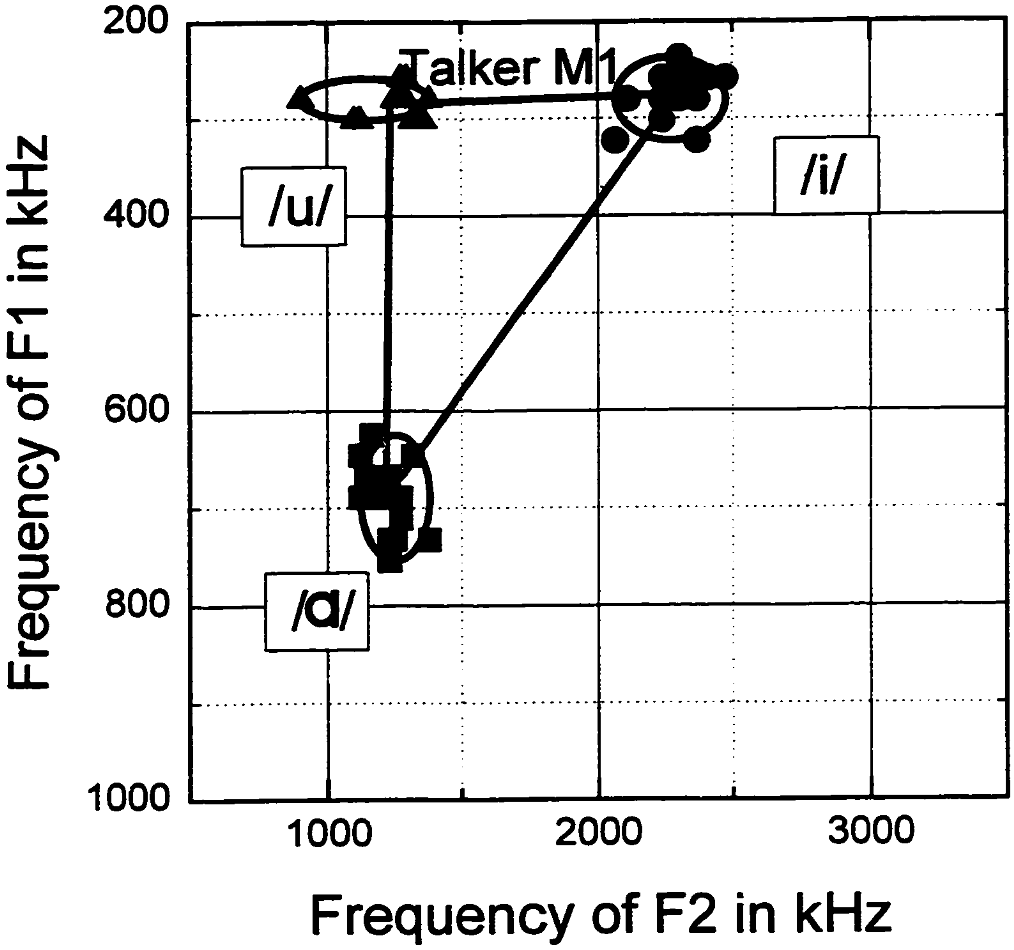


Figure 5.2. F1 versus F2 scatter plots for 3 vowels produced in consonant-vowel-consonant words by talker M1. Straight lines join means for the three vowels. Axes of the ellipses extend to the maximum and minimum values for each vowel.

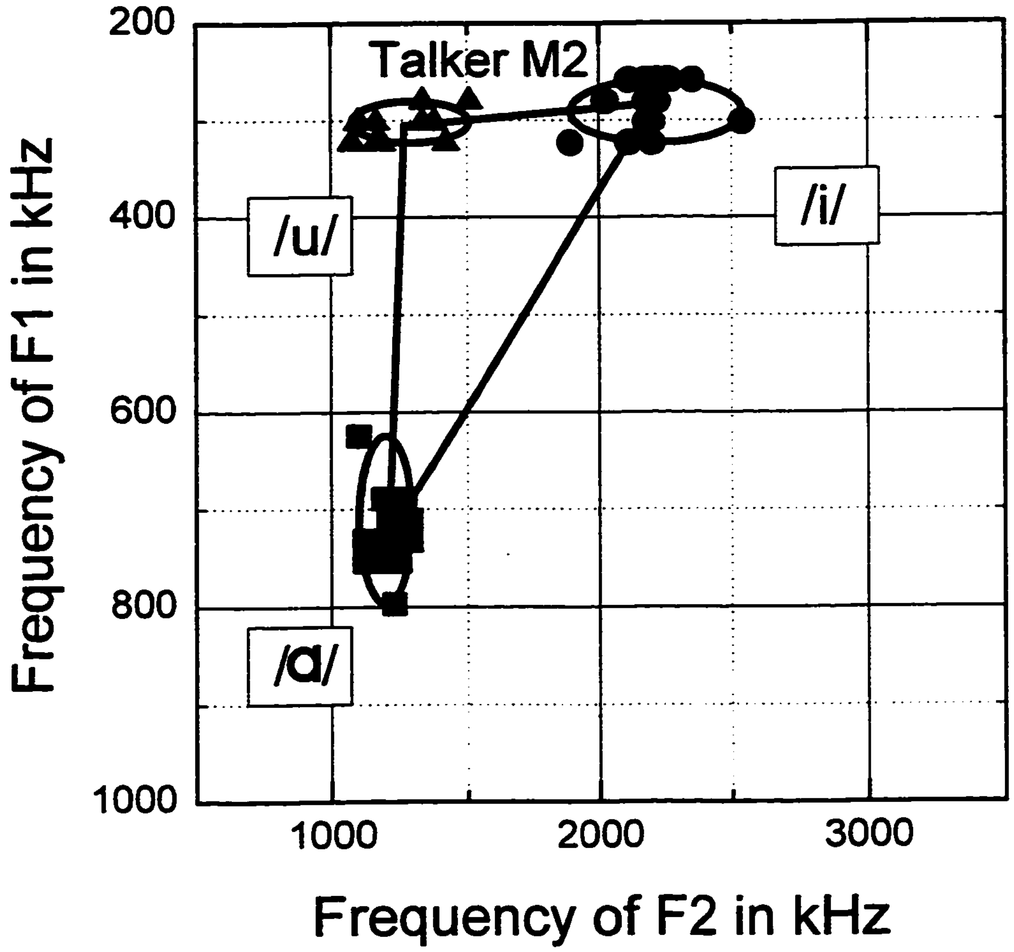


Figure 5.3. F1 versus F2 scatter plots for 3 vowels produced in consonant-vowel-consonant words by talker M2. Straight lines join means for the three vowels. Axes of the ellipses extend to the maximum and minimum values for each vowel.

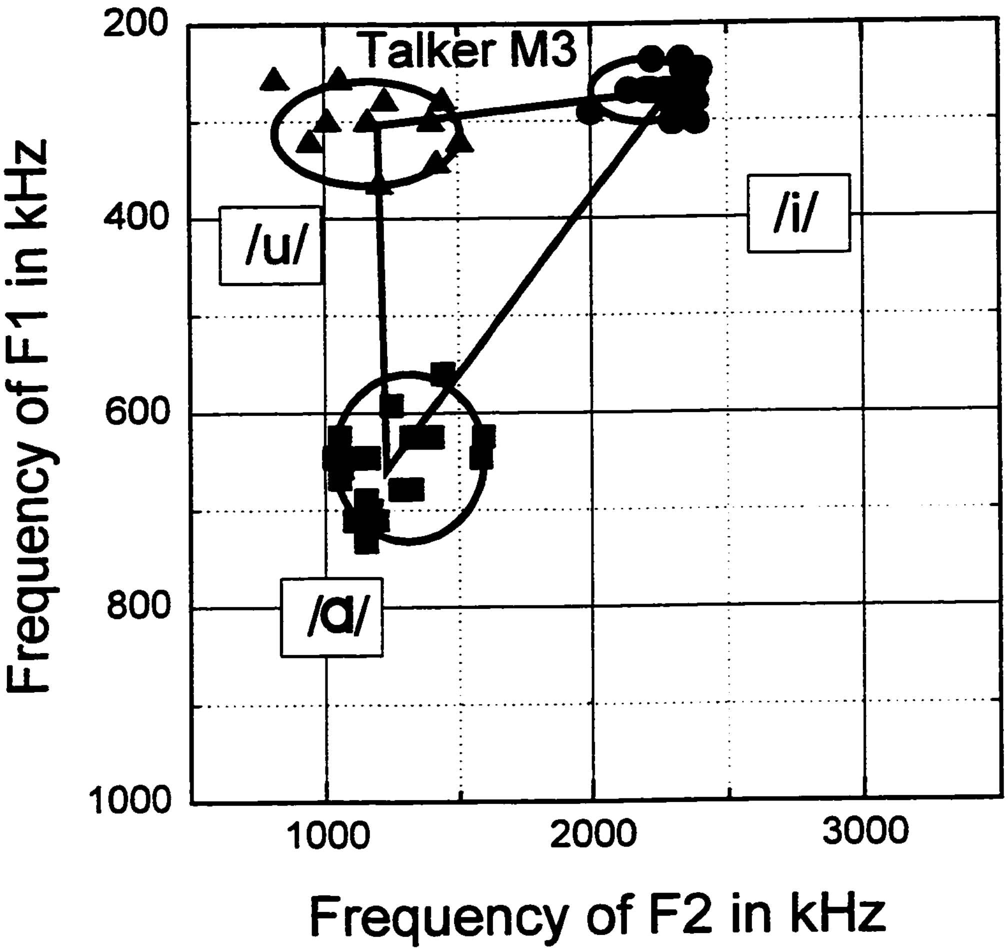


Figure 5.4. F1 versus F2 scatter plots for 3 vowels produced in consonant-vowel-consonant words by talker M3. Straight lines join means for the three vowels. Axes of the ellipses extend to the maximum and minimum values for each vowel.

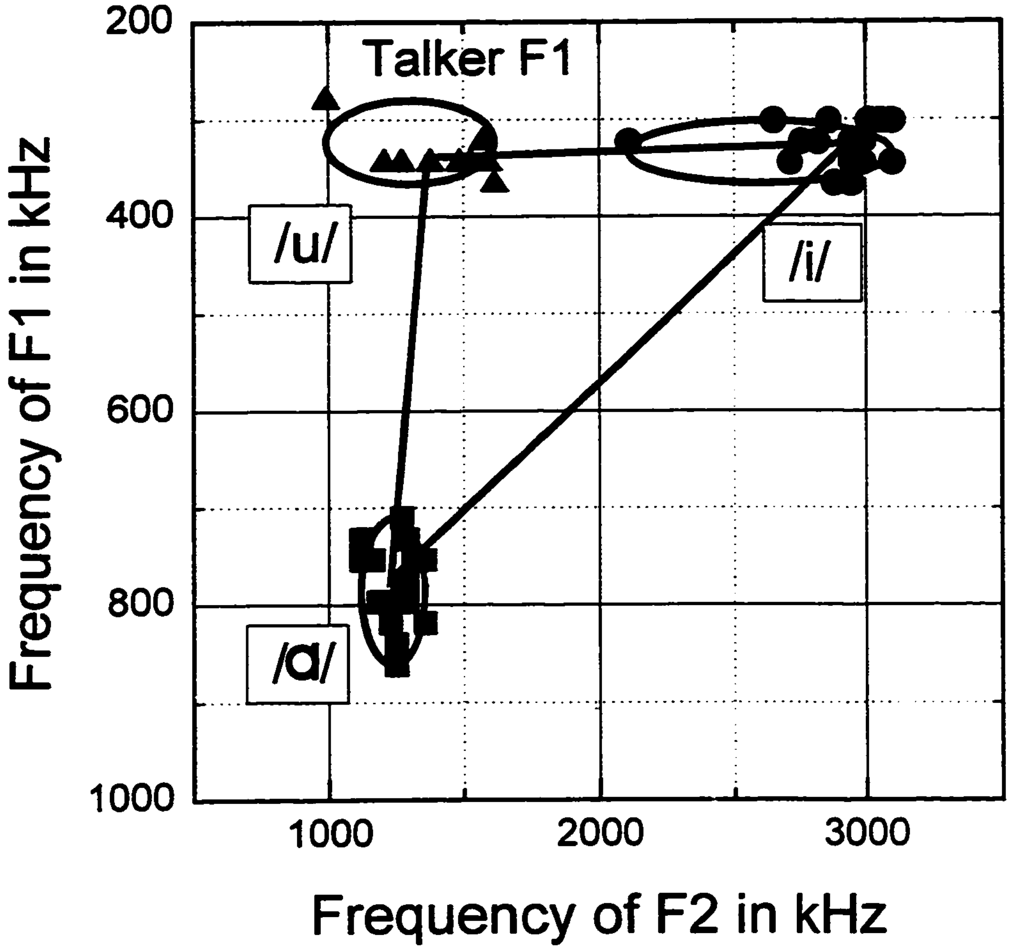


Figure 5.5. F1 versus F2 scatter plots for 3 vowels produced in consonant-vowel-consonant words by talker F1. Straight lines join means for the three vowels. Axes of the ellipses extend to the maximum and minimum values for each vowel.

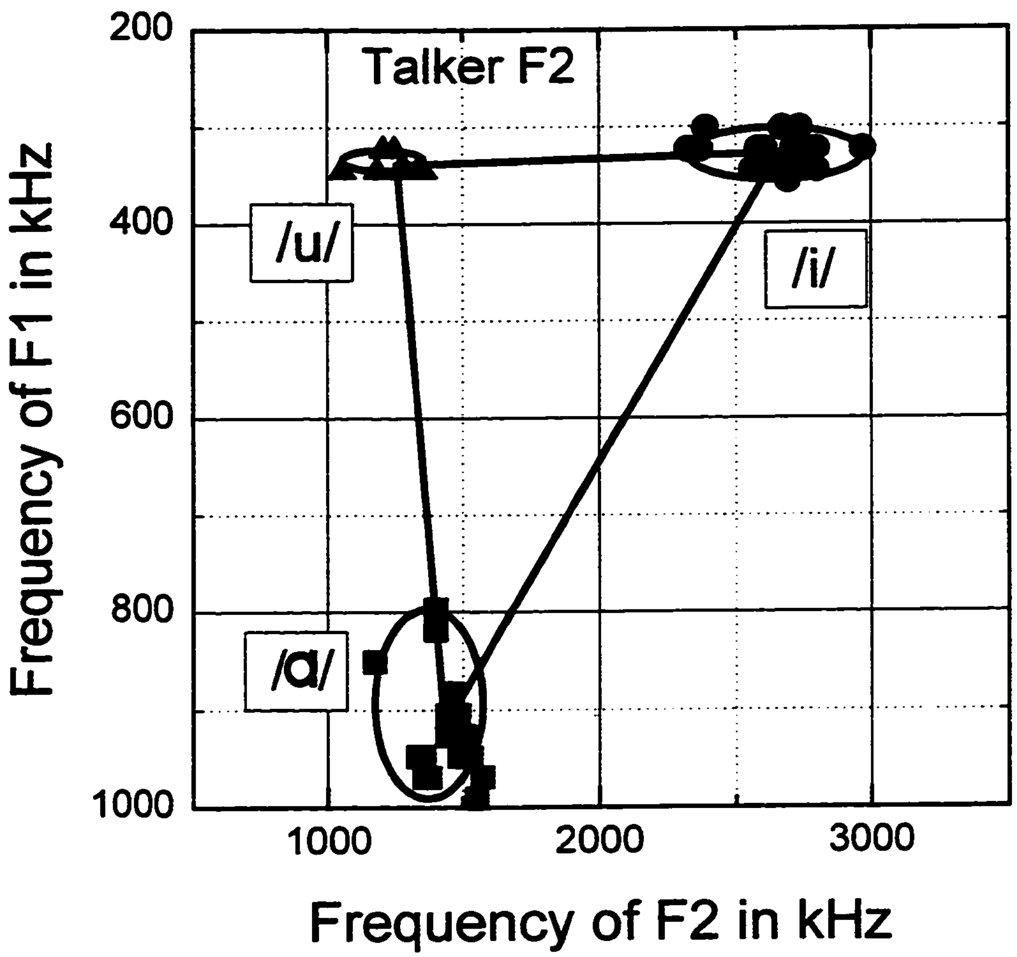


Figure 5.6. F1 versus F2 scatter plots for 3 vowels produced in consonant-vowel-consonant words by talker F2. Straight lines join means for the three vowels. Axes of the ellipses extend to the maximum and minimum values for each vowel.

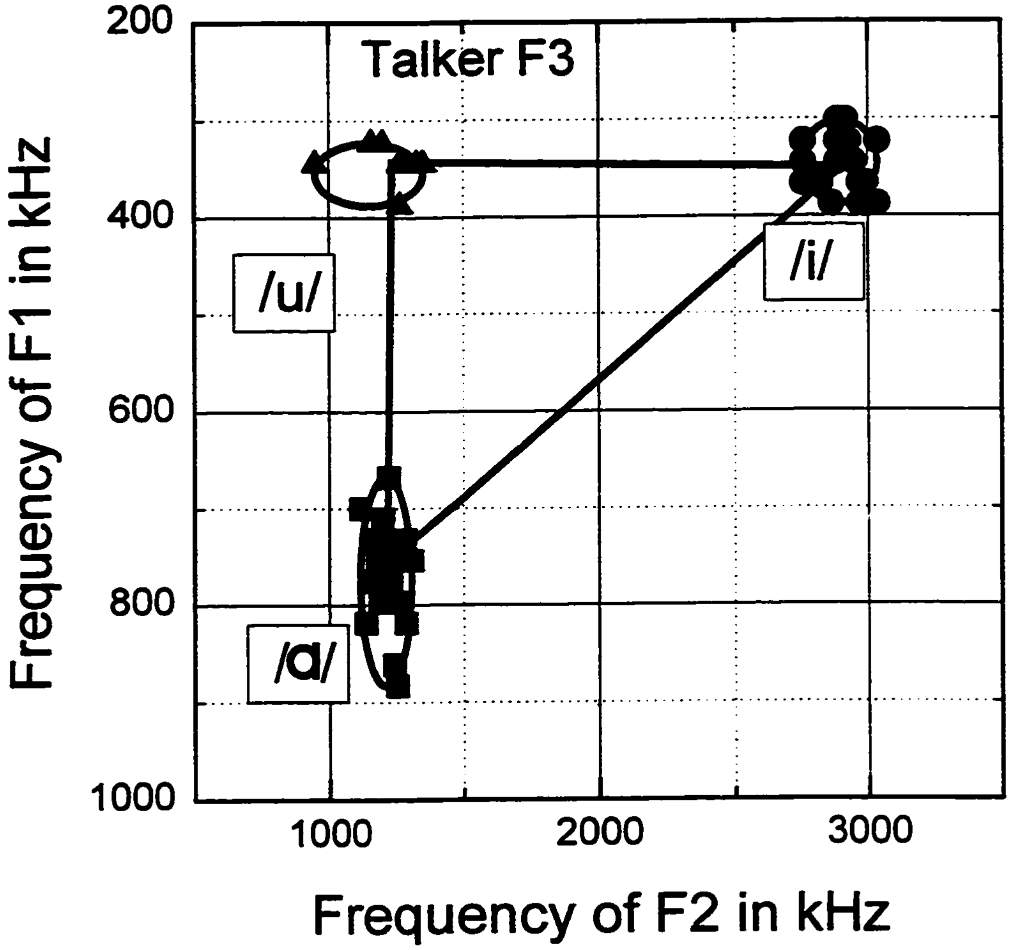


Figure 5.7. F1 versus F2 scatter plots for 3 vowels produced in consonant-vowel-consonant words by talker F3. Straight lines join means for the three vowels. Axes of the ellipses extend to the maximum and minimum values for each vowel.

2. Long-term average speech spectrum

Figure 5.8 shows the long-term average speech spectrum for each of the six talkers. The top three graphs illustrate results for the three female talkers and the bottom three graphs illustrate results for the three male talkers. The y-axis represents the relative spectrum level in decibels integrated over 194 Hz and the x-axis represents frequency in kHz. The numbers in the boxes on each graph represent the gradients in dB per octave below and above 250 Hz and 500 Hz. While female talkers TF2 and TF3 and male talkers TM2 and TM3 have similar slopes (between 7 and 8 dB per octave), female talker TF1 had the lowest slope (5.7 dB per octave) and male talker TM1 had the highest slope (9.2 dB per octave). Note that these slopes refer, essentially, to spectrum level (level per cycle).When expressed in terms of 1/3 octave levels, they would be decreased by about three dB per octave.

To investigate other possible differences in the spectra of the talker's voices, the energy ratio of the high and low portion of each talkers' voice was measured. The energy below 1000 Hz was compared to the energy above 1000 Hz and the low/high energy ratio was calculated. Figure 5.9, (shown below) illustrates the low/high energy ratio in dB for each of the six talkers. It will be seen that the range of ratios is from 11 dB to 6 dB. There is a tendency for higher ratios in male talkers than female talkers but the data for the two groups overlap. Note that gender-related differences of spectral shape have been reported previously by Pearsons, et. al, (1977).

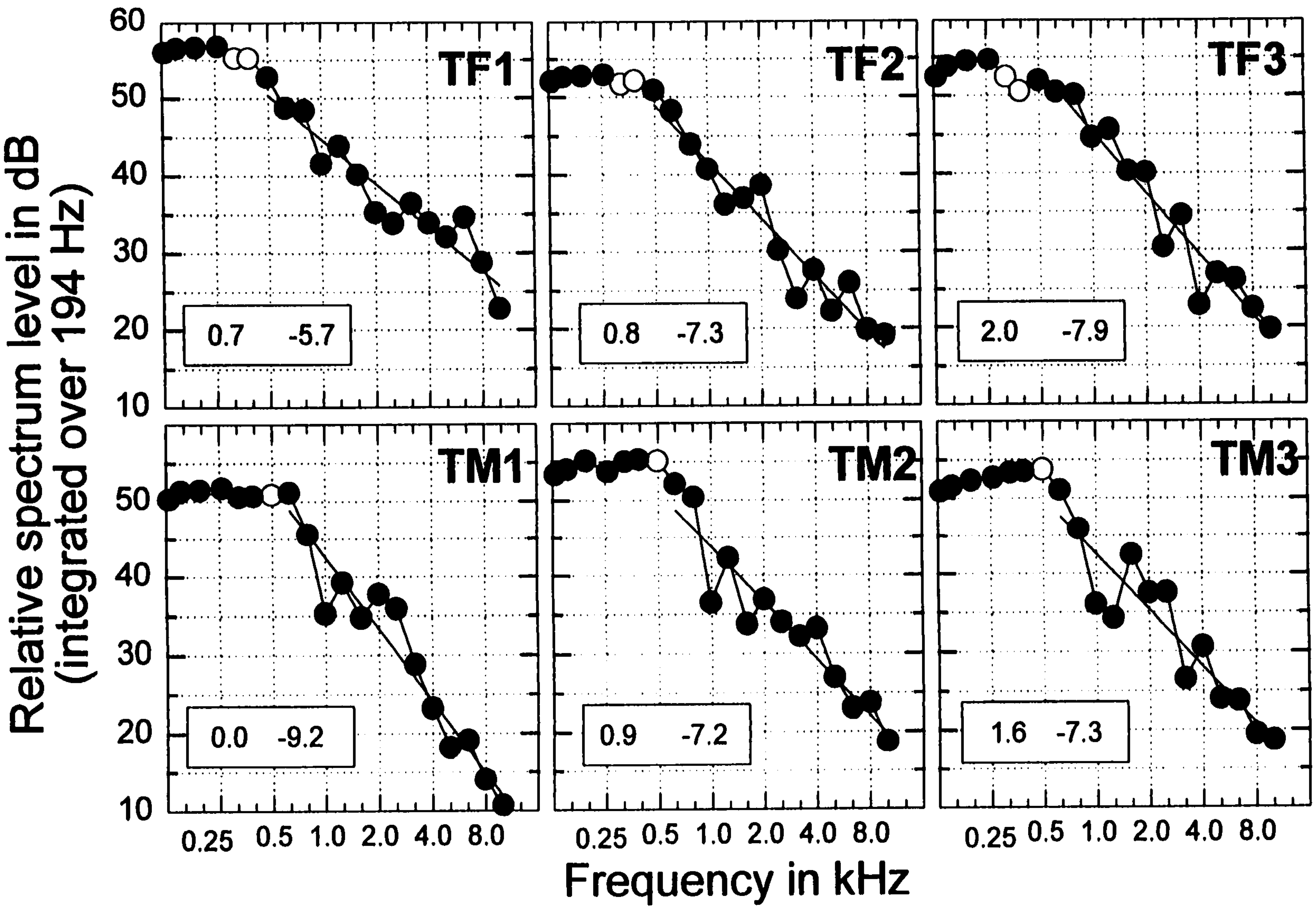


Figure 5.8. Long-term average spectrum of the speech of six talkers. Numbers show gradients in dB/oct below and above 250 Hz (500 Hz for Male talkers).

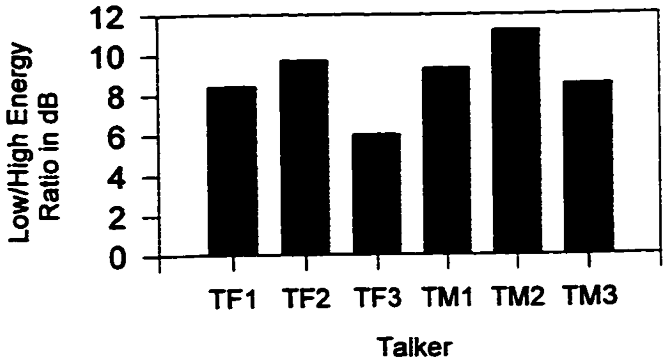


Figure 5.9. Energy ratio of the spectrum of the high and low portion of each talker's voice. The energy below and above 1000 Hz was compared and the low/high energy ratio was calculated.

3. Cross-over frequency and half-band phoneme intelligibility as functions of all six acoustic measures

Figure 5.10 shows the two behavioral measures, cross-over frequency and half-band phoneme intelligibility score, as functions of all the acoustic measures taken. The top six graphs represent the cross-over frequency as a function of fundamental frequency in Hz, first, second, and third formants in Hz, the mean formant frequency in Hz, and the long-term average speech spectrum slope in dB per octave above 500 Hz. The fundamental and formant frequency measures were collapsed across the three vowels measured (/ɑ,i,u/).

Also shown on each graph are the linear correlations and regression functions. The critical values for r (for four degrees of freedom) are: 0.81 for $p = 0.05$, and 0.73 for

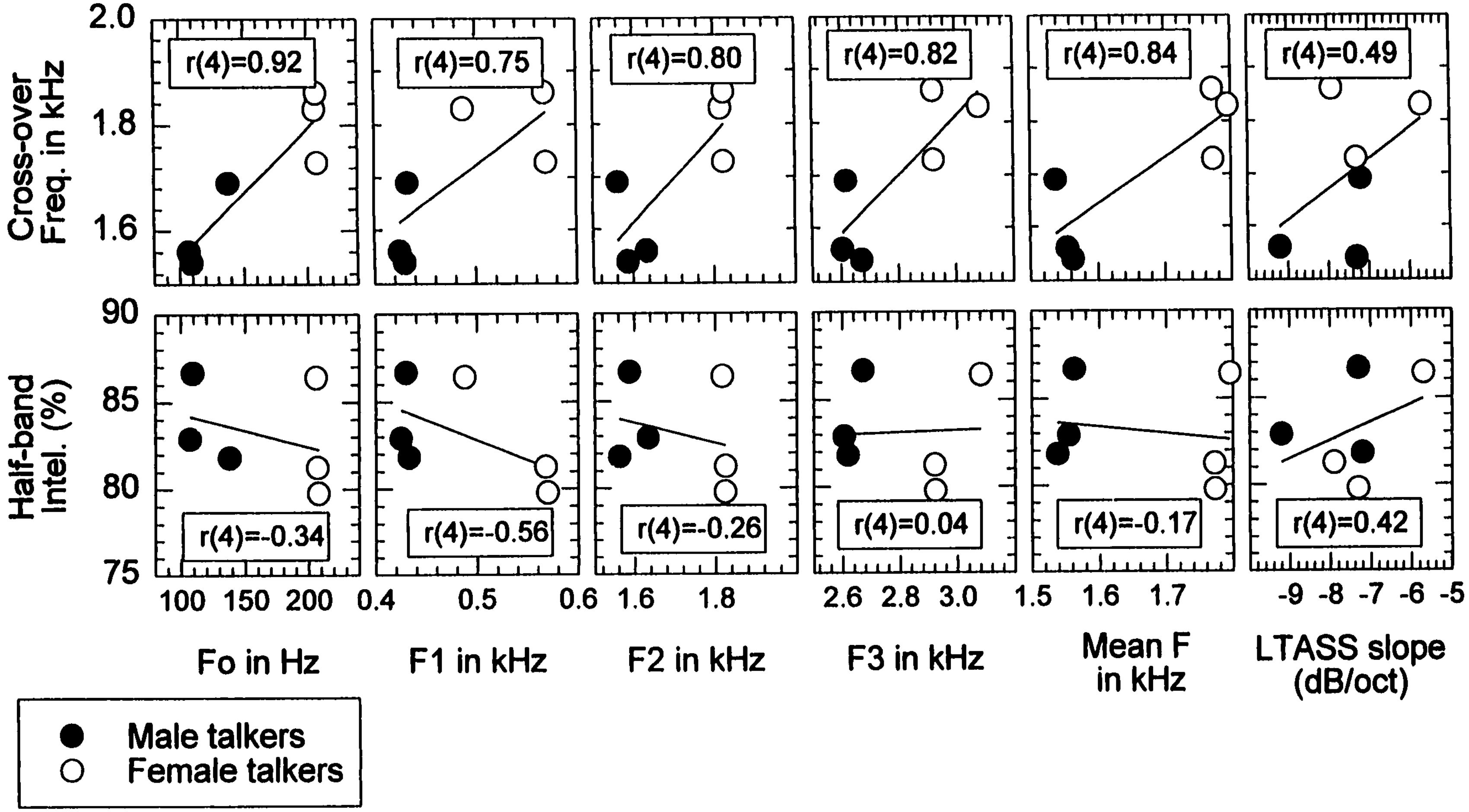


Figure 5.10. Cross-over frequency and 1/2-band phoneme intelligibility as functions of 6 acoustic measures in 6 talkers. Linear correlations and regression functions are shown. Critical values for r = 0.81 and 0.73 (for p=0.05 and 0.1, respectively).

$p = 0.01$. A strong positive correlation for all acoustic measures of fundamental and formant frequency values and cross-over frequency should be noted. The correlation for long-term average speech spectrum and cross-over frequency was not significant ($r(4)=0.49$). The strongest correlation was seen between fundamental frequency and cross-over frequency ($r(4)=0.92$). The linear correlations for the half-band intelligibility scores and the six acoustic measures were not significant

IV. Discussion

The correlation between cross-over frequency and various frequency measures is not surprising considering the large effect of gender. Cross-over occurred at a higher frequency for female talkers - who also have higher fundamental and formant frequency values. Because female talkers tend to have higher cross-over frequencies and higher voice fundamental frequencies, a positive correlation was inevitable and does not necessarily reveal a cause-effect relationship. The high information content of the second formant makes it reasonable to infer a cause-effect relationship for this variable. The danger of confounding correlation with cause, however, is emphasized by the fact that the best predictor of cross-over frequency was fundamental frequency, for which no obvious cause-effect relationship can be offered.

None of the acoustic measures offered was able to predict half-band phoneme intelligibility. Presumably, one would need to examine other acoustic parameters in order to identify potential correlates. It is possible, for example, that individual properties of articulation the primary determinants of phoneme intelligibility. These issues will be examined at more length in the general discussion.

V. Conclusions

1. Formant frequencies and fundamental frequency were higher for the female talkers than for the male talkers.
2. Formant frequencies and fundamental frequency accounted for significant amounts of variance in measured cross-over frequency.
3. Because the formant patterns carry considerable phonemic information, it is reasonable to assume that higher formant values cause higher cross-over frequencies, i.e., an upward shift of the frequency-importance function, as seen in the female talkers.
4. The correlation of cross-over frequency with fundamental frequency, however, most likely reflects the fact that women have a higher fundamental frequency value. There is no obvious way in which the value of the fundamental frequency might influence the frequency-importance function.
5. There was no evidence of an association between formant or fundamental frequencies and half-band intelligibility.

CHAPTER 6

GENERAL DISCUSSION

I. Cross-over frequency

The principal purpose of this study was to determine the presence of inter-talker differences - including effects of gender - and their effects on phoneme intelligibility. The dependent variable of primary interest was cross-over frequency, or the point on the importance function for which high- and low-pass filtering provided identical phoneme intelligibility.

The mean cross-over frequency for the six talkers was 1702 Hz. The mean cross-over frequency for the female talkers was about 200 Hz higher than for the male talkers (1806 Hz and 1599 Hz, respectively). Hirsh, et al. (1954) reported an average cross-over frequency of 1675 Hz for male and female talkers. The average cross-over frequency for the male talkers only (1599 Hz) is similar to the 1450 Hz found by Black (1959) and the cross-over frequency for the female talkers only (1806 Hz) is similar to the 1920 Hz found by Pollack and Pickett (1964). When Studebaker, et. al (1987) reviewed cross-over frequencies generated by male and female talkers in several studies, they found that the studies using only female talkers reported higher cross-over frequencies than those using both male and female talkers or male talkers. In each of these studies, the test material was meaningful consonant-vowel-consonant words. Therefore, the cross-over frequency values reported in the literature support those found in this study.

The results of the present study provide strong evidence to support the conclusion that the cross-over frequency is higher for female talkers than for male talkers. Considering the small sample size, the finding of statistical significance suggests that this effect is very robust.

The finding of a higher cross-over frequency for female talkers is not surprising. Many acoustic cues known to contribute to phoneme recognition are higher in frequency for females than for males (Eguchi and Hirsh, 1969; Hillenbrand, et.al, 1995; Peterson and Barney, 1952). The higher formant frequency values seen for females in the studies reviewed in Chapter 2 were also seen in the acoustical analyses done for this study. This male/female formant frequency difference is attributed to anatomical differences between males and females, namely, the shorter vocal tracts of the females. Formant patterns are known to be primary cues for vowel recognition, while formant transitions contribute to consonant recognition. Moreover, there is evidence to show that the fricative spectra of /s/ and /ʃ/ are higher for females than for males (Boothroyd and Medwetsky, 1992).

The mean second formant value for the female talkers was 12% higher than the male talkers. Moreover, the mean cross-over frequency for the female talkers was 13% higher than the male talkers. It is tempting to take the quantitative similarity of these findings as support for the conclusion that second formant frequency of the talker determines the cross-over frequency. It should be noted, however, that because of the small sample size there were wide confidence limits for estimates of mean cross-over frequencies for the populations of men and women represented by these samples. It would be wrong, therefore, to attach too much significance to quantitative similarities between second formant frequency and differences in cross-over frequency.

Another reason for caution in interpreting the cross-over frequency values obtained in this study is the interpolation method used. The logarithmic scale was chosen for frequency in Figures 4.2 - 4.4 because this transformation increases the symmetry of the frequency-importance function (according to the published literature). The interpolation used in this study, however, (see Appendix F) assumes a linear relationship between phoneme intelligibility and the logarithm of frequency between 1414 and 2828 Hz. The data of French and Steinberg (1947) and others in fact show a curvilinear relationship. It is expected, therefore, that this method of interpolation may have produced some systematic error of estimation in addition to unavoidable random error of measurement.

Lastly, while the mean values of cross-over frequency for female talkers and male talkers were significantly different, there was overlap in the estimated ranges. Therefore, there are some female talkers with cross-over frequencies lower than some male talkers, and vice versa. Due to considerable variability within gender group, it is clear that gender alone does not determine cross-over frequency.

II. Half-band phoneme intelligibility

Based on the data from the six talkers in this study, it was concluded that 95% of adults would have half-band phoneme intelligibility scores between about 76% and 90%, which represents a 14 percentage point range. There was considerable individual variation in the half-band phoneme intelligibility scores. While some of this range might be attributable to measurement error, a major portion must be attributed to genuine differences in the intelligibility of these six talkers. Additionally, the gender effect that was seen for cross-over frequency was not seen for half-band phoneme intelligibility. Thus, this study finds that while there were differences among the talkers, there was no evidence of inherent intelligibility differences between male and female talkers.

There was a significant effect of replication in this study, which suggests the presence of a learning effect. The implications of a learning effect on half-band phoneme intelligibility will be discussed later in this chapter in greater detail.

One aspect of the half-band intelligibility data warrants further discussion. Full-band intelligibility was predicted from the half-band intelligibility. Listeners in this study performed better than the predicted phoneme recognition scores in the two most favorable listening conditions: high-pass 707 Hz and low-pass 2828 Hz. For these two conditions many of the scores were near 100%. One explanation could be that the overlapping high- and low-pass filter skirts (see Figure 3.1, Chapter 3) would have allowed some redundancy in the estimates of half-band intelligibility. The result of such an error would, however, be the overestimation of full-band intelligibility because a small part of the acoustic spectrum would be represented twice. It could be argued that the underlying assumptions of independence in Articulation Theory, and embodied in equation (2), are incorrect. To test this possibility would require more detailed data than are available from the present study.

A more likely reason for the discrepancy between predicted and measured full-band intelligibility is an underestimation of half-band intelligibility. The full-band intelligibility scores were predicted from the half-band intelligibility scores (see Equation

1). The interpolation method used to determine the half-band phoneme intelligibility scores assumed a linear relationship between phoneme intelligibility and the logarithm of frequency between 1414 and 2828 Hz. The data reported in the literature by other researchers, such as French and Steinberg (1947), indicate a curvilinear relationship between phoneme intelligibility and the logarithm of frequency (see discussion above). The predicted full-band phoneme intelligibility scores from this study tended to be somewhat lower than if a curvilinear interpolation was used.

III. Possible learning effects

A concern about the design of this study was the repetition of test materials. Each listener heard a total of 72 word lists throughout testing, and because there are only 20 lists, each list was heard about three times. Although efforts were made to reduce effects of learning, by randomization of all talkers, filtering conditions, and test sessions, the possibility of learning effects should not be excluded. There was evidence of a small but significant increase of mean phoneme intelligibility between the first and second test sessions, suggestive of a learning effect. In addition, in the second testing session, listeners were more familiar with the test procedure, which also may have improved performance. In spite of the evidence of a learning effect, there was no evidence that its magnitude was a function of filtering condition or gender, or that it affected cross-over frequency. The presence of this effect does not, therefore, threaten the validity of findings related to cross-over frequency.

The learning effect did, however, have a small but significant influence on interpolated half-band intelligibility and inferred full-band intelligibility. The absence of interactions justified the collapsing of data across two test sessions. There is no reason to question the validity of the qualitative findings. It must be noted that when generalizing to clinical situations, listeners do not typically have the amount of exposure to either the task or the test materials that they did in the present study. Therefore, in drawing inferences about clinical testing from the present data, it might be more appropriate to use the qualitative results from the first test session only.

IV. Clinical Implications

The issue of clinical application was central to the motivation of the present study. The underlying question was how much influence do inter-talker differences have on the reliability of measures of phoneme intelligibility in hearing-impaired individuals. While the present study did not include hearing-impaired individuals, it did provide data on the performance of adults with normal linguistic knowledge who are presented with the task of interpreting incomplete acoustic signals. Most hearing-impaired individuals suffer a greater loss of acuity in the high frequencies than in the low frequencies. Therefore, the data from this study on low-pass filtering at 707 and 1414 Hz offer the best simulation. For the present purposes, therefore, the data from these two low-pass conditions were collapsed to give a single estimate of phoneme intelligibility. These values are shown in Table 6.1.

Table 6.1: Low-pass filtered percent phoneme recognition for the speech of 6 talkers during two test sessions. The data for a single talker and session are collapsed across 8 listeners and 2 filter cut-off frequencies: 707 and 1414 Hz.

	Replication		
Talker	First	Second	Mean
TF1	64.0	66.9	65.4
TF2	57.5	59.2	58.3
TF3	59.2	61.3	60.2
TM1	63.7	68.3	66.0
TM2	66.0	63.5	64.8
TM3	64.8	72.9	68.9
Mean	62.5	65.3	63.9
St.Dev.			3.9

In Table 6.1, the standard deviation of the mean percent intelligibilities of the six talkers under low-pass filtering was 3.9 percentage points. Using the student's value of 't' = 2.57 for 5 degrees of freedom, the best estimate of 95% confidence limits for the range of intelligibilities in the population of adults represented by this sample of six talkers is +/- 10.0 percentage points around a mean of 64%. It must be noted that some of this range includes the between-listener talker differences and effects of inherent measurement error (the 95% confidence limits for a talker for half-band phoneme intelligibility are +/-1.9 percentage points). Since the latter sources of error are small, it can be assumed, for the moment, that inter-talker differences are the primary determining factor. These numbers suggest that, by picking a random talker, we create an uncertainty in the estimate of the "true[3]" phoneme recognition performance capability of an individual in the region of 10 percentage points. While differences of this magnitude may not be considered serious in terms of diagnosis, they may be considered serious in other applications, such as establishing the effects of changes in amplification characteristics or effects of training.

---

[3] The qualifier "true" is used here to represent the hypothetical mean phoneme intelligibility score for a single listener averaged across all talkers for a large number of replications, in the absence of learning effects.

As talker gender was found to be a significant contributor to inter-talker differences of the frequency-importance functions, its removal would be expected to reduce the confidence limits of scores under a given filtering condition. In other words, picking a talker at random from a single gender group will reduce the uncertainty in "true" phoneme intelligibility. From the data in Table 6.1, we obtain a pooled within-gender estimate of standard deviation of 3.0 with 4 degrees of freedom. Using the student's value of 't' = 2.78 for 4 degrees of freedom, the best estimate of 95% confidence limits for the range of intelligibilities in the population of adults represented by these talkers is +/- 8.3 percentage points around a mean of 64%. Therefore, by picking a talker at random from within a fixed gender group the uncertainty in the estimate of a listener's "true" phoneme intelligibility can be reduced from around 10 percentage points to around 8 percentage points. Although an improvement of this magnitude in the precision with which one can estimate an individual's "true" phoneme recognition skills may be of clinical significance, it is hardly dramatic. If precision cannot be improved by adequately by restricting the choice of talkers to a specific gender, then perhaps more dramatic improvements can be obtained by eliminating all talker differences. This possibility was examined by subjecting the data from Table 6.1 to an analysis of variance with gender as a grouping variable and replication as a repeated measure. The residual error term gives an estimated standard deviation of repeated scores (averaged across two filtering conditions) of 1.36 percentage points. Applying the 't' value of 2.78 for 4 degrees of freedom gives 95% confidence limits for the estimate of "true" phoneme intelligibility of +/- 3.8 percentage points around a mean score of 64%.

Thus, to summarize: ignoring talker differences and picking a talker at random (equivalent to live-voice clinical testing) yields a precision of measurement of phoneme intelligibility of around 10 percentage points for scores in the 64% region. This precision can be improved to around 8 percentage points by picking a talker at random within a single gender group. Use of a single talker (equivalent to standardized recorded materials), however, improves precision to around 4 percentage points. Therefore, there is evidence that complete elimination of talker differences leads to considerable improvements in precision. Also, from a practical viewpoint, use of a single talker can be achieved through the use of recorded test materials. Although the data in this study were obtained with simulated hearing loss, these data support the conclusions drawn from previous studies mentioned in Chapter 1 (Bess, 1983; Brandy, 1966; Gengle & Kupperman, 1980; Hood & Poole, 1980; Kirk, Pisoni, & Miyamoto, 1997; Kruel, Bell, and Nixon, 1969; Penrod, 1979).

It should be pointed out, however, that the 4 percentage point value may be overly optimistic because it assumes scores are collapsed across 32 presentations (eight listeners x two conditions x two replications). This issue is explored further below.

V. Inherent measurement error

It was pointed out in the previous section that the estimates of confidence limits, although based on the means of 32 presentations, will contain a certain amount of inherent measurement error. That error arises from the variability of a single percent phonemes recognized score obtained from a single listener. It has been pointed out previously that this variability can be estimated from the binomial model under the assumption that a single score is based on a set of independent test items (Boothroyd, 1968; Dubno, Lee, Klein, Matthews & Lam, 1995; Gelfand, 1998; Olsen, Van Tasell, & Speaks, 1997; Thornton & Raffin, 1978).

The appropriate formula is:

sd = sqrt(p/100*(1-p/100)/n)*100....................................(3)

where:

sd = predicted standard deviation of repeated scores in percentage points

p = probability of a correct response in percent

n = the number of independent items on which a score is based

For each filter condition in this study a list of ten words was presented. While the number of phonemes in that list was 30, this value should be reduced to around 25 to account for inter-phonemic dependencies (Boothroyd and Nittrouer, 1988). Substituting n = 25 into equation (3), together with a value of 64% for p, gives an estimated standard deviation of 9.6 percentage points. Using a value of 't' = 1.98 gives 95% confidence limits of +/-19 percentage points around a score of 64%. Thus, the measurement error for a single estimate of phoneme intelligibility using a ten word list is about twice that introduced by talker variability. Using scores based on 32 word lists (eight listeners x two replications x two conditions), gives an estimated standard deviation of 9.6/sqrt(32) percentage points, which equals 1.70 percentage points. The 95% confidence limits are +/- 3.4 percentage points. This value of 3.4 percentage points is similar to the 3.8 percentage point estimate obtained in the previous section when all talker variability was removed.

Therefore, the conclusion is that removal of inter-talker differences will dramatically enhance the precision of a test only if other steps are also taken to increase test-retest reliability on a single measure of phoneme intelligibility. One way to achieve this is to increase the number of test items. It has been suggested that the number of test items can be increased without significantly increasing test time by using phoneme based scoring, rather than whole word scoring (Boothroyd, 1968; Gelfand, 1998; Olsen, Van Tasell, & Speaks, 1997). In fact, Gelfand recommends the use of 150 consonant-vowel-consonant words scored phonemically, which yields 450 test items, to obtain a single estimate of phoneme intelligibility.

VI. Additional Practical Concerns

Even if the problem of inherent test-retest reliability could be solved, the use of a standard talker would automatically imply recorded materials. Along with the use of a single talker in phoneme intelligibility testing, several other variables need careful control to improve test reliability. Talker microphone position and distance should remain constant. Also, studies have shown that attempts to make speech clear for the listener, especially by decreasing rate of speech, can effect phoneme intelligibility (Howell and Bonnett,1997; Picheny, Durlach, & Braida,1985; Picheny, Durlach, & Braida, 1986; Picheny, Durlach, & Braida,1989).

VII. Further research

Subsequent research in this area is needed. The small number of talkers used in this study, limited by the test materials used, create the need for a follow-up study using more talkers. Although statistically significant gender differences were seen for cross-over frequency, it is difficult to make conclusions about the magnitude of the average male/female difference with such a limited sample. Also, with a greater number of talkers, the range of cross-over frequencies will be expanded, and it is possible that there would be more overlap between the male and female cross-over frequencies.

There is also a need for a study using male and female talkers of different ages. The differences between acoustic characteristics of adult and child talkers are described in Chapter 2. The use of all adult talkers in this study provides no information about a listener's performance on phoneme intelligibility measures using child talkers.

A study using multiple talkers and listeners with sensorineural hearing loss is also warranted. One approach to this type of study might be to investigate the effect of different talkers on listeners using various sensory aids, such as hearing aids and cochlear implants. It would be interesting to determine the effect of the acoustic input to these sensory aids on the phoneme intelligibility of multiple talkers.

Methods must be found of increasing test-retest reliability without unreasonable increases in test items, and therefore, test time. For example, a performance-intensity function using the same test materials repeatedly may offset the need for additional test materials by presenting the same word lists at ascending sound pressure levels, essentially retaining the novelty of the test materials and thus increasing usable test items. These data can be compared to a function using a descending sound pressure level and a different set of unchanging test materials. A variation on this design may be to increase or decrease amount of filtering rather than intensity.

VIII. Summary and Conclusions

Based on the results of this study, the following conclusions may be drawn:

1. Using these test materials and procedures, the median frequency in the importance function for phoneme intelligibility (i.e. cross-over frequency) lies in the range of 1482 Hertz and 1923 Hz.
2. Half-band phoneme intelligibility lies in the range of 74.9% to 88% - in subjects who have not overlearned the test material.
3. Predicted full-band phoneme intelligibility lies in the range of 96% to 98%.
4. These three measures, cross-over frequency, half-band phoneme intelligibility, and full-band phoneme intelligibility, vary across talkers.
5. Gender contributes significantly to cross-over frequency. In this study, the female talkers, on average, have a cross-over frequency in the region of 13% higher than the male talkers.
6. There was no evidence from this study of a gender contribution to half-band phoneme intelligibility or predicted full-band phoneme intelligibility.
7. There was weak evidence, however, of a gender effect on intelligibility under specific conditions: female talkers tend to be perceived better under high-pass filter conditions, and male talkers tend to be perceived better under low-pass filter conditions.
8. For filter conditions giving phoneme intelligibility in the 50% to 80% range (low-pass 707 and high-pass 2828 Hz), individual talker intelligibility varies over a range of 9-13 percentage points.

9. Elimination of talker differences alone would not have a major effect on confidence limits for an estimate of phoneme intelligibility based on a list of ten words.
10. The first step in improving reliability in clinical measures of phoneme intelligibility is to reduce inherent variability by increasing number of test items on which a score is based.
11. Further benefits may then come from using either a single talker or a carefully chosen "average talker".

APPENDIX A:

AB isophonemic word lists used as test stimuli.
Each list contains ten words consisting of the same
ten vowels and 20 consonants in various combinations

APPENDIX A - AB ISOPHONEMIC WORD LISTS

LIST 1	LIST 4	LIST 7	LIST 10	LIST 13	LIST 16	LIST 19
SHIP	FUN	BADGE	JUG	KISS	WAGE	VASE
RUG	WILL	HUTCH	LATCH	BUZZ	RAG	CAB
FAN	VAT	KILL	WICK	HASH	BEACH	TEACH
CHEEK	SHAPE	THIGHS	FAITH	THIEVE	CHEF	DEATH
HAZE	WREATH	WAVE	SIGN	GATE	DIME	NICE
DICE	HIDE	REAP	BEEP	WIFE	THICK	FIG
BOTH	GUESS	FOAM	HEM	POLE	LOVE	RUSH
WELL	COMB	GOOSE	ROD	WRETCH	ZONE	HOPE
JOT	CHOOSE	NOT	VOTE	DODGE	HOP	LODGE
MOVE	JOB	SHED	SHOES	MOON	SUIT	WOMB
LIST 2	**LIST 5**	**LIST 8**	**LIST 11**	**LIST 14**	**LIST 17**	**LIST 20**
FISH	FIB	BATH	MATH	WISH	JADE	CAVE
DUCK	THATCH	HUM	HIP	DUTCH	CASH	RASH
PATH	SUM	DIG	GUN	JAM	THIEF	TEASE
CHEESE	HEEL	FIVE	RIDE	HEATH	SET	JELL
RACE	WIDE	WAYS	SIEGE	LAZE	WINE	GUIDE
HIVE	RAKE	REACH	VEIL	BIKE	GIVE	PIN
BONE	GOES	JOKE	CHOSE	ROVE	RUB	FUSS
WEDGE	SHOP	NOOSE	SHOOT	PET	HOLE	HOME
LOG	VET	POT	WEB	FOG	CHOP	WATCH
TOMB	JUNE	SHELL	COUGH	SOON	ZOOM	BOOTH
LIST 3	**LIST 6**	**LIST 9**	**LIST 12**	**LIST 15**	**LIST 18**	
THUG	FILL	HUSH	HAVE	HUG	SHAVE	
WITCH	CATCH	GAS	WIG	DISH	JAZZ	
TEAK	THUMB	THIN	BUFF	BAN	THEME	
WRAP	HEAP	FAKE	MICE	RAGE	FETCH	
VICE	WISE	CHIME	TEETH	CHIEF	HEIGHT	
JAIL	RAVE	WEAVE	JAYS	PIES	WIN	
HEN	GOT	JET	POACH	WET	SUCK	
SHOWS	SHOWN	ROB	RULE	COVE	ROBE	
FOOD	BED	DOPE	DEN	LOOSE	DOG	
BOMB	JUICE	LOSE	SHOCK	MOTH	POOL	

APPENDIX B:

Commands from DaDisp worksheet for equalizing the average rms amplitude of the word lists

APPENDIX B: Commands from DaDisp worksheet for equalizing AB words

```
W1-10: readwav(xxx)
W11: Elliptic(2,22050.0,150.0,1.0,50.0,10.0)
W12:Cascade(W1,W11)
W13: Cascade(W2,W11)
W14: Cascade(W3,W11)
W15: Cascade(W4,W11)
W16: Cascade(W5,W11)
W17: Cascade(W6,W11)
W18: Cascade(W7,W11)
W19: Cascade(W8,W11)
W20: Cascade(W9,W11)
W21: Cascade(W10,W11)
W22: Extract(W12,2200,length(W12)-4400)
W23: Extract(W13,2200,length(W13)-4400)
W24: Extract(W14,2200,length(W14)-4400)
W25: Extract(W15,2200,length(W15)-4400)
W26: Extract(W16,2200,length(W16)-4400)
W27: Extract(W17,2200,length(W17)-4400)
W28: Extract(W18,2200,length(W18)-4400)
W29: Extract(W19,2200,length(W19)-4400)
W30: Extract(W20,2200,length(W20)-4400)
W31: Extract(W21,2200,length(W21)-4400)
W32: Concat(W22..w31)
W33: W32+gnormal(length(W32),1/22050)
W34: gseries(stdev(W33))
W35: clip(W12/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W36: Clip(W13/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W37: Clip(W14/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W38: Clip(W15/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W39: Clip(W16/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W40: Clip(W17/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W41: Clip(W18/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W42: Clip(W19/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W43: Clip(W20/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W44: Clip(W21/max(W33)*10^(73/20),-2^15+1000,2^15-1000)
W45:clip(concat(20*log10(transpose(colstdev(Ravel((w33)/max(w34)*10^(70/20),880,1,660))))
,gseries(0)),0,90
W46: delay(extract(Ampdist(w44,.25),2,90*4),1)|setxoffset(0)
W47:Partsum(w45)/sum(w45)*100|sethunits("dB")|setvunits("Percentile")|setxy(0,90,0,100)|setx
tic(5)|setytic(5)
W48:20*log10(transpose(sqrt(colmean(transpose(spectrum(hanning(Ravel(w32,1024,1,512))))))
```

W1-10: reads in each AB word in each 10 word list

W11: sets up high pass filter

W12-21: executes filtering

W22-31: removes100 ms before and after each word

W32: concatenates words

W35-44: equalizes each list's amplitude

W46-48: calculates the amp. distribution, cumulative amp. distribution, and long-term average speech spectrum

APPENDIX C:

Commands from DaDisp worksheet for creating low-pass and high-pass filters

APPENDIX C: Commands from DaDisp worksheet for creating low-pass and high-pass filters

```
W1: readwav(xxx)
W2: Elliptic(1,22050.0,707.0,1.0,100.0,1414.0)
W3: Elliptic(1,22050.0,1414.0,1.0,100.0,2828.0)
W4: Elliptic(1,22050.0,2828.0,1.0,100.0,5656.0)
W5: Elliptic(2,22050.0,1414.0,1.0,100.0,707.0)
W6: Elliptic(2,22050.0,2828.0,1.0,100.0,1414.0)
W7: Elliptic(2,22050.0,5656.0,1.0,100.0,2828.0)
W8: Cascade(W1,W2)
W9: Cascade(W1,W3)
W10: Cascade(W1,W4)
W11: Cascade(W1,W5)
W12: Cascade(W1,W6)
W13: Cascade(W1,W7)
```

W1: reads in AB words

W2-7: sets up low-pass and high-pass filters using different cut-off Hz's

W8-13: executes the low-pass and high-pass filtering

APPENDIX D:

Randomizations used in testing

APPENDIX D: Talker and condition randomizations for 8 listeners, 2 replications

Replication 1													
L1-1	HPH-LPH						L2-1	HPE-LPE					
kHz	TF1	TM1	TM2	TM3	TF3	TF2	kHz	TM3	TM2	TF2	TF3	TF1	TM1
HP2.828	15	20	2	17	1	6	HP.707	4	10	20	15	8	5
HP1.414	11	8	19	20	15	5	HP1.414	20	19	1	17	4	12
HP.707	12	18	10	11	7	9	HP2.828	11	12	2	3	19	16
LP.707	6	9	13	12	3	8	LP2.828	1	18	14	20	15	6
LP1.414	7	17	5	3	16	2	LP1.414	10	11	6	9	3	8
LP2.828	16	1	14	4	13	18	LP.707	9	5	16	13	7	18
L3-1	LPH-HPE						L4-1	HPE-LPH					
kHz	TF3	TM1	TF2	TM2	TM3	TF1	kHz	TF2	TF3	TF1	TM1	TM3	TM2
LP.707	20	8	13	2	9	4	HP.707	3	18	12	20	2	11
LP1.414	14	10	2	11	6	19	HP1.414	18	2	15	14	7	6
LP2.828	6	15	7	10	20	3	HP2.828	6	4	11	16	1	15
HP.707	19	18	1	5	17	13	LP.707	1	19	17	12	19	8
HP1.414	1	16	4	12	14	7	LP1.414	11	6	9	3	5	13
HP2.828	18	3	5	16	15	8	LP2.828	20	7	10	6	8	4
L5-1	LPH-HPH						L6-1	HPE-LPE					
kHz	TF1	TF3	TM3	TF2	TM1	TM2	kHz	TM2	TF1	TM1	TM3	TF2	TF3
LP.707	12	11	8	1	19	4	LP2.828	10	14	5	2	15	13
LP1.414	8	7	6	20	14	7	LP1.414	12	17	8	16	3	9
LP2.828	20	9	19	16	10	2	LP.707	19	11	15	10	1	8
HP2.828	6	1	3	13	9	5	HP.707	16	20	9	4	12	18
HP1.414	2	5	4	12	15	17	HP1.414	1	2	18	19	17	5
HP.707	13	17	14	11	18	3	HP2.828	3	7	6	13	15	4
L7-1	HPH-LPE						L8-1	LPE-HPH					
kHz	TF2	TM3	TM1	TF3	TM2	TF1	kHz	TM1	TF2	TM3	TF3	TF1	TM2
HP2.828	13	14	20	5	6	12	LP2.828	15	1	11	6	14	20
HP1.414	18	4	17	9	3	11	LP1.414	12	4	7	15	18	10
HP.707	7	19	10	16	18	8	LP.707	6	14	9	1	7	19
LP2.828	6	16	12	19	10	13	HP2.828	16	13	10	8	17	15
LP1.414	8	3	1	20	15	2	HP1.414	18	19	3	2	4	13
LP.707	11	15	2	4	14	7	HP.707	17	20	5	12	9	3

CODES:
HPH=high-pass filtering starting with 2828 cut-off Hz
HPE=high-pass filtering starting 707 cut-off Hz
LPH=low-pass filtering starting with 707 cut-off Hz
LPE=low-pass filtering starting with 2828 cut-off Hz
TF1-3=female speakers 1-3
TM1-3=male speakers 1-3
L1-1=listener 1, replication 1
L1-2=listener 1, replication 2, and so on...

APPENDIX D: Talker and condition randomizations for 8 listeners, 2 replications

Replication 2													
L1-2	LPE-HPE						L2-2	LPH-HPH					
kHz	TF3	TM3	TF1	TM2	TF2	TM1	kHz	TF1	TM1	TF3	TM2	TF2	TM3
LP2.828	4	16	18	12	10	17	LP.707	14	9	10	8	13	5
LP1.414	13	6	14	1	2	16	LP1.414	4	6	13	17	15	19
LP.707	9	19	17	7	15	20	LP2.828	18	4	17	20	14	11
HP.707	15	10	1	4	19	14	HP2.828	3	11	15	2	10	12
HP1.414	7	2	11	18	8	3	HP1.414	16	5	9	3	18	7
HP2.828	5	12	20	3	13	11	HP.707	7	12	19	1	8	16
L3-2	HPH-LPE						L4-2	LPE-HPH					
kHz	TM1	TF2	TM3	TF1	TM2	TF3	kHz	TM3	TF2	TM1	TM2	TF3	TF1
HP2.828	14	18	9	5	15	10	LP2.828	14	3	1	2	9	15
HP1.414	20	8	12	1	14	17	LP1.414	2	5	13	20	11	6
HP.707	2	7	17	16	12	13	LP.707	13	17	3	19	8	1
LP2.828	15	16	11	9	4	3	HP2.828	7	9	20	4	5	17
LP1.414	1	5	4	13	20	9	HP1.414	4	10	18	14	12	16
LP.707	19	12	6	10	3	11	HP.707	8	16	19	15	10	7
L5-2	HPE-LPE						L6-2	HPH-LPH					
kHz	TF2	TM3	TM2	TM1	TF3	TF1	kHz	TF3	TM2	TF2	TF1	TM3	TM1
HP.707	16	1	18	19	17	6	HP2.828	15	20	4	16	13	5
HP1.414	2	14	7	9	19	10	HP1.414	12	10	14	2	1	11
HP2.828	13	11	3	14	20	18	HP.707	9	19	7	4	3	13
LP2.828	17	8	15	13	11	1	LP.707	10	11	8	20	14	19
LP1.414	9	20	6	4	15	5	LP1.414	18	12	3	13	6	8
LP.707	4	10	12	3	16	7	LP2.828	1	9	5	15	17	7
L7-2	LPH-HPE						L8-2	HPE-LPH					
kHz	TM1	TF1	TM2	TF2	TF3	TM3	kHz	TM2	TF3	TF1	TM3	TM1	TF2
LP.707	17	4	12	15	5	6	HP.707	1	13	18	11	2	13
LP1.414	19	11	18	2	4	10	HP1.414	17	19	20	4	16	7
LP2.828	5	10	14	20	7	8	HP2.828	8	6	1	9	5	12
HP.707	1	20	13	14	19	2	LP.707	12	5	14	15	4	18
HP1.414	3	8	15	1	20	13	LP1.414	20	2	16	5	3	6
HP2.828	15	17	3	11	9	16	LP2.828	11	18	19	2	17	13

<u>CODES:</u>
HPH=high-pass filtering starting with 2828 cut-off Hz
HPE=high-pass filtering starting 707 cut-off Hz
LPH=low-pass filtering starting with 707 cut-off Hz
LPE=low-pass filtering starting with 2828 cut-off Hz
TF1-3=female speakers 1-3
TM1-3=male speakers 1-3
L1-1=listener 1, replication 1
L1-2=listener 1, replication 2, and so on...

APPENDIX E1:

Raw data showing percent phonemes recognized by six talkers under three high-pass and three low-pass filter conditions for eight listeners for the first of two replications.

Table E1. Raw data - Percent phonemes recognized for 6 talkers under 6 filter conditions for 8 listeners for Replication 1

Replication 1 - Female talkers

Listener	Talker and Filter condition																	
	F1H28	F1H14	F1H70	F1L28	F1L14	F1L70	F2H28	F2H14	F2H70	F2L28	F2L14	F2L70	F3H28	F3H14	F3H70	F3L28	F3L14	F3L70
L1	46.7	96.7	96.7	90.0	66.7	50.0	63.3	90.0	96.7	100.0	73.3	53.3	63.3	96.7	93.3	96.7	63.3	53.3
L2	46.7	93.3	100.0	100.0	80.0	46.7	30.0	83.3	96.7	100.0	53.3	40.0	40.0	90.0	93.3	100.0	56.7	56.7
L3	83.3	100.0	100.0	100.0	80.0	56.7	53.3	93.3	93.3	96.7	80.0	40.0	53.3	96.7	100.0	93.3	63.3	46.7
L4	43.3	90.0	100.0	96.7	70.0	56.7	40.0	73.3	100.0	100.0	66.7	43.3	40.0	93.3	96.7	93.3	53.3	46.7
L5	53.3	100.0	96.7	100.0	73.3	40.0	56.7	83.3	100.0	100.0	76.7	40.0	43.3	90.0	100.0	96.7	80.0	46.7
L6	73.3	93.3	100.0	96.7	86.7	36.7	60.0	86.7	100.0	100.0	80.0	23.3	80.0	93.3	100.0	93.3	63.3	50.0
L7	80.0	100.0	100.0	100.0	90.0	63.3	66.7	80.0	100.0	100.0	80.0	40.0	66.7	93.3	96.7	96.7	86.7	60.0
L8	80.0	100.0	96.7	93.3	80.0	46.7	70.0	86.7	96.7	100.0	66.7	63.3	83.3	93.3	100.0	100.0	70.0	50.0
m	63.3	96.7	98.8	97.1	78.3	49.6	55.0	84.6	97.9	99.6	72.1	42.9	58.7	93.3	97.5	96.2	67.1	51.3
sd	17.4	4.0	1.7	3.8	8.0	9.0	13.7	6.2	2.5	1.2	9.4	11.6	17.4	2.5	3.0	2.8	11.3	5.0
se	6.1	1.4	0.6	1.3	2.8	3.2	4.8	2.2	0.9	0.4	3.3	22.4	6.1	0.9	1.0	1.0	4.0	1.8

Replication 1 - Male talkers

Listener	Talker and Filter condition																	
	M1H28	M1H14	M1H70	M1L28	M1L14	M1L70	M2H28	M2H14	M2H70	M2L28	M2L14	M2L70	M3H28	M3H14	M3H70	M3L28	M3L14	M3L70
L1	66.7	90.0	100.0	100.0	66.7	43.3	40.0	90.0	100.0	93.3	73.3	46.7	73.3	93.3	100.0	96.7	90.0	43.3
L2	43.3	86.7	96.7	93.3	83.3	36.7	46.7	76.7	93.3	100.0	83.3	36.7	40.0	80.0	90.0	100.0	76.7	33.3
L3	63.3	96.7	96.7	100.0	83.3	53.3	66.7	96.7	96.7	100.0	90.0	70.0	73.3	90.0	96.7	96.7	86.7	53.3
L4	40.0	76.7	100.0	100.0	90.0	30.0	50.0	86.7	100.0	93.3	63.3	46.7	43.3	96.7	96.7	96.7	73.3	73.3
L5	40.0	83.3	96.7	90.0	93.3	76.7	50.0	83.3	93.3	100.0	73.3	63.3	40.0	80.0	90.0	100.0	66.7	43.3
L6	60.0	90.0	100.0	100.0	53.3	60.0	43.3	90.0	93.3	100.0	60.0	50.0	50.0	90.0	96.7	93.3	80.0	56.7
L7	63.3	90.0	93.3	100.0	90.0	53.3	63.3	96.7	100.0	100.0	73.3	66.7	60.0	96.7	93.3	90.0	90.0	46.7
L8	46.7	86.7	100.0	96.7	63.3	43.3	63.3	86.7	96.7	96.7	90.0	70.0	66.7	86.7	86.7	96.7	86.7	36.7
m	52.9	87.5	97.9	97.5	77.9	49.6	52.9	88.3	96.7	97.9	75.8	56.3	55.8	89.2	93.8	96.3	81.3	48.3
sd	11.5	5.8	2.5	3.9	14.8	14.6	10.1	6.7	3.1	3.1	11.2	12.8	14.3	6.6	4.5	3.3	8.5	12.7
se	4.1	2.1	0.9	1.4	5.2	5.2	3.6	2.4	1.1	1.1	4.0	4.5	5.1	2.3	1.6	1.2	3.0	4.5

CODES: **M1H28** = Male talker 1, high-pass cut-off 2828 Hz, **M2H28** = Male talker 2, high-pass cut-off 2828 Hz, and so on...

APPENDIX E2:

Raw data showing percent phonemes recognized by six talkers under three high-pass and three low-pass filter conditions for eight listeners for the second of two replications.

Table E2. Raw data - Percent phonemes recognized for 6 talkers under 6 filter conditions for 8 listeners for Replication 2

Replication 2 - Female talkers

Listener	Talker and Filter condition																	
	F1H28	F1H14	F1H70	F1L28	F1L14	F1L70	F2H28	F2H14	F2H70	F2L28	F2L14	F2L70	F3H28	F3H14	F3H70	F3L28	F3L14	F3L70
L1	66.7	100.0	100.0	100.0	80.0	43.3	73.3	93.3	96.7	100.0	73.3	40.0	53.3	93.3	96.7	86.7	73.3	30.0
L2	46.7	100.0	100.0	100.0	73.3	36.7	43.3	86.7	100.0	96.7	70.0	40.0	56.7	93.3	100.0	100.0	80.0	46.7
L3	76.7	96.7	100.0	96.7	90.0	80.0	56.7	93.3	93.3	100.0	70.0	43.3	66.7	96.7	100.0	100.0	93.3	60.0
L4	53.3	96.7	100.0	100.0	73.3	63.3	50.0	100.0	100.0	100.0	63.3	60.0	53.3	90.0	93.3	83.3	70.0	46.7
L5	90.0	100.0	100.0	100.0	76.7	43.3	50.0	96.7	96.7	100.0	70.0	43.3	63.3	93.3	100.0	100.0	76.7	53.3
L6	83.3	93.3	100.0	96.7	83.3	40.0	63.3	90.0	100.0	96.7	73.3	50.0	60.0	96.7	90.0	80.0	73.3	30.0
L7	90.0	93.3	100.0	100.0	83.3	56.7	76.7	90.0	96.7	100.0	76.7	40.0	86.7	93.3	96.7	93.3	76.7	36.7
L8	76.7	100.0	100.0	100.0	86.7	60.0	73.3	93.3	100.0	93.3	76.7	56.7	80.0	100.0	100.0	100.0	80.0	53.3
m	72.9	97.5	100.0	99.2	80.8	52.9	60.8	92.9	97.9	98.3	71.7	46.7	65.0	94.6	97.1	92.9	77.9	44.6
sd	16.2	3.0	0.0	1.5	6.1	14.7	12.7	4.2	2.5	2.5	4.4	8.0	12.3	3.1	3.8	8.4	7.1	11.3
se	5.7	1.0	0.0	0.5	2.2	5.2	4.5	1.5	0.9	0.9	1.5	2.8	4.4	1.1	1.3	3.0	2.5	4.0

Replication 2 - Male talkers

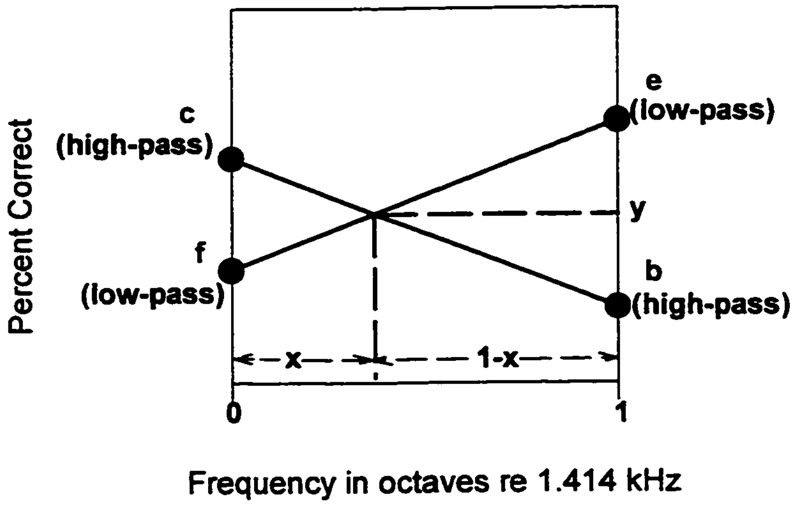
Listener	Talker and Filter condition																	
	M1H28	M1H14	M1H70	M1L28	M1L14	M1L70	M2H28	M2H14	M2H70	M2L28	M2L14	M2L70	M3H28	M3H14	M3H70	M3L28	M3L14	M3L70
L1	56.7	86.7	93.3	100.0	90.0	60.0	66.7	83.3	96.7	100.0	66.7	50.0	63.3	86.7	100.0	100.0	73.3	50.0
L2	40.0	73.3	100.0	100.0	76.7	33.3	63.3	90.0	100.0	96.7	90.0	43.3	50.0	100.0	93.3	100.0	90.0	46.7
L3	63.3	90.0	100.0	100.0	96.7	66.7	76.7	100.0	100.0	96.7	86.7	80.0	53.3	100.0	96.7	100.0	93.3	60.0
L4	60.0	86.7	96.7	100.0	70.0	60.0	53.3	93.3	100.0	100.0	66.7	53.3	40.0	73.3	100.0	93.3	90.0	66.7
L5	70.0	93.3	100.0	100.0	90.0	46.7	60.0	90.0	100.0	100.0	70.0	50.0	53.3	96.7	90.0	96.7	86.7	50.0
L6	53.3	93.3	100.0	80.0	90.0	46.7	73.3	90.0	96.7	100.0	73.3	36.7	63.3	96.7	100.0	100.0	93.3	50.0
L7	66.7	86.7	100.0	100.0	66.7	70.0	70.0	100.0	100.0	96.7	86.7	40.0	63.3	96.7	100.0	90.0	100.0	56.7
L8	56.7	90.0	100.0	100.0	83.3	46.7	63.3	80.0	100.0	100.0	73.3	50.0	66.7	93.3	96.7	96.7	86.7	73.3
m	58.3	87.5	98.8	97.5	82.9	53.8	65.8	90.8	99.2	98.8	76.7	50.4	56.7	92.9	97.1	97.1	89.2	56.7
sd	9.3	6.4	2.5	7.1	10.8	12.4	7.5	7.1	1.5	1.7	9.6	13.3	9.1	9.0	3.8	3.8	7.7	9.4
se	3.3	2.2	0.9	2.5	3.8	4.4	2.7	2.5	0.5	0.6	3.4	4.7	3.2	3.2	1.3	1.3	2.7	3.3

CODES: M1H28 = Male talker 1, high-pass cut-off 2828 Hz, **M2H28** = Male talker 2, high-pass cut-off 2828 Hz, and so on...

APPENDIX F:

Derivation of the formula used to determine linear and logarithmic cross-over frequency

Appendix F: Derivation of the formula used to calculate the linear and logarithmic cross-over frequencies



$$(e-f)/1 = (y-f)/x \tag{1}$$

$$(c-b)/1 = (y-b)/(1-x) \tag{2}$$

$$y = x(e-f) + f \tag{3}$$

$$y = (1-x)(c-b) + b \tag{4}$$

$$x(e-f)+f = (1-x)(c-b)+b \tag{5}$$

$$x(e-f) + f = c-b - x(c-b) + b \tag{6}$$

$$x(e-f) + (c-b) = (c-b) + (b-f) \tag{7}$$

$$x = (c-b) + (b-f)/(e-f) + (c-b) \quad \text{(in octaves)} \tag{8}$$

$$x = c-f/(e-f) + (c-b) \tag{9}$$

$$F_x = 1.414 * 2^\wedge x \tag{10}$$

APPENDIX G:

DaDisp worksheet used for determining fundamental and formant frequency values for AB words

APPENDIX G: Commands from DaDisp worksheet for determining fundamental and formant frequency values

```
W1: Readwav("c:\bam\f31010eq.wav")                        W1: imports equalized sound files
W2: Extract(w1,256+(curpos(w10)-1)*64-512,1024)           W2: extracts temp. center of vowel
W3: Spectrum(hanning(W2))                                 W3: displays frequency spectrum
W4: 20*log10(W3)                                          W4,6,7,8: converts to dB scale
W5: Extract(movavg(w3,15),8,length(W3))|setxoffset(0)W5: smoothes spectrum
W6: 20*log10(w5)
W7: 20*log10(W5)                                          W9: creates spectrum of 21 Hz BW
W8: 20*log10(W5)                                          W10: converts y-axis to dB scale
W9: Gseries(curpos(w4)-1,curpos(w6)-1,curpos(w7)-1,curpos(w8)-1)*11025/512
W10: Transpose(20*log10(1+spectrum(hanning(Ravel(W1,512,1,448)))))
```

APPENDIX H:

DaDisp worksheet used for calculating long-term average speech spectrum

Appendix G: Commands from DaDisp worksheet calculating the long-term average speech spectrum

```
W1: spectq.6.01
W2: spectq.6.02
W3: spectq.6.03
W4: spectq.6.04
W5: spectq.6.05
W6: spectq.6.06
W7: spectq.6.07
W8: spectq.6.08
W9: spectq.6.09
W10: spectq.6.10
W11: spectq.6.11
W12: spectq.6.12
W13: spectq.6.13
W14: spectq.6.14
W15: spectq.6.15
W16: spectq.6.16
W17: spectq.6.17
W18: spectq.6.18
W19: spectq.6.19
W20: spectq.6.20
W21: Sums(w1..w20)/20|setytic(5)
W22: W21|setxy(100,10000,15,65)
```

W1-W20: Imports long-term average speech spectrum for each of the 20 AB word lists

W21: Calculates average of 20 lists

W22: Graphs long-term average

REFERENCES

Bess, F.H. (1983). Clinical assessment of speech recognition. In Konkle, D.F., Rintelmann, W.F.,(eds.): *Principles of Speech Audiometry*. Baltimore: University Park Press, 127-201.

Bilger, R.C. and Wang, M.D. (1976). Consonant confusions in patients with sensorineural hearing loss. *Journal of Speech and Hearing Research*, 19, 718-48.

Boothroyd, A. (1967). The discrimination by partially hearing children of frequency distorted speech. *International Audiology*, 6(2), 136-45.

Boothroyd, A. (1968a). Developments in speech audiometry. *Sound*, 2, 3-10.

Boothroyd, A. (1968b.) Statistical theory of the speech discrimination score. *Journal of the Acoustical Society of America*, 43, 362-367.

Boothroyd, A. (1984). Auditory perception of speech contrasts by subjects with hearing sensorineural loss. *Journal of Speech and Hearing Research*,27, 134-44.

Boothroyd, A.; Erickson, F.N.; and Medwetsky, L.(1994). The hearing aid input: a phonemicapproach to assessing spectral distribution of speech. *Ear and Hearing* 15(6), 432-42.

Boothroyd, A. and Medwetsky, L.(1992). Spectral distribution of /s/ and the frequency response of hearing aids. *Ear and Hearing*, 13(3), 150-157.

Boothroyd, A. and Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America*, 84, 101-114.

Brandy, W.T. (1966). Reliability of voice tests in speech discrimination. *Journal of Speech and Hearing Research*, 9, 461-465.

Coleman, R.O.(1976). A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research*, 19, 168-180.

Coleman, R.O.(1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research*, 14, 565-577.

Cox, R.M. and McDaniel, D.M.(1984). Intelligibility ratings of continuous discourse: Application to hearing aid selection. *Journal of the Acoustical Society of America*, 76(3), 758-766.

Curry, E.T.(1940). The pitch characteristics of the adolescent male voice. *Speech Monographs*, 7, 48-62.

Diehl, R.L.; Lindblom, B.; Hoemeke, K.A., and Fahey, R.P.(1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, 24, 187-208.

Dubno, J.R.; Dirks, D.D.; and Ellison, D.E. (1988). Stop consonant recognition for normal listeners and listeners with high-frequency hearing loss. I: The contribution of selected frequency regions. *Journal of the Acoustical Society of America*, 85(1), 347-354.

Dubno, J.R.; Lee, F.; Klein, A.J.; Matthews, L.J.; and Lam, C.F. (1995). Confidence limits for maximum word recognition scores. *Journal of Speech and Hearing Research*, 38, 490-502.

Eguchi, S. and Hirsh, I.J.(1969). Development of speech sounds in children. *Acta Otolaryngologica Supplement*, 257.

Eklund, I.; Traunmuller, H. (1997). Comparative study of male and female whispered and phonated versions of long vowels of Swedish. *Phonetica,* 54, 1-21.

Fairbanks, G. (1960). Voice and articulation drillbook, 2nd ed. (New York), 127.

Fairbanks, G.; Herbert, E.L.; and Hammond, J.M.(1949). An acoustical study of vocal pitch in seven- and eight-year old girls. *Child Development*, 20(2), 71-78.

Fairbanks, G.; Wiley, J.H.; and Lassman, F.M.(1949). An acoustical study of vocal pitch in seven- and eight-year old boys. *Child Development*, 20(2), 63-69.

French, N.R. and Steinberg, J.C.(1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America*, 19(1), Jan., 90-119.

Gelfand, S.A. (1998). Optimizing the reliability of speech recognition scores. *Journal of Speech-Language-Hearing Research*, 41, 1088-1102.

Gengel, R. and Kupperman, G.L. (1980). Word discrimination in noise: effect of different speakers. *Ear and Hearing*, 1, 156-60.

Hillenbrand, J.M., Getty, L.A., Clark, M.J. and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, (5), Pt. 1, May, 3099-3111.

Hirsh, I.J.; Reynolds, E.G.; and Joseph, M. (1954). Intelligibility of different speech materials. *Journal of the Acoustical Society of America*, 26(4), July, 530-38.

Hollien, H. and Paul, P.(1969). A second evaluation of the speaking fundamental frequency of post-adolescent girls. *Language and Speech*, 12, 119-124.

Hollien, H. and Shipp, T.(1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research*, 15, 155-159.

Hood, J.D. and Poole, J.P. (1980). Influence of the speaker and other factors affecting speech intelligibility. *Audiology*, 19, 434-455.

Howell, P. and Bonnett, C. (1997). Speaking clearly for the hearing impaired: intelligibility differences between clear and less clear speakers. *European Journal of Disorders of Communication*, 32, 89-97.

Klatt, D.H. and Klatt, L.C.(1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87(2), 820-857.

Kirk, K.I.; Pisoni, D.B.; and Miyamoto, R.C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech-Language-Hearing Research*, 40, 1395-1405.

Kruel, E.J.; Bell, D.W.; Nixon, J.C. (1969). Factors affecting speech discrimination test difficulty. *Journal of Speech and Hearing Research*, 12, 281-287.

Kryter, K.(1962a). Methods for the calculation and use of the Articulation Index. *Journal of the Acoustical Society of America*, 34(11), Nov., 1689-1697.

Kryter, K.(1962b). Validation of the Articulation Index. *Journal of the Acoustical Society of America*, 34(11), Nov., 1698-1702.

Levitt, H. and Webster, J.C. (1991). Effects of Noise and Reverberation on Speech. from Harris,C. *Handbook of Acoustical Measurements and Noise Control*, 3rd edition. New York: McGraw-Hill.

Linke, E. (1953). A study of pitch characteristics of female voices and their relationship to vocal effectiveness. Unpublished doctoral dissertation, State University of Iowa.

Margulies, M. (1980). Effects of talker differences on speech intelligibility in the hearing impaired. Unpublished doctoral dissertation, City University of New York.

McGlone,R.E. and Hollien, H.(1963).Vocal pitch characteristics of aged women. *Journal of Speech and Hearing Research*,6 (2), 164-70.

McGowan, R. and Nittrouer, S. (1988). Differences in fricative production between children and adults: evidence from acoustic analysis of /sh/ and /s/. *Journal of the Acoustical Society of America A*, 83, 229-236.

Michel, J.F.; Hollien, H., and Moore, P.(1966). Speaking fundamental frequency characteristics of 15,16, and 17 year old girls. *Language and Speech*, 9, 46-51.

Miller, G.A. and Nicely, P.E.(1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27(2), 338-352.

Milner, P.; Braida, L.D.; Durlach, N.I.; and Levitt, H.(1984). Perception of filtered speech by hearing impaired listeners. Appendix to L.D. Braida: "Articulation testing methods for evaluating speech reception by hearing impaired listeners" in *Speech recognition by the hearing impaired*, edited by E.Elkins, ASHA reports 14, 30-41.

Monsen, R.B. and Engebretson, A.M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America*, 62(4), October, 981-993.

Mysak, E.D.(1959). Pitch and duration characteristics of older males. *Journal of Speech and Hearing Research*, 2, 46-59.

Nittrouer, S. (1992). Age-related differences in perceptual effects on formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20, 351-382

Nittrouer, S.; Studdert-Kennedy, M; and McGowan, R.S. (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 32, March, 120-32.

Olsen, W.O.; Van Tasell, D.J.; and Speaks, C.E. (1997). Phoneme and word recognition for words in isolation and in sentences. *Ear and Hearing*, 18, 175-188.

Owens, E.; Talbott, C.B.; and Schubert, E.D. (1968). Vowel discrimination of hearing-impaired listeners. *Journal of Speech and Hearing Research*, 11, 648-655.

Pavlovic, C. and Studebaker, G.A.(1984). An evaluation of some assumptions underlying the articulation index. *Journal of the Acoustical Society of America*, 75(5), 1606-1612.

Penrod, J. (1979). Talker effects on word discrimination scores of adults with sensorineural hearing impairment. *Journal of Speech and Hearing Disorders*, 44, 340-349.

Peterson, G.E. and Barney, H.L.(1952). Control methods used in the study of the vowels. *Journal of the Acoustical Society of America*, 24(2), March, 175-184.

Picheny, M.A.; Durlach, N.I.; and Braida, L.D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28, 96-103.

Picheny, M.A.; Durlach, N.I.; and Braida, L.D. (1986). Speaking clearly for the hard of hearing II: Acoustic correlates of clear and conversational speech. *Journal of Speech and Hearing Research*, 29, 434-446.

Picheny, M.A.; Durlach, N.I.; and Braida, L.D. (1989). Speaking clearly for the hard of hearing III: an attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, 32, 600-603.

Pollack, I. and Pickett, J.M.(1964). Frequency importance functions for isolated words and for conversation of female talkers. *Language and Speech*, 7, 71-75.

Raffin, M.J.M. and Thornton, A. (1980). Confidence levels for differences between speech discrimination scores. *Journal of Speech and Hearing Research*, 23, 5-18.

Sher, A.E. and Owens, E. (1974). Consonant confusions associated with hearing loss above 2000 Hz. *Journal of Speech and Hearing Research*, 17, 669-81.

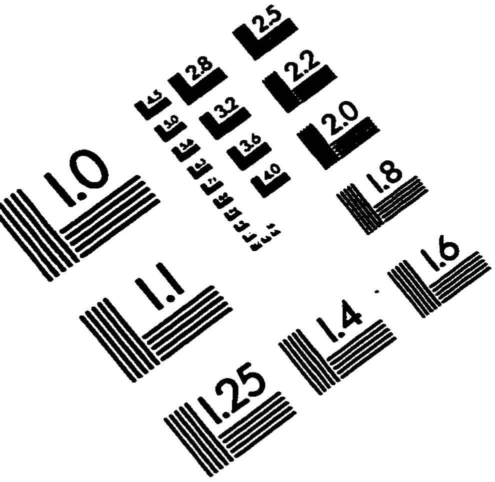
Speaks, C. (1967). Intelligibility of filtered synthetic sentences. *Journal of Speech and Hearing Research*, 10, 289-98.

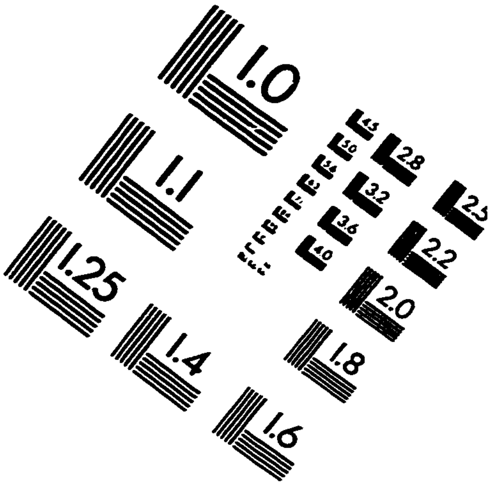
Studebaker, G.A.; Pavlovic, C.V.; and Sherbecoe, R.L. (1987). A frequency importance function for continuous discourse. *Journal of the Acoustical Society of America*, 81(4), April, 1130-1138.

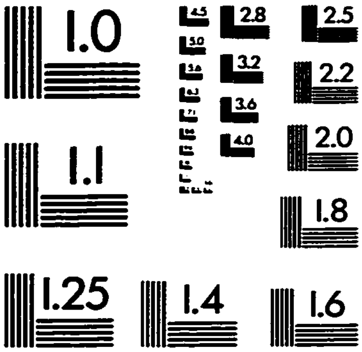
Thornton, A.R. and Raffin, M.J.M. (1978). Speech-discrimination scores modeled as a binomial variable. *Journal of Speech and Hearing Research*, 21, 507-518.

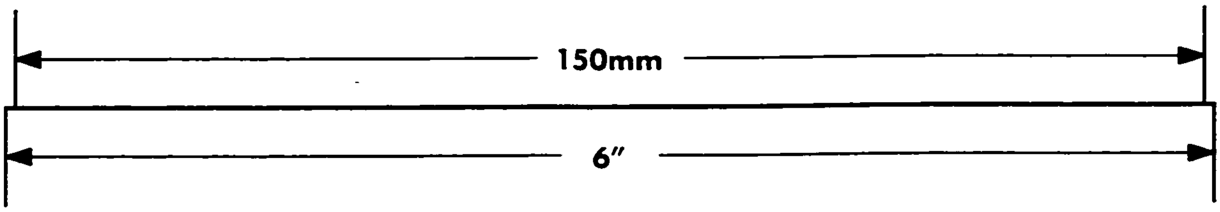
Wang, M.D.; Reed, C.M.; and Bilger, R.C. (1978). A comparison of the effects of filtering and sensorineural hearing loss on patterns of consonant confusions. *Journal of Speech and Hearing Research* 21, 5-36.

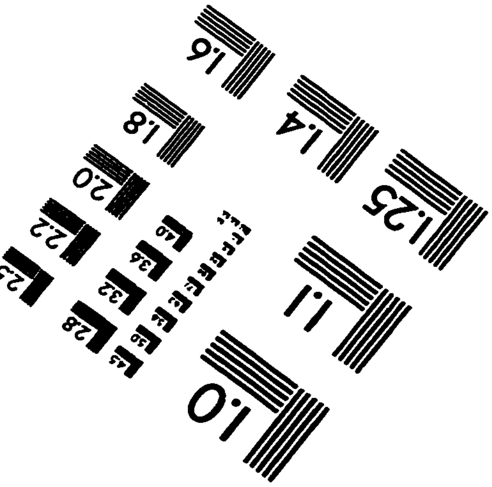
IMAGE EVALUATION
TEST TARGET (QA-3)

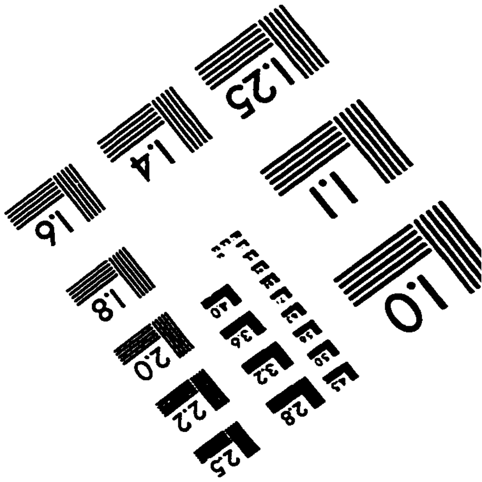












APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989