

## **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.**

**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.**

**Bell & Howell Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600**

**UMI<sup>®</sup>**



GENDER EFFECTS ON THE RESISTANCE OF SPEECH TO SPECTRAL  
SMEARING

By

CHUNG-I ANGELA LI

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

2000

**UMI Number: 9986352**

**Copyright 2000 by  
Li, Chung-I Angela**

**All rights reserved.**

**UMI<sup>®</sup>**

---

**UMI Microform 9986352**

**Copyright 2000 by Bell & Howell Information and Learning Company.  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.**

---

**Bell & Howell Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346**

Copyright 2000

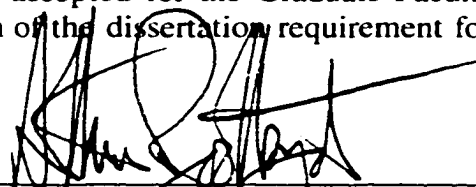
**CHUNG-I ANGELA LI**

All Right Reserved

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

9-15-00

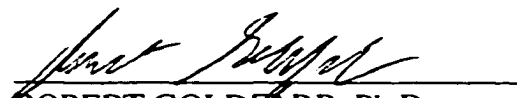
Date



ARTHUR BOOTHROYD, Ph.D.  
Chair of Examining Committee

9/15/00

Date



ROBERT GOLDFARB, Ph.D.  
Executive Officer

LAWRENCE RAPHAEL, Ph.D.  
WINIFRED STRANGE Ph. D.  
Examining Committee

The City University of New York

## ABSTRACT

GENDER EFFECTS ON THE RESISTANCE OF SPEECH  
TO SPECTRAL SMEARING

By

CHUNG-I ANGELA LI

Advisor: Professor Arthur Boothroyd

It is generally believed that most of the difficulties of speech perception accompanying sensorineural hearing loss are caused by a loss of frequency resolution. This belief is supported by experiments with normally hearing listeners in which poor frequency resolution is simulated by artificial smearing of the spectral envelope of speech.

The purpose of the present study was to test the prediction that women's speech is more resistant to the negative effects of spectral smearing than is men's speech. This prediction was based on the hypothesis that spectral smearing degrades speech perception by obscuring the relationships among formant frequencies.

As predicted, scores for women's speech were significantly higher than for men's speech. The smeared bandwidth producing 50% phoneme recognition averaged 2300 Hz for the female talkers and 1800 Hz for the male talkers. The difference of 0.35 octaves (or 28%) was close to that predicted on the basis of gender differences of average formant spacing.

Acoustical measures of the unprocessed speech confirmed that the formant frequencies in the speech of this sample of talkers were in keeping both with that reported in the literature and with predictions based on known anatomical gender differences. The findings support the hypothesis that spectral smearing degrades speech

perception by obscuring formant patterns. It was found, however, that the frequency of the second vocal tract formant was a somewhat better predictor of the effects of smearing than any measure of formant spacing.

The smearing used in the present study was of constant bandwidth, regardless of frequency. When measured in terms of octaves, however, the amount of smearing fell with increasing frequency. It is possible, therefore, that the present findings reflect the importance of the second formant in phoneme recognition and the fact that the average frequency of the second formant in women's speech is higher than in men's speech. Before this work can be applied to speech processing in hearing aids, further studies will require more realistic simulations of the effects of sensorineural hearing loss and the control of gender differences other than those of formant frequency.

## ACKNOWLEDGEMENTS

I am grateful to Veterans General Hospital-Taipei for providing me the opportunity to pursue the advanced study in the area that has just begun to bloom in Taiwan.

My courage and perseverance to complete the study came from my husband, my daughter and son, their love and understanding kept me moving forward. My special thanks go to the most important person, my mother-in-law, without her support in caring my children, I would not be able to do the task.

Thanks go to Jodi, Nina, Bethany, Tom, Chris, Frank, and Dorothy for their kind help in preparation for the recordings of the talkers. Thanks also go to my listeners who participated in these experiments.

Appreciation is extended to Dr. Harry Levitt who has brought me into the program, Dr. Hwei-bing Lin who has shared my joy and stress that I have gone through during my stay in the program. Regretfully, I could never have a chance to say thanks to late executive officer, Dr. Irving Hochberg, whose humor and kindness had kept me returning to the program every semester after a home visit.

Particularly, I want to express my deep appreciation to Drs. Winifred Strange, Lawrence Raphael, and Robert Goldfarb for their valuable insights, comments, and suggestions that secured my dissertation to a nearly impeccable manner.

Finally, my immeasurable gratitude is indebted to my mentor, Dr. Arthur Boothroyd, for his guidance and full support to work on my weaknesses throughout the stage of dissertation and study at the Graduate Center, and his dedication to research is a model for me to follow.

## TABLE OF CONTENTS

CHAPTER .....	PAGE
1. INTRODUCTION AND PURPOSE .....	1
1.1 Statement of purpose.....	1
1.2 Spectral smearing.....	1
1.3 Rationale for present study .....	2
1.4 Design of the study .....	3
1.5 Summary .....	4
2. REVIEW OF THE LITERATURE .....	6
2.1 Background .....	6
2.1.1 Hearing loss and speech perception .....	6
2.1.2 Threshold elevation.....	6
2.1.3 Loss of auditory resolution .....	6
2.1.4 Frequency resolution.....	7
2.1.5.Spectral aspects of speech.....	8
2.1.6 Establishing the connection between frequency resolution and speech perception.....	9
2.2 Studies of spectral smearing .....	10
2.2.1 Villchur .....	10
2.2.2 Gagné and Erber.....	11
2.2.3 Celmer and Bienvenue .....	11
2.2.4 Ter Keurs, Festen, and Plomp.....	13
2.2.5 Baer and Moore.....	13
2.2.6 Howard-Jones and Summers.....	15
2.2.7 Summers et al.....	15
2.2.8 Shannon et al.....	16
2.2.9 Boothroyd et al.....	16
2.3 Summary .....	18
2.4 Rationale for present study .....	19
3. METHOD.....	20
3.1 Outline.....	20
3.2 Subjects .....	20
3.2.1 Talkers .....	20
3.2.2 Listeners .....	21
3.3 Test Materials.....	21
3.4 Preparation of recordings .....	22
3.5 Processing of stimuli .....	22
3.5.1 Digitization .....	22
3.5.2 Spectral smearing.....	23
3.6 Design .....	23
3.6.1 Dependent and independent variables.....	23
3.6.2 Counterbalancing of lists .....	24
3.6.3 Presentation order of smearing conditions.....	24

3.7 Procedure .....	25
3.7.1 Replication .....	25
3.7.2 Test stimuli presentation .....	25
3.7.3 Scoring .....	25
3.8 Response judgements .....	26
3.9 Spectrographic analysis .....	26
<b>4. RESULTS OF SMEARING STUDY .....</b>	<b>28</b>
4.1 Raw data .....	28
4.2 Analysis of variance .....	28
4.3 Replication .....	29
4.4 Main effect of gender .....	30
4.5 Effect of phoneme position .....	31
4.5.1 Main effect of phoneme position .....	31
4.5.2 Interaction with gender .....	31
4.5.3 Interaction with smeared bandwidth .....	32
4.6. Effect of smearing .....	33
4.6.1 Main effect of smeared bandwidth .....	33
4.6.2 Smearing threshold and gender .....	34
4.7 Summary .....	34
<b>5. ACOUSTIC ANALYSIS OF THE SPEECH STIMULI .....</b>	<b>36</b>
5.1 Purpose .....	36
5.2 Method .....	36
5.2.1 Selection of vowels .....	36
5.2.2 Procedure .....	36
5.3 Results .....	37
5.3.1 Raw data .....	37
5.3.2 Individual and group mean formant frequencies .....	37
5.3.3 The main effect of gender .....	39
5.3.4 Formant spacing .....	40
5.3.5 Prediction of phoneme recognition in spectrally smeared speech .....	41
5.4 Discussion .....	44
5.5 Summary .....	44
<b>6. DISCUSSION AND CONCLUSIONS .....</b>	<b>46</b>
6.1 Phoneme recognition data .....	46
6.1.1 General effect of smearing .....	46
6.1.2 Smearing thresholds .....	46
6.1.3 Slope of the smearing function .....	47
6.2 Replication .....	47
6.2.1 Causes of replication effect .....	47
6.2.2 Implications of replication effect .....	48
6.3 Phoneme differences .....	48
6.3.1 Vowels versus consonants .....	48

6.3.2 Initial versus final consonants.....	49
6.4 The gender effect .....	50
6.4.1 Existence of a gender effect .....	50
6.4.2 Magnitude of the gender effect .....	50
6.4.3 Gender effect and phoneme position .....	51
6.5 Cause of the gender effect.....	51
6.5.1 Initial hypothesis .....	51
6.5.2 Support from absolute formant data.....	52
6.5.3 Support from correlational data .....	52
6.5.4 Perceptual spectral smearing.....	53
6.5.5 Other gender differences .....	54
6.6 Further research.....	55
6.6.1 Smearing algorithms .....	55
6.6.2 Speech processing .....	55
6.6.3 Studies with hearing-impaired subjects .....	56
6.6.4 Intra-talker effects .....	56
6.7 Clinical implications .....	57
6.7.1 Speech perception testing.....	57
6.7.2 Psychoacoustic evaluation .....	57
6.7.3 Teachers of hearing-impaired children .....	57
6.7.4 Amplification .....	58
6.8 Conclusions.....	58
APPENDICES .....	60
A. Subject consent form.....	60
B. Instructions to subjects .....	61
C. AB isophonemic word lists .....	62
C1—List 1, 2, 3, 4, 5, 6.....	62
C2—List 8, 9, 10, 11, 12.....	63
C3—List 13, 14, 15.....	64
D. Spectrographic examples of smearing .....	65
D1—Spectrograms of the word “shed” .....	66
D2—Smearred and unsmearred vowel “ε” .....	67
E. Dadisp worksheet .....	68
F. Counterbalancing test sequence .....	72
G. Subject answer sheet.....	74
H. Linear predictive coding analysis.....	75
I. Raw data of spectral smearing results .....	77
I1—Phoneme scores .....	77
I2—Initial consonant scores.....	78
I3—Vowel nuclei scores .....	79
I4—Final consonant scores .....	80
J. Raw data of formant results.....	81
J1—Formant data for Talker 1 .....	81
J2—Formant data for Talker 2 .....	82
J3—Formant data for Talker 3 .....	83

J4—Formant data for Talker 4 .....	84
J5—Formant data for Talker 5 .....	85
J6—Formant data for Talker 6 .....	86
References .....	87

## List of Tables

Table	Page
4.1 Analysis of variance in phoneme recognition scores.....	29
5.1 Mean frequencies, in Hertz, of formants 1 through 4 for 3 female and 3 male talkers .....	39
5.2 Two-way repeated-measures analysis of variance in the log-transformed mean formant data.....	40
5.3 Female-male formant difference in percent and octaves .....	40
5.4 Five estimates of mean formant spacing, in Hertz, for 6 talkers together with mean phoneme recognition of the various spectral smearing conditions .....	41
5.5 Predictors of phoneme recognition in smeared speech.....	42

## List of Figures

Figure	Page
4.1 Phoneme recognition in smeared speech as a function of replication .....	30
4.2 Phoneme recognition in smeared speech as a function of phoneme position and talker gender .....	32
4.3 Phoneme recognition in smeared speech as a function of smearing condition and phoneme position .....	33
4.4 Phoneme recognition in smeared speech as a function of talker gender and smeared bandwidth in Hertz.....	35
5.1 Average values of the frequencies of formants 1 through 4 for 3 female and 3 male talkers estimated from recordings of 7 monophthongs in each of 3 recorded words lists .....	38
5.2 Phoneme recognition in smeared speech as a function of the frequency spacing of the first and second formants.....	43
5.3 Phoneme recognition in smeared speech as a function of the frequency of the second formant.....	43

## CHAPTER 1 INTRODUCTION AND PURPOSE

### 1.1. Statement of purpose

The purpose of this study was to test the prediction that the speech of female talkers is more resistant to the negative effects of spectral smearing on phoneme recognition than is the speech of male talkers.

### 1.2. Spectral smearing

Spectral smearing refers to a computerized technique that is used to process the acoustic speech signal so as to artificially reduce spectral resolution in normally hearing subjects. It has been demonstrated in research studies that this method can simulate one of the known consequences of sensorineural hearing loss, namely, a loss of frequency resolution. By measuring the recognition of smeared speech in normally hearing listeners, it is possible to test hypotheses about the nature and causes of poor speech perception by hearing-impaired listeners. It is also possible to perform pre-clinical trials of hearing aid processing schemes that are designed to reduce the negative effects of poor spectral resolution.

Previous work on this topic has shown that, in normally hearing listeners, spectral smearing does not cause a serious reduction in phoneme and word recognition ability until the amount of smearing is roughly equal to the average spacing between vocal tract formants. This finding strongly suggests that spectral smearing (and, by implication, sensorineural hearing loss) degrades speech perception mainly because it reduces the

listener's ability to perceive the formant patterns of speech. This conclusion is in keeping with much of what is known about the acoustics of speech and the contribution of formant patterns to speech perception. Vowel identification is highly dependent on the relationships among formant frequencies. In addition, the place of articulation of consonants is often cued by formant transitions between vocalic and consonantal segments of the acoustic signals.

Formant patterns are not the only cues involved in speech perception. Periodicity and temporal amplitude envelope are two other important sources of information. Nevertheless, it is clear that the instantaneous spectral envelope, and its variation over time, play the major role. The normal human hearing mechanism has remarkable spectral resolution, - that is, the ears can detect very small changes in frequency and spectral envelope. Any physical distortion or physiological damage that reduces the ability to detect changes of spectral envelope will have negative effects on the listener's ability to recognize speech sounds and the words that contain them.

### 1.3 Rationale for present study

If it is true that the effects of reduced spectral resolution are mediated by an inability to resolve formant patterns of speech, it follows that speech with more widely spaced formants should be more resistant to spectral smearing than speech in which the formants are closer together. This argument leads to the prediction that increasing the formant spacing will enhance the intelligibility of speech for persons with sensorineural hearing loss. Before testing this last prediction on hearing-impaired subjects, however, it is

appropriate to test the underlying theory that the effects of reduced spectral resolution are directly related to formant spacing.

In fact, there is a ready-made opportunity to test this theory by taking advantage of the known differences between male and female talkers. The vocal tracts of adult female talkers are 15 to 20% (2.5 cm to 3 cm) shorter than those of adult male talkers, leading to a corresponding elevation of the mean frequencies of all formants, and therefore, a corresponding increase in mean formant separation. A finding that the speech of female talkers is more resistant to the effects of smearing than the speech of male talkers will support (but by no means to prove) the theory that the effects of reduced spectral resolution are mediated by a reduced ability to perceive formant patterns.

#### 1.4. Design of the study

In the study to be described, spectral smearing was accomplished by multiplying the speech waveform by low-pass-filtered noise. Using this technique, each pure tone component of the speech spectrum is, in effect, replaced by a band of noise. This band of noise is centered on the frequency of the tone it is replacing, and its bandwidth is equal to twice that of the low-pass-filtered noise. It should be noted that this approach to smearing results in a fixed amount of smearing, in Hz, at all frequencies. When expressed as a percentage, or in octaves, the amount of smearing decreases with increasing frequency. The implications of this fact will be explored in the discussion.

Speech perception performance was measured in this study in terms of the percentage of correctly identified phonemes in lists of consonant-vowel-consonant words. This measure is relatively free of the effects of linguistic context and linguistic knowledge and, therefore, provides a sensitive indicator of the consequences of spectral degradation.

The talkers in this study were three adult males and three adult females. Although the sample size is small, it was felt that the spectral differences between male and female talkers are robust enough to provide an adequately powerful test for the experimental prediction.

To shed further light on the relationship between the spectral properties of speech and the effects of spectral smearing, samples of the recordings of the individual talkers were examined in terms of formant frequencies and formant spacing. The relationships between these acoustic measures and the phoneme recognition data were then explored.

### 1.5. Summary

In summary, sensorineural hearing loss involves reduced frequency resolution, which is believed to be the primary cause of poor speech perception. Simulation of reduced frequency resolution by spectral smearing has been shown to affect speech perception in normally hearing listeners. Experimental data suggest that the effects of reduced spectral resolution are mediated by a loss of ability to perceive the formant patterns in speech. This explanation is in keeping with current knowledge of acoustic phonetics. It leads to the prediction that increasing the formant spacing in speech will increase its resistance to

the negative effects of spectral smearing. To test this prediction, the effects of smearing on phoneme recognition were measured in eight normally hearing listeners, using speech samples from male and female talkers, where differences of formant spacing are known to exist. Acoustic measurements were made both to verify that the expected gender effects were present, and to explore potential predictors of the effects of spectral smearing.

## CHAPTER 2 REVIEW OF THE LITERATURE

### 2.1 Background

#### 2.1.1. Hearing loss and speech perception

The impact of sensorineural hearing damage depends mainly on its effect on the auditory perception of speech. In adults, the problems of speech perception can affect socialization, emotional status, and employment (Vernon and Andrews, 1990). In children, the problems of speech perception may make it difficult to develop spoken language skills (Fry, 1978; Ross, 1982).

#### 2.1.2. Threshold elevation

Sensorineural hearing damage affects speech perception because of diminished psychoacoustic ability. The most obvious psychoacoustic effect is the elevation of threshold. This elevation (expressed as hearing loss in decibels) makes it impossible to hear some, or all, of the conversational speech signal. Fortunately, this problem can largely be addressed with hearing aids (Studebaker and Bess, 1982; Studebaker and Hochberg, 1993).

#### 2.1.3. Loss of auditory resolution

An additional psychoacoustic effect of sensorineural hearing damage, however, is a reduction in the ability to perceive fine detail in the acoustic signal (Zwicker and Schorn, 1978). Unfortunately, hearing aids can do nothing to restore detail in the perceived signal. Even if full audibility of the speech signal is restored by hearing aids, the subject still has

difficulty differentiating among the different sounds in the speech signal. An additional problem is an increase in the interfering effects of noise. The problems of speech perception and noise interference can range from mild to very severe. To a certain extent, the magnitude of these problems depends on the degree of hearing loss - that is, the greater the threshold elevation, the greater the loss of auditory resolution (Bonding, 1979; Dreschler and Plomp, 1980, 1985; Festen and Plomp, 1983; Patterson, Nimmo-Smith, Weber, and Milroy, 1982; Tyler, Summerfield, Wood, and Fernandes, 1982). The correlation, however, is not perfect. It is not uncommon to find two individuals with the same hearing loss in decibels but very different speech perception ability.

#### 2.1.4. Frequency resolution

Sensorineural damage affects the perception of acoustic signals in several ways. For example, there is an abnormal relationship between intensity and loudness - commonly referred to as loudness recruitment (Simon, 1963; Villchur, 1974). There is also a loss of temporal resolution - revealed by such things as poor gap detection or excessive forward masking (Fitzgibbons and Wightman, 1982; Florentine and Buus, 1984). It is generally believed, however, that the most serious psychoacoustic consequence of sensorineural damage is a loss of frequency resolution - that is, the subject has difficulty discriminating sounds with different spectral characteristics (Miller and Nicely, 1955; Bilger and Wang, 1976; Hack and Erber, 1982; Dubno, Dirks, and Schaefer, 1987; Dubno, Dirks, and Ellison, 1989; Turner and Henn, 1989). Tyler (1986) has stated "... reduced frequency resolution may be the most important consequence of cochlear hearing dysfunction." The conclusion that reduced frequency resolution is so serious is based on several

considerations. First, the normal hearing mechanism appears to be specially adapted to provide extremely fine frequency resolution (Evans, 1978). Second, speech scientists have established that much (though not all) of the information in the acoustic speech signal is found in the frequency domain - that is the spectrum and the changes of spectrum over time (Fletcher, 1953; French and Steinberg, 1947). Third, there is ample evidence of reduced frequency resolution in cases of sensorineural hearing loss (Bonding, 1979; Florentine, Buus, Scharf, and Zwicker, 1980; Carney and Nelson, 1983; Glasberg and Moore, 1986; Stelmachowicz, Jesteadt, Gorga, and Mott, 1985; Zwicker and Schorn, 1978). Fourth, speech perception performance has been shown to be correlated with more direct measures of reduced frequency resolution in hearing-impaired subjects (Pick, Evans, and Wilson, 1977; Moore and Glasberg, 1987).

#### 2.1.5. Spectral aspects of speech

The spectral characteristics of speech are dominated by the resonant properties of the vocal tract. Resonance in the vocal tract results in the concentration of energy in specific frequency bands known as formants (Borden and Harris, 1984). There are several formants, each representing a different mode of vibration of the air in the vocal tract. Starting with the lowest frequency, the formants are numbered - F1, F2, F3 etc. The first formant, F1, mainly reflects the raising and lowering of the jaw and tongue. The second formant, F2, mainly reflects the positioning of the tongue towards the front or back of the mouth. Because they convey information about jaw and tongue position, these two formants, and their variations over time, provide a major portion of the acoustic information in speech. Their importance has been established in studies using both

synthetic speech and modified natural speech (Cooper, Delattre, Liberman, Borst, and Gerstman, 1952; Liberman, Delattre, Gerstman, and Cooper, 1956; Delattre, Liberman, and Cooper, 1955; Kewley-Port, 1983). The instantaneous values of these formants provide information on the current position of the tongue, jaw, and lips. The variations over time provide information on where the tongue has been and where it is headed (Fox, 1989; Rakerd and Verbrugge, 1985). This information contributes to the identification of both vowels and consonants (Raphael, 1972, 1980; Strange, 1989; Strange, Verbrugge, Shankweiler, and Edman, 1976; Nearey, 1989). The range covered by F1 and F2 is approximately 300Hz to 3000Hz (depending on age and gender). It is no coincidence that in early research on the development of the telephone, it was found that transmission of the frequency range between 300 and 3000 Hz provided almost perfect intelligibility for connected speech (French and Steinberg, 1947).

#### 2.1.6. Establishing the connection between frequency resolution and speech perception

It was mentioned above that correlations exist between reduced frequency resolution and reduced speech perception in hearing-impaired individuals. Correlation, however, does not guarantee cause and effect. In other words, one can make a strong case for the argument that it is the poor frequency resolution that causes the poor speech perception but other possibilities exist. When the hearing mechanism is damaged, frequency resolution is not the only ability to be affected. Dynamic range (audibility-to-discomfort), intensity/loudness relationships, temporal resolution, and frequency resolution are all correlated with degree of hearing loss (Moore and Glasberg, 1987; Villchur, 1977). It is quite possible that some of these other psychoacoustic factors influence speech

perception, either alone, or in interaction with reduced frequency resolution. Moreover, spectral cues are not the only significant cues in speech perception. Fundamental frequency, amplitude, and their variations of over time have all been shown to contribute (Boothroyd, 1984; Tyler, Summerfield, Wood, and Fernandes, 1982; Rosen and Fourcin, 1986). One way to establish and quantify the cause-effect relationship between frequency resolution and speech perception is through experiments on spectral smearing. In these experiments, the loss of spectral resolution can be simulated without introducing changes of temporal resolution or of the intensity/loudness relationship. By manipulating the simulated spectral resolution and measuring its effect on speech perception in normally hearing subjects, the details of any cause-effect relationship can be determined. This is the rationale behind the studies of spectral smearing reviewed below.

## 2.2. Studies of spectral smearing

### 2.2.1. Villchur (1977)

Villchur (1977) reported the first study that used an electronic device to simulate loss of hearing sensitivity, loss of frequency resolution, and abnormal loudness growth. He developed a simulation that blurred speech by a 16-channel amplitude-modulated noise-carrier system. These carriers allowed noise bands to overlap one another, and center frequencies and bandwidths were adjustable. The greater the overlap between bands the less distinct was the frequency information in the coded speech. Data were obtained from three subjects writing each phoneme in PB-50 word lists being heard, and the percentage of correctly identified phonemes was calculated. Phoneme recognition was found to decrease as the bandwidth of the noise bands was increased.

### 2.2.2. Gagné and Erber (1987)

Gagné and Erber (1987) used a "time/frequency jittering" processing device in which a modified analog delay line as well as low-pass filtered noise were used to simulate reduced frequency selectivity. Consonant, vowel, and whole word recognition scores were measured, using monosyllabic word lists, in 10 normally hearing subjects. Four types of distortion were used to simulate different degrees of hearing loss and audiogram configuration. Condition B simulated a high frequency hearing loss without loss of frequency resolution. Conditions C and D simulated a moderate hearing loss with flat configuration, but C differed from D in the amount of frequency jittering applied to the signal. For condition E, the device was set to simulate a severe to profound hearing loss with poor frequency resolution. As would be expected, the last condition, with the most severe spectral smearing had the greatest effect on phoneme recognition (35% and 19% for vowels and consonants, respectively). The other three conditions had some effect on phoneme recognition but were not very different from each other (approximately 75% and 45% for vowels and consonants, respectively). Place of articulation was the most difficult feature for consonant identification in all conditions, and many errors were made in identifying fricatives. The overall results were similar to those found in subjects with sensorineural hearing loss.

### 2.2.3. Celmer and Bienvenue (1987)

Celmer and Bienvenue (1987) processed speech signals by separating them into 13 ms overlapping Hamming-windowed segments and computing the spectrum using a Fast Fourier Transform. The discrete frequency amplitudes were averaged and set equal to a

root-mean-square (rms) value, followed by a modification of the spectrum, then converted back to waveform with inverse transformation. It should be noted that the smearing bandwidth was proportional to frequency (i.e. this was logarithmic smearing). Forty-eight normally hearing young adults and sixty-eight persons with sensorineural hearing loss served as subjects. Subjects' word recognition scores were highly correlated with smearing bandwidth, - that is, word recognition scores decreased as the bandwidth increased. For the hearing-impaired group, the "knee" at which the performance started to drop coincided with their critical bandwidth - as measured using a loudness summation test. In addition, a significant correlation was found between critical band measures and speech intelligibility. The authors concluded that the integrity of the critical band is an important factor in speech intelligibility, and that speech intelligibility suffers when the smearing bandwidth is greater than the critical bandwidth. The critical band is the range of frequencies over which noise energy contributes to the masking of a pure tone located at the spectral center of the noise. Energy within the band contributes to masking. Energy outside the band does not. The critical band phenomenon has been explained in terms of a hypothetical "auditory filter". The ear's behavior is similar to that of a bank of overlapping filters - each filter responding to tones over a restricted frequency range. The filter width and the critical bandwidth are not exactly the same but are closely related. Both are examples of the ear's "frequency integration" - that is the summing of energy over a narrow band of frequencies roughly 1/6 to 1/3 of an octave wide. In a very real sense, this behavior represents natural "spectral smearing". It is not surprising that artificial smearing would need to exceed the ear's own spectral smearing in order for it to have a significant effect on speech perception.

#### 2.2.4. Ter Keurs, Festen, and Plomp (1992)

Ter Keurs, Festen, and Plomp (1992) investigated the effect of spectral envelope smearing on speech perception in 16 normal-hearing subjects. The spectral envelope of the signal was smeared over 1/8, 1/2, and 2 octaves. With this method, the spectrum was smeared, resulting in reduced spectral contrasts, but the signal's phase spectrum and harmonic structure were preserved. Many measures of speech perception were used in several experiments. Speech-reception threshold (SRT) for sentences was found to deteriorate when the degree of smearing was between 1/3 and 1/2 octaves - which is close to the normal ear's critical bandwidth. When examining vowel and consonant identification as a function of the amount of spectral smearing, the authors found that a large number of vowels were identified as the back vowels /ɔ/ and /u/. Overall, initial consonant identification was higher than final consonant identification. The average identification score for consonants in noise was much lower (about 30 percentage points) when smearing exceeded 2 octaves. The average score for consonants was almost twice the average vowel score in the same condition (61% vs. 32%), suggesting that vowel identification was more susceptible to loss of spectral contrasts than was consonant identification. Confusion matrices for consonants showed that consonants were confused primarily with respect to place of articulation.

#### 2.2.5. Baer and Moore (1993)

The effect of smearing has been studied on sentence intelligibility in quiet and in noise has been studied by Baer and Moore (1993). The authors applied several different types

of smearing to simulate different degrees of broadening and asymmetry of the auditory filter. The complex spectrum was separated into power and phase components before smearing. Then the asymmetrically smeared power spectrum and the unmodified phase spectrum were recombined and an overlapping-addition algorithm was used to produce the final waveforms. Each of nine normally hearing subjects listened to 18 sentence lists that were processed in three conditions: no noise, 0 dB S/N, and -3 dB S/N. For symmetrical smearing, intelligibility of sentences in quiet was hardly affected by spectral smearing, even when simulating auditory filters that were six times broader than normal. However, speech intelligibility in noise was significantly affected by smearing of larger degrees and at low signal-to-noise ratio. The results of the asymmetrical broadening simulation showed a decreased performance with increasing degree of smearing even in the quiet condition (though the difference was seen only in the widest smeared condition). The effect was greater for the lower side broadening than for the upper side broadening of the filter. This last finding can be explained in terms of cochlear physiology. It has long been known that the pattern of excitation along the basilar membrane, by a pure tone, is not symmetrical but spreads basally (towards higher frequencies) from the point of maximum stimulation. This pattern of excitation results in the phenomenon of "upward spread of masking". In other words, the normal cochlea already "smears" frequency in the direction of high frequencies – but not low frequencies. It is reasonable to expect, therefore, that the introduction of artificial low-frequency smearing will have more effect on speech perception than similar amounts of high-frequency smearing.

### 2.2.6. Howard-Jones and Summers (1992)

Howard-Jones and Summers (1992) have pointed out that the simulation of reduced frequency resolution by spectral smearing also degrades the temporal properties of speech. They set out to determine the relative importance of the spectral and temporal distortions. Monosyllabic word recognition was measured in normally hearing subjects under 4 conditions. Each condition involved spectral smearing but the four processing methods produced different amounts of temporal distortion. Word recognition was severely degraded but scores under the four conditions were very similar. The authors concluded that the loss of intelligibility was caused mainly by "...loss of spectral information, with loss of temporal structure being of secondary importance". The results of this study support the validity of spectral smearing as a simulation of the effects of reduced frequency resolution in sensorineural hearing loss.

### 2.2.7. Summers (1991), Summers and Al-Dabbagh (1985), and Summers, Al-Dabbagh, and Garnham (1986)

Summers (1991), Summers and Al-Dabbagh (1985), and Summers, Al-Dabbagh, and Garnham (1986) investigated three smearing techniques that had different relationships between the amount of smearing and frequency. In the "uniform smearing" condition, the smearing bandwidth in Hz did not vary with frequency, while in the "constant-percentage smearing", or "logarithmic smearing", the smearing bandwidth was proportional to frequency (e.g., at 2000 Hz, the smeared bandwidth was twice that at 1000 Hz). In the "mixed smearing" condition, the bandwidth in Hz was constant for frequencies below 630 Hz and increased proportionally with frequency for frequencies above 630 Hz.

Summers examined word recognition in 20 normally hearing subjects under those three smearing conditions. The results showed the expected trend- word scores fell as the smeared bandwidth ( $\Delta f$ ) increased for all processed conditions. Unfortunately, it is not possible to draw conclusions from these data about the relative merits of the different smearing techniques because there is no condition under which logarithmic and linear smearing can be considered equivalent.

#### 2.2.8. Shannon, Zeng, Kamath, Wygonski, and Ekelid (1995).

Shannon, Zeng, Kamath, Wygonski, and Ekelid (1995) have used spectral smearing to simulate perception via multi-channel cochlear implants. They employ a noise-band technique similar to that of Villchur (1977) in which the amplitude envelopes in a series of non-overlapping frequency bands of speech are used to modulate bands of noise. They showed that high levels of sentence recognition are possible when speech is represented by as few as 4 bands of amplitude-modulated noise. They do not report, however, on the effects of smearing on the perception of words and phonemes without benefit of sentence context.

#### 2.2.9. Boothroyd, Mulhearn, Ostroff, and Gong (1996)

The most recent study of the simulation of reduced frequency resolution by spectral smearing was carried out by Boothroyd, Mulhearn, Ostroff, and Gong (1996) who investigated the effects of spectral smearing on phoneme and word recognition by normally hearing adults. This study employed uniform smearing, i.e., the amount of spectral smearing, in Hz, was independent of frequency. Smearing was accomplished by

multiplying the speech waveform by a random-noise waveform that had been low-pass filtered. The result was that each tonal component of the speech spectrum was replaced with a band of noise whose bandwidth was twice that of the modulating noise (i.e., a 500 Hz low-pass filtered noise would produce a smeared bandwidth of 1000 Hz). The bandwidth of the low-pass filtered noise ranged from 125 to 4000 Hz in half octave steps - to give smeared bandwidths of 250 through 8000 Hz in one octave steps. An unfiltered noise was also used to produce complete smearing. Note that this technique degraded the spectral envelope, the fine temporal structure, and the fine harmonic structure of the speech signal-- but not the gross temporal envelope. Speech of a single male talker was used in this study. Phoneme recognition scores (in consonant-vowel-consonant words) fell from 97% to 12% as the amount of smearing increased from zero (unprocessed) to complete (speech signal modulated by the unfiltered random noise). The smeared bandwidth required for 50% phoneme recognition was about 1400 Hz. Word recognition scores fell from 91% to near 0% as the smeared bandwidth increased from zero to 2000 Hz. A smeared bandwidth of only 720 Hz was required to reduce word recognition to 50%. Vowel recognition was more affected by spectral smearing than was consonant recognition. A smeared bandwidth of 1150 Hz was needed to reduce vowel recognition to 50%, whereas a smeared bandwidth of 1600 Hz was required for consonants (1350 Hz and 2000 Hz for initial and final consonants, respectively). The effects of smearing on speech perception were also explored in the presence of noise. The signal-to-noise ratio required for 50% phoneme recognition was increased by 12.9 dB and 16.4 dB for smearing at 707 Hz and 2000 Hz, respectively, when compared with unsmeared speech. The effect of spectral smearing on the perception of three consonant features was also

examined. The findings showed that the perception of voicing was the least affected by smearing, while the perception of place of articulation was the most affected. The perception of manner fell between the other two.

### 2.3. Summary

Previous research confirms that artificially reduced frequency resolution has a negative effect both on speech perception and on the ability to perceive speech in noise. The findings, therefore, offer strong support for the conclusion that reduced frequency resolution is a major cause of the speech perception difficulties encountered by persons with sensorineural hearing loss. Further support comes from the data of Ter Keurs et al., (1992) and Celmer and Bienvenue (1987), in which speech perception was first affected when the smeared bandwidth exceeded the critical bandwidth.

The study of Boothroyd et al., (1996) provided an estimate of smearing threshold - that is, the smeared bandwidth needed to reduce phoneme recognition to 50%. This threshold was in the region of 1400 Hz (1150 Hz for vowels and 1600 Hz for consonants). Although by no means conclusive, these numbers are highly suggestive of the notion that the principal effect of smearing is to obliterate the details of formant structure in the acoustic speech signal. In a typical male talker (as used in the Boothroyd et al. study) the average formant spacing is around 1000 Hz (Peterson and Barney, 1952). The spacing between F1 and F2 will, therefore, be expected to be below 1000 Hz for roughly 50% of the time. A smeared bandwidth of 1000 Hz would, therefore, make it impossible for the listener to differentiate F1 from F2, or to estimate the relationship between them, roughly 50% of the time.

#### 2.4. Rationale for present study

The data of Boothroyd et al., (1996) suggest that spectral smearing degrades speech perception when it interferes with the listener's ability to perceive formant patterns. A prediction of this hypothesis is that the effects of reduced spectral resolution can be reduced by increasing the formant spacing in speech. The existing differences of speech acoustics between men and women provides a ready-made opportunity to test this prediction (Peterson and Barney, 1952). Evidence that the speech of women is more resistant to the effects of spectral smearing than is the speech of men would provide support for the theory that reduced spectral resolution affects speech perception by making it difficult to perceive the formant patterns.

## CHAPTER 3 METHOD

### 3.1. Outline

Two groups of subjects participated in this study. One group served to produce the test stimuli (6 talkers), and the other served to perceive them (8 listeners). Test materials consisted of 150 consonant-vowel-consonant (CVC) words that were recorded on digital audiotape by each talker. Spectral smearing was introduced off-line by multiplying the digitized recordings by a set of low-pass filtered noises. The smeared stimuli were converted back to analog form and presented to the 8 listeners monaurally, under headphones, through the speech circuit of a clinical audiometer. The listeners' task was to repeat and write what they heard. The independent variables were talker gender and amount of smearing. The dependent variable was the percentage of phonemes recognized. In addition, a sample of unsmeared recordings was analyzed spectrographically in order to obtain comparative formant frequency data from the 6 talkers. These data were to be used to examine inter-talker differences in the effects of spectral smearing on phoneme recognition.

### 3.2. Subjects

#### 3.2.1. Talkers

There were 6 talkers, 3 male and 3 female, recruited from the Program in Speech and Hearing Sciences at the Graduate Center of the City University of New York. An additional talker was recorded so as to provide practice materials. Talkers were unpaid volunteers who met the following criteria:

1. Native speaker of American English, without regard for regional accent.
2. No apparent speech, language, or voice disorder.
3. No evidence of current respiratory infection.
4. No known hearing disorder.
5. Age 45 years or less.

### 3.2.2. Listeners

There were 8 listeners, 4 male and 4 female, recruited from the graduate student dormitory of the City University of New York. Listeners were paid volunteers who met the following criteria:

1. Native speaker of American English, without regard for regional accent.
2. No apparent speech, language, or voice disorder.
3. No evidence of current respiratory infection.
4. Ages between 20 and 50 years.
5. Naïve with respect to the procedures, the distorted speech, and methods.
6. Hearing thresholds of 15 dB or less at octave intervals from 250 through 8000Hz, bilaterally.

All subjects gave their informed consent. (The consent forms are shown in Appendix A).

Also, all subjects read instructions prior to the experiments. (The instructions are shown in Appendix B).

### 3.3. Test Materials

The test material consisted of 15 lists of consonant-vowel-consonant (CVC) words developed by Boothroyd (1968) and later modified for North American usage

(Boothroyd, 1984). Each list consists of 10 vowels and 20 consonants that are arranged to form 10 words. The phonemes are those that occur most commonly in English CVC words. Because each list contains the same 30 phonemes, Boothroyd refers to them as *isophonemic* word lists. Note that the lists are balanced for phonemic content only. There is no attempt to balance them for frequency of word occurrence, lexical neighborhood, or word type. Nor is consonant position (pre- or post-vocalic) controlled. These materials, which have a long history of use in both clinical and research applications, are shown in Appendix C.

### 3.4. Preparation of recordings

Speech materials were recorded on digital audiotape, using a Panasonic Digital Audio Tape recorder (model SV-255) and an Electrovoice dynamic omni-directional microphone (model 635A). All recordings were made in a commercial sound-treated room. Talker-to-microphone distance was maintained at 6 inches and the talkers were instructed to maintain constant effort. Each of the 6 "experimental" talkers and the additional "practice" talker recorded all 150 words: preceded by "number X" where X was the number of the word within its list. During recording, the experimenter monitored the VU meter of the recorder to ensure that no vowel peaks exceeded the 0 dB mark.

### 3.5. Processing of stimuli

#### 3.5.1. Digitization

The recorded test words were digitized at 22050 samples per second with a resolution of 16 bits. Digitization was performed in Wave for Windows (Turtle Beach, version 2.01).

Gain levels were adjusted so that the instantaneous amplitude did not exceed 80% of the dynamic range of the analog-to-digital converter. Once adjusted, these levels remained constant for all talkers and all tokens. At the time of digitization, the carrier phrases were removed from the test words.

### 3.5.2. Spectral smearing

Spectral smearing of the digital files was accomplished in DaDisp (DSP Corporation, version 3.0). Four Finite Impulse Response low-pass filters were created, with 100 dB stop-band attenuation, 1 dB pass-band ripple, and a stop-band edge at 300 Hz above the cut-off frequency. The cut-off frequencies were 354, 500, 707, and 1000 Hz. Random noise was processed by each filter to produce 4 low-pass filtered noises. Each word file was then multiplied by each of the 4 noises to produce spectral smearing. Note that the smeared bandwidths were twice the noise bandwidths. Thus, the processing produced four sets of test materials for each talker, with smeared bandwidths of 707, 1000, 1414, and 2000 Hz, in addition to the original unsmeared recordings. Pilot experiments showed that the smeared bandwidth of 707 Hz had only a small effect on phoneme recognition so this condition was omitted from the main experiment. Spectrograms and spectra of a sample of smeared words are shown in Appendix D. Details of the DaDisp worksheet used for processing are given in Appendix E.

## 3.6. Design

### 3.6.1. Dependent and independent variables

Phoneme scores, in percent correct, collected from 8 listeners were used as the dependent variable. Smearing condition (at 3 levels) and Gender were the principal independent

variables. In a more detailed analysis, the effects of Replication (4 levels) and Phoneme Position in a word (3 levels) were also examined. Each of the 8 listeners listened to all six talkers under each of 3 smearing conditions. Four word lists were presented under each condition for each talker - giving 4 replications. The main effects of talker gender and smearing condition, and the interaction between the two, were examined in a repeated-measures analysis of variance with gender as a grouping variable and smearing condition as a repeated measure. As indicated earlier, replication and phoneme position were also included as repeated measures.

### 3.6.2. Counterbalancing of lists

Because lists 1 to 3 were heard without smearing during practice, these lists were not used for testing under the smeared conditions. The order of occurrence of the remaining 12 test lists ( i.e., 4 through 15) was counterbalanced across talker, smearing condition, replication, and listener. This counterbalancing is shown in Appendix F. In addition, the order of presentation of the 10 words in each list was randomized during presentation. The purpose was to minimize the effects of list-differences and learning on the results of the experiment.

### 3.6.3. Presentation order of smearing conditions

The order of presentation of the smearing conditions was held constant for each replication. That is, each talker was presented with the 1000 Hz smeared bandwidth, followed by the 1414 Hz and the 2000 Hz bandwidths in that order. The limited number of test lists (12) and the large number of subjects x conditions x replications (72) required each word to be heard 6 times by each listener during the experiment. In part, the 4 replications were included to provide the opportunity to examine the magnitude of this

and other potential learning effects.

### 3.7. Procedure

#### 3.7.1. Replication

Each listener was tested in 2 sessions separated by at least one week. At the beginning of the first session, listeners read and signed the informed consent form and pure tone thresholds were checked to confirm that the listeners met the auditory criterion. In addition, listeners heard lists 1 to 3 in the unsmearred condition to establish familiarity with the materials and procedures. During each session listeners were presented with 1 word list for each of the 6 talkers under each of 3 smearing conditions. After a short break, this process was repeated. Each session took approximately 3 hours. The outcome of the two sessions was 4 phoneme recognition scores for each talker under each of the 3 smearing conditions. These are the data that will be presented in Chapter 4.

#### 3.7.2. Test stimuli presentation

Stimuli were presented monaurally via a Grason-Stadler Audiometer (model 10). Input level was adjusted to give 0 dB deflection for the vowel peak in the word "wise" spoken by talker BM. This word was chosen for calibration because it had the highest peak vowel level in the total corpus. Stimuli were presented monaurally at 65 dBHL using TDH-49P earphones fitted with MX41/AR ear cushions. The choice of test ear was alternated across listeners.

#### 3.7.3. Scoring

At the beginning of each session the listener was provided with pencils and answer sheets. Each sheet contained 4 blocks with 10 cells in each block, as shown in Appendix

G. Listeners were instructed to write down what they thought they heard in each cell. Guessing was encouraged so that no cell should be left blank. Listeners were also asked to repeat the word. The tester monitored the spoken response through the talkback system and also recorded it on digital audiotape to help with later interpretation of written responses. Listeners received no feedback on the accuracy of their responses.

### 3.8. Response judgements

For each listener, the number of phonemes (out of 30) correctly identified for each talker, each condition, and each replication was totaled and converted to a percentage score. Two scorers served to judge the listeners' response sheets. The recordings were used as an aid in interpreting the written responses when a phonetic interpretation was unclear. When there was disagreement, the experimenter's advisor acted as a final judge. Phoneme responses were classified as errors when there were omissions (e.g., "ed" for "bed"), additions (e.g., "blaze" for "laze"), or substitutions (e.g., "rise" for "wise"). Misspellings (e.g., "coff" for "cough") and homonyms (e.g., "knot" for "not") were counted as correct responses.

### 3.9. Spectrographic analysis

Spectrographic analysis of 180 words from each subject's unsmeared recordings was carried out with an analysis system from Kay Elemetrics (model CSL 4300B). Frequencies of the first 4 formants were measured at a single 20 msec time slice in the temporal center of the vowel, using Linear Predictive Coding (LPC) analysis. The parameters of the LPC analysis are provided in Appendix H, together with an example of

one of the analyses. The results of these analyses and examination of the relationship between formant values and the effects of smearing will be presented in Chapter 5.

## CHAPTER 4

### RESULTS OF SMEARING STUDY

#### 4.1. Raw data

Individual data for percent recognition of initial consonants, vowels, and final consonants under the 3 smearing conditions are shown in Appendix I. Data are shown separately for the 4 replications and also collapsed across replications.

#### 4.2. Analysis of variance

The percent correct data were subjected to a repeated-measures analysis of variance with 1 grouping variable (Gender at 2 levels) and 3 repeated measures: Phoneme position at 3 levels (initial consonant, vowel, and final consonant), smeared bandwidth at 3 levels (1.0, 1.4, and 2.0 kHz), and replication at 4 levels. The result is shown in Table 4.1. It will be seen that all four main effects reached significance at or well beyond the 1% level. The interaction between Gender and Phoneme position reached significance at the 5% level and the interaction between Phoneme position and Smeared Bandwidth reached significance at well beyond the 1% level.

Table 4.1. Analysis of variance in phoneme recognition scores. Shading indicates probability of 0.05 or less.

	df	MS	df	MS	F	p-level
	Effect	Effect	Error	Error		
Gender	1	2220.2	4	93.0	23.87	0.008126
Phoneme position	2	4844.4	8	48.9	99.05	0.000002
Smearing	2	5696.8	8	55.2	103.18	0.000002
Replication	3	2176.8	12	28.9	75.23	0.000000
GxP	2	267.6	8	48.9	5.47	0.031802
GxS	2	6.7	8	55.2	0.12	0.886783
PxS	4	197.4	16	26.2	7.54	0.001298
GxR	3	11.4	12	28.9	0.39	0.759839
PxR	6	34.8	24	26.3	1.32	0.286060
SxR	6	33.3	24	26.3	1.27	0.309641
GxPxS	4	71.4	16	26.2	2.73	0.066381
GxPxR	6	25.3	24	26.3	0.96	0.472297
GxSxR	6	38.6	24	26.3	1.47	0.230783
PxSxR	12	16.1	48	20.4	0.79	0.659315
GxPxSxR	12	27.2	48	20.4	1.33	0.231359

#### 4.3. Replication

The effect of replication was striking, as illustrated in Figure 4.1. Collapsed across subjects and smearing conditions, the mean phoneme recognition score rose from 52.1% at the beginning of testing to 67.0% at the end, an increase of almost 15 percentage points. Moreover, the improvement was steady (at around 5 percentage points) from replication to replication, with no evidence of slowing. The sources of this apparent learning effect will be explored in the discussion. For present purposes, however, the important observation is that none of the interactions involving replication reached the 5% level of significance. In other words, there is no evidence that the relationships among the two most important variables - gender and smearing bandwidth - changed

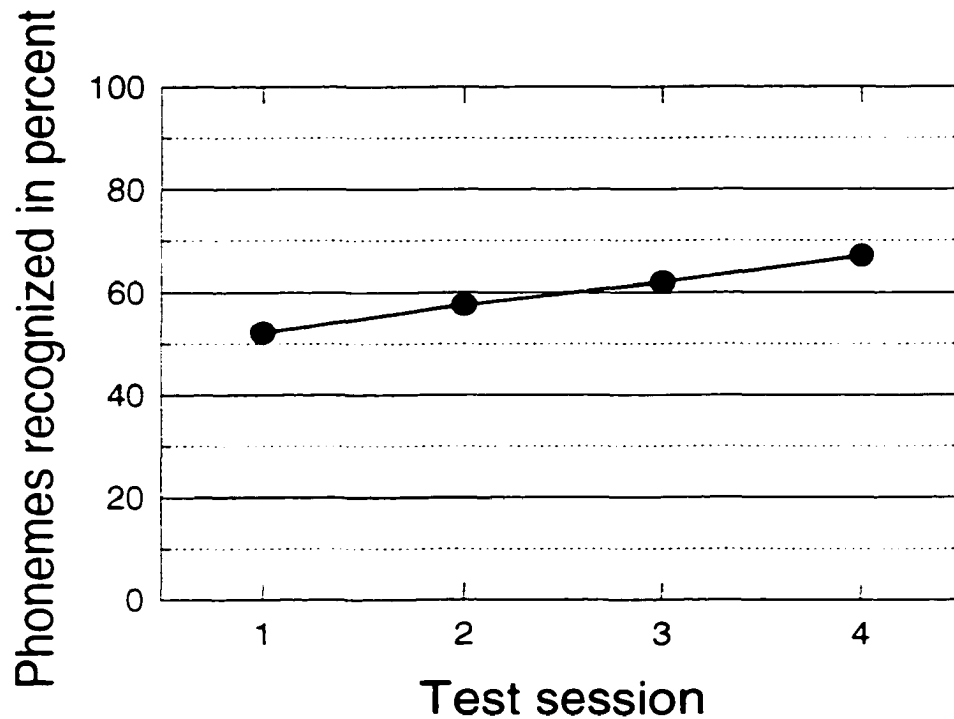


Figure 4.1. Phoneme recognition in smeared speech as a function of replication. Data are collapsed across listener, talker and smearing condition. Each listener heard one word list at each replication for each talker x smearing condition combination.

during the course of the experiment. For this reason it is appropriate to address the main research question using data that are collapsed across replication.

#### 4.4. Main effect of gender

When collapsed across replication and smeared bandwidth, the average phoneme recognition scores for the female and male talkers were 62.9% and 56.4%, respectively. The difference of 6.5 percentage points, though small, is significant at the 1% level. This

finding is in keeping with the prediction that the speech of female talkers would be more resistant to the effects of spectral smearing than the speech of male talkers.

#### 4.5. Effect of phoneme position

##### 4.5.1 Main effect of phoneme position

When collapsed across subjects, smearing conditions, and replications, the average scores for initial consonant, vowels, and final consonants were 64.8, 50.2, and 64.0%, respectively. Post-hoc testing, using the least-significant-difference test, failed to reveal a significant difference between initial and final consonants. On average, however, the consonant scores were 14.2 percentage points higher than the vowel scores. This difference was highly significant ( $p = 0.000005$ ). The fact that consonants are more resistant than vowels to the effects of the type of spectral smearing used in this study has been noted previously (Boothroyd et al. 1996). This issue will be explored in the discussion.

##### 4.5.2 Interaction with gender

There was evidence of an interaction between phoneme position and gender ( $p=0.03$ ). Details of this interaction are illustrated in Figure 4.2. In post hoc testing, using the least-significant-difference test, the gender effect did not reach the 5% level of significance for the initial consonant. There were, however, significant differences for vowels ( $p=0.0004$ ) and final consonants ( $p=0.0016$ ). Thus, the conclusion that the speech of female talkers is more resistant to spectral smearing than that of male talkers is not supported for initial

consonants in these monosyllables. It is, however, supported for the vowel nucleus and the final consonant, and the average phoneme score.

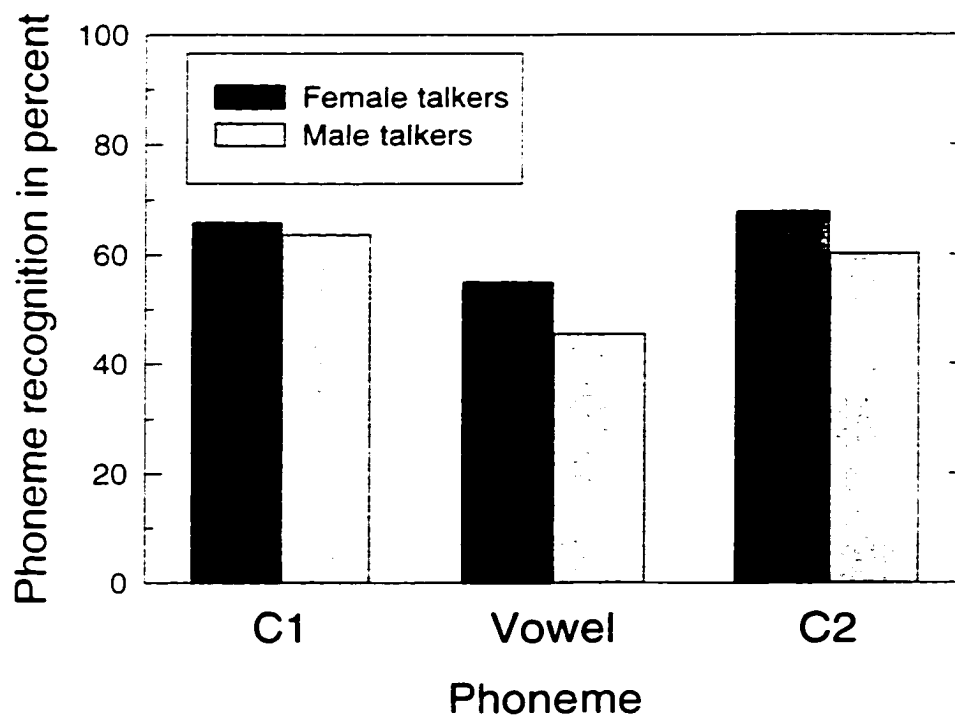


Figure 4.2. Phoneme recognition in smeared speech as a function of phoneme position and talker gender.

#### 4.5.3. Interaction with smeared bandwidth

The significant interaction between phoneme position and smearing indicates that the effect of smearing was not the same across the three phonemes. Details of this interaction are illustrated in Figure 4.3. It will be seen that scores for vowels fell more rapidly with increasing smeared bandwidth (24 percentage points per octave) than did scores for consonants (14 percentage points per octave).

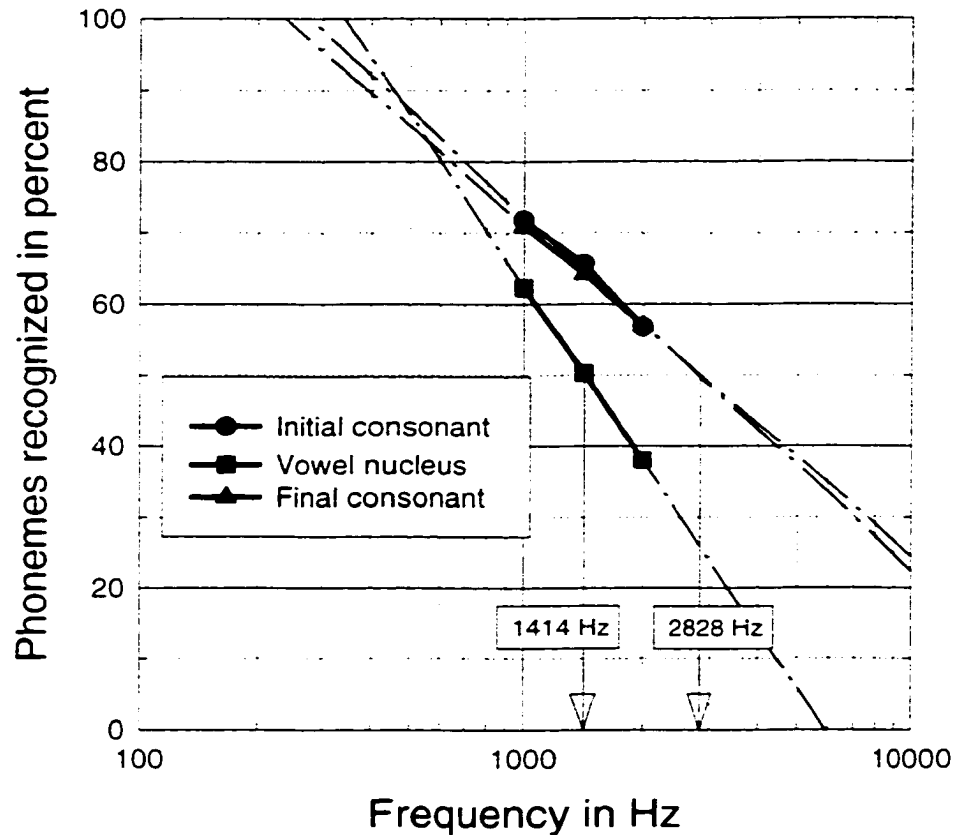


Figure 4.3. Phoneme recognition in smeared speech as a function of smearing condition and phoneme position. Data are collapsed across 6 talkers and 4 replications. Arrows show the estimated smearing thresholds for vowels and consonants.

#### 4.6. Effect of smearing

##### 4.6.1. Main effect of smeared bandwidth

When collapsed across subjects and replications, the average phoneme recognition scores were 68.4, 60.0, and 50.6% for smeared bandwidths of 1.0, 1.4, and 2.0 kHz, respectively. Thus, mean phoneme recognition fell almost linearly at a rate of 17.8 percentage points per octave as smeared bandwidth increased. Note that there are no significant interactions between smeared bandwidth and gender, or between smeared bandwidth and replication.

Thus there is no evidence to suggest that this slope of 17.8 percentage points per octave changes with either gender or replication.

#### 4.6.2. Smearing threshold and gender

The main effect of gender amounted to a 6.5 percentage point difference between the female and male talkers. Combined with the 17.8 percentage point per octave slope just derived, this difference gives an estimated shift of  $6.5/17.8$ , or 0.37 of an octave, in the functions relating phoneme score to smeared bandwidth. This shift is illustrated in Figure 4.4, which shows average phoneme recognition score, collapsed across replications, as a function of smeared bandwidth. Data are shown for the 6 individual talkers together with means for the males and the females. The data for the two genders are also extrapolated using linear regression functions for score versus the logarithm of frequency. It will be seen that the smearing threshold (the smeared bandwidth providing 50% phoneme recognition) averaged 2300 Hz for the female talkers and 1800 Hz for the male talkers. The difference is 0.35 octaves or 28%.

#### 4.7. Summary

These findings support the conclusion that the speech of female talkers is more resistant to spectral smearing –as implemented in this study - than is the speech of male talkers. To the extent that the average formant spacing in the speech of female talkers is greater than in the speech of male talkers, the findings also support the hypothesis that spectral smearing interferes with speech perception by obscuring the formant structure in speech.

The following chapter will explore the formant patterns of the speech in this sample of talkers so as better to define the relationship between formant patterns and the effects of spectral smearing.

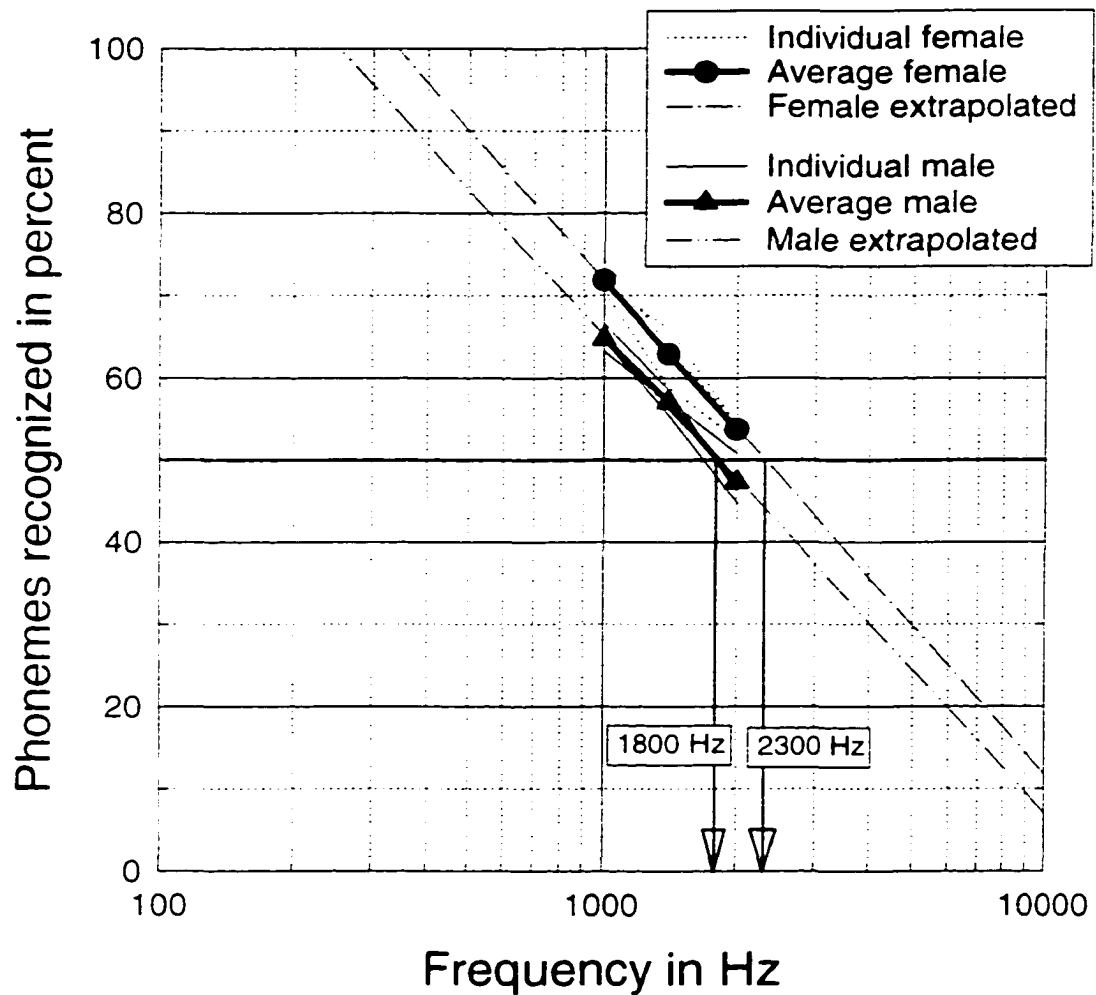


Figure 4.4. Phoneme recognition in smeared speech as a function of talker gender and smeared bandwidth in Hertz. Data are collapsed across phoneme position and replication. Extrapolations are linear regressions on log frequency. Arrows show best estimates of smearing thresholds for the two gender groups.

## CHAPTER 5

### ACOUSTIC ANALYSIS OF THE SPEECH STIMULI

#### 5.1. Purpose

The present study was based on the assumption that the formant spacing in the speech of female talkers would be greater than in the speech of male talkers - and the prediction that this difference would cause female speech to be more resistant to the effects of spectral smearing than male speech. The goals of the work described in this chapter were:

- a) to measure actual formant values in the speech of the talkers used in this study
- b) to compare the results with published data
- c) to examine various measures of formant frequency and formant spacing as potential predictors of the intelligibility of spectrally smeared speech - as reported in chapter 4.

#### 5.2. Method

##### 5.2.1. Selection of vowels

Each AB word list contains an example of each of 10 vowels. Three of these vowels, however, are diphthongs, which are difficult to quantify in terms of steady-state formant frequencies. Analysis was, therefore, restricted to the remaining 7 monophthongs which were: /i/ as in "cheek", /ɪ/ as in "ship", /ɛ/ as in "well", /æ/ as in "fan", /ɑ/ as in "jot", /ʌ/ as in "rug", and /u/ as in "move".

##### 5.2.2. Procedure

Recordings of the first three word lists were used to provide three estimates of the formant frequencies in tokens of these 7 vowels, spoken by each of the 6 talkers. For each

token. estimates of the frequencies of the first 4 formants were obtained from a 20 msec sample taken from the temporal center of the vocalic segment. The measurements were made with the Kay Elemetrics CSL 4300B speech analysis system. Linear predictive coding analysis was used. An example is provided in Appendix H.

### 5.3. Results

#### 5.3.1. Raw data

The raw formant data will be found in Appendix J. For each talker and each vowel, the three estimates of the 4 formant frequencies were averaged. The values for the 7 vowels were then averaged to provide an estimate of the four mean formant frequencies for each talker. The results are included in Appendix J.

#### 5.3.2. Individual and group mean formant frequencies

The mean values of formants 1 through 4 for each of the 6 talkers are shown in Table 5.1. Also shown are the means for the two gender groups. The individual data are illustrated in Figure 5.1, together with corresponding data for formants 1 through 3 taken from Peterson and Barney (1952). For reference, the Peterson and Barney data are also shown in Table 5.1. It will be seen that the present data agree quite well with the Peterson Barney data for formants 1 through 3, in spite of major differences in procedure.

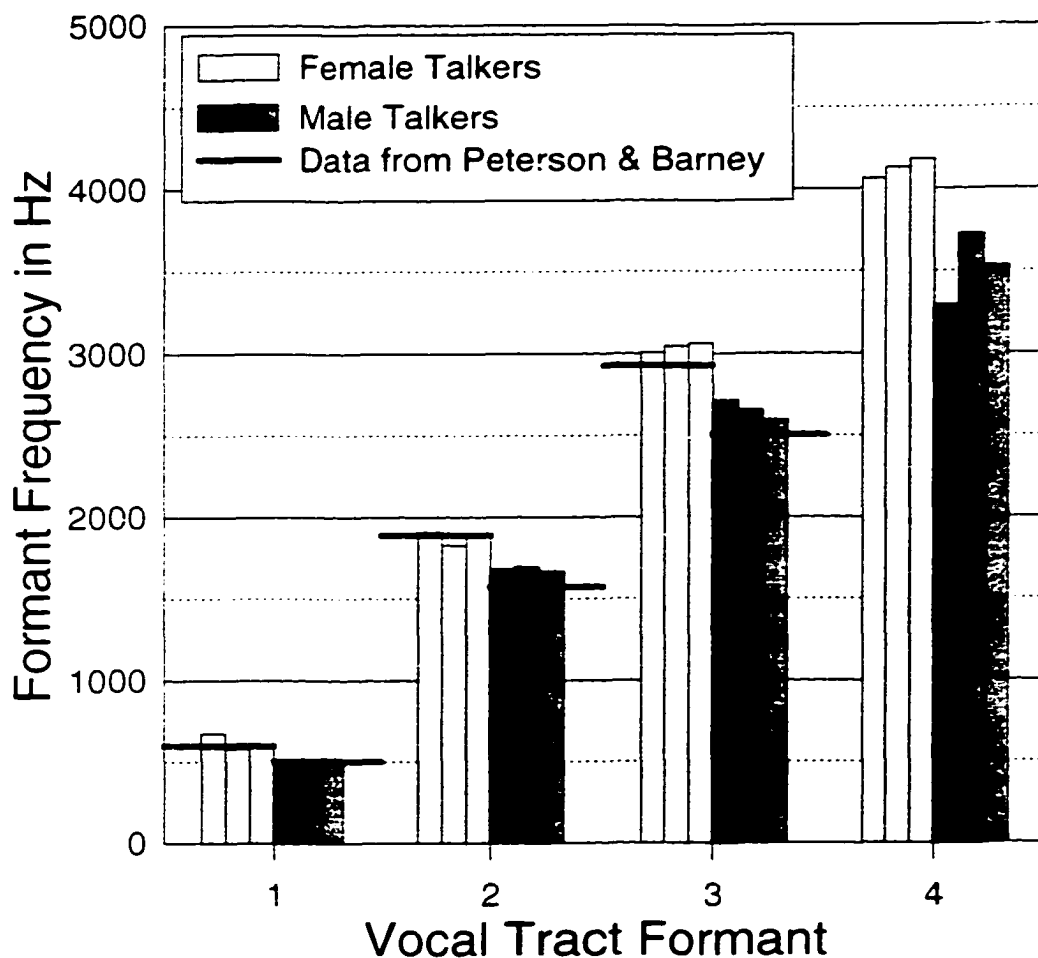


Figure 5.1. Average values of the frequencies of formants 1 through 4 for 3 female and 3 male talkers estimated from recordings of 7 monophthongs in each of 3 recorded word lists. Also shown are published data from Peterson and Barney (1952).

Table 5.1 Mean frequencies, in Hz. of formants 1 through 4 for 3 female and 3 male talkers. Each mean is averaged across 7 vowels from each of 3 lists. Also shown are the means for the two gender groups together with equivalent data from Peterson and Barney (1952) for formants 1 through 3.

Talker	Formant			
	F1	F2	F3	F4
T1	676	1905	3006	4063
T2	583	1826	3045	4128
T3	603	1883	3064	4180
Mean Female	621	1872	3038	4123
Peterson/Barney Female	599	1889	2926	-----
T4	507	1679	2713	3289
T5	509	1693	2654	3723
T6	519	1666	2592	3528
Mean Male	512	1679	2653	3513
Peterson/Barney Male	503	1570	2503	-----

### 5.3.3. The main effect of gender

The individual values of average formant frequency in Table 5.1 were log transformed and subjected to a 2-way, repeated-measures analysis of variance with gender as a grouping variable at 2 levels and formant number as a repeated measure at 4 levels. The result is shown in Table 5.2. It will be seen that the main effect of gender is highly significant ( $F[1,4]=106.8$ ,  $p=.0005$ ), in spite of the small number of subjects in each gender group. The interaction between gender and formant number, however, fails to reach the 5% level of significance. In other words, there is no evidence that the effect of gender, when expressed in octaves, is different for the different formants. These findings are in keeping with the conclusion that the effect of gender is to shift all frequencies, on average, by the same proportional amount.

Table 5.2. Two-way repeated-measures analysis of variance in the log-transformed mean formant data.

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
Gender	1	0.27744	4	0.002599	106.77	0.00050
Formant	3	8.81876	12	0.003103	2841.99	<.000005
GxF	3	0.00393	12	0.003103	1.27	0.33024

Table 5.3 shows the effect of gender on the frequencies of formants 1 through 4, expressed both as a percentage increase and an octave increase when shifting from male to female. It will be seen that, on average, the female formants were 16.2%, or 0.22 octaves, higher than those of the male talkers.

Table 5.3. The amount by which the average frequencies of the formants of the female talkers in the present study exceed those of the male talkers. The female-male difference is expressed both in percent and in octaves.

Formant	octaves	%
F1	0.28	21.3
F2	0.16	11.4
F3	0.20	14.5
F4	0.23	17.4
Mean	0.22	16.2

#### 5.3.4. Formant spacing

Two methods were used to obtain estimates of formant spacing. In one method, the frequency difference between adjacent formants was calculated. In the second method,

the difference between each formant frequency and the frequency of the first formant was calculated and then divided by  $(n-1)$  where  $n$  was the formant number. The results for the 6 talkers are shown in Table 5.4, together with the average phoneme recognition scores taken from Chapter 4. It will be seen that the estimates of mean formant spacing ranged from 860 to 1167 Hz for the male talkers and from 1085 to 1251 Hz for the female talkers. The amount by which the mean female spacing exceeded the mean male spacing averaged 17% with a range from 7 to 26%. These data are all in keeping with the expectation that the formant spacing would be higher for female than for male talkers.

Table 5.4. Five estimates of mean formant spacing, in Hz, for 6 talkers together with mean phoneme recognition of the various spectral smearing conditions. Data are also shown for the group means. Where available, spacing is also reported for the Peterson and Barney data.

Talker	Estimate of formant Spacing (Hz)					Phoneme Score (%)
	F2-F1	F3-F2	F4-F3	(F3-F1)/2	(F4-F1)/3	
T1	1229	1101	1057	1165	1129	63.7
T2	1243	1219	1083	1231	1182	60.6
T3	1280	1181	1116	1231	1192	64.1
Mean Female	1251	1166	1085	1209	1167	62.8
Peterson/Barney Female	1290	1037	-----	1164	-----	-----
T4	1172	1034	576	1103	927	57.1
T5	1184	961	1069	1073	1071	57.1
T6	1147	926	936	1037	1003	55.0
Mean Male	1167	974	860	1071	1000	56.4
Peterson/Barney Male	1067	933	-----	1000	-----	-----

### 5.3.5. Prediction of phoneme recognition in spectrally smeared speech

The final step was to test the various measures of formant frequency and spacing as predictors of resistance to spectral smearing - regardless of gender. Taking formant frequencies from Table 5.1 and formant spacing from Table 5.4, Pearson product-

moment correlations were calculated using the average smeared phoneme recognition scores shown in Table 5.4. The results are shown in Table 5.5. It will be seen that all measures of formant frequency and formant spacing but one were predictive of phoneme recognition. The exception was the spacing of F4 and F3. Of the spacing estimates, the separation of F2 and F1 was the best predictor ( $r[4] = 0.93$ ,  $p = 0.008$ ). This observation is in keeping with the experimental prediction, but suggests that the separation of F1 and F2 is more important than the separation of other formants. Figure 5.2 illustrates the relationship between the F2/F1 spacing and phoneme recognition.

Table 5.5. Predictors of phoneme recognition in smeared speech.

Acoustic Measure	Correlation (df=4)	p value
F1	.89	.0164
F2	.98	.0006
F3	.94	.0059
F4	.85	.0329
F2-F1	.93	.0076
F3-F2	.81	.0516
F4-F3	.52	.2878
(F3-F1)/2	.86	.0275
(F4-F1)/3	.79	.0608

It is interesting to note that the best single predictor of smeared phoneme recognition was not a measure of formant spacing, but the frequency of the second formant ( $r[4]=0.98$ ,  $p=0.0006$ ). In stepwise multiple regression, no other measures added significantly to the predictive value of second formant frequency. Figure 5.3 illustrates the relationship between second formant frequency and smeared phoneme recognition.

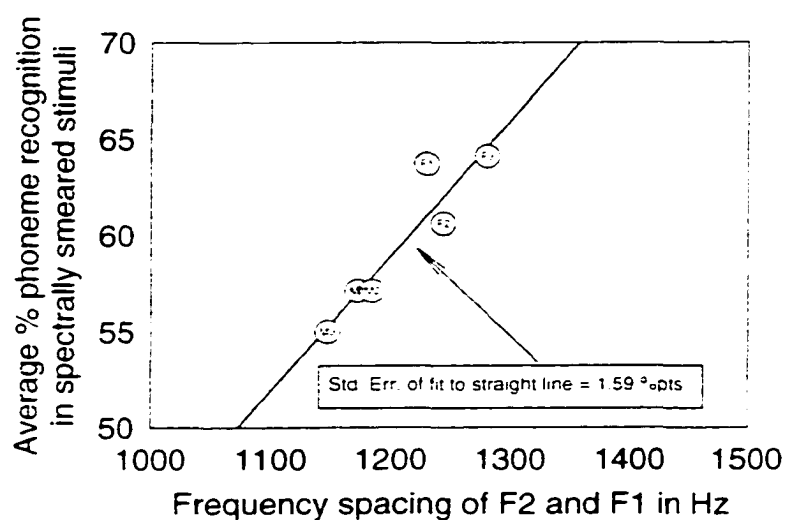


Figure 5.2. Phoneme recognition in smeared speech as a function of the frequency spacing of the first and second formants, for 3 female and 3 male talkers. Also shown is the least-squares fit to a linear relationship.

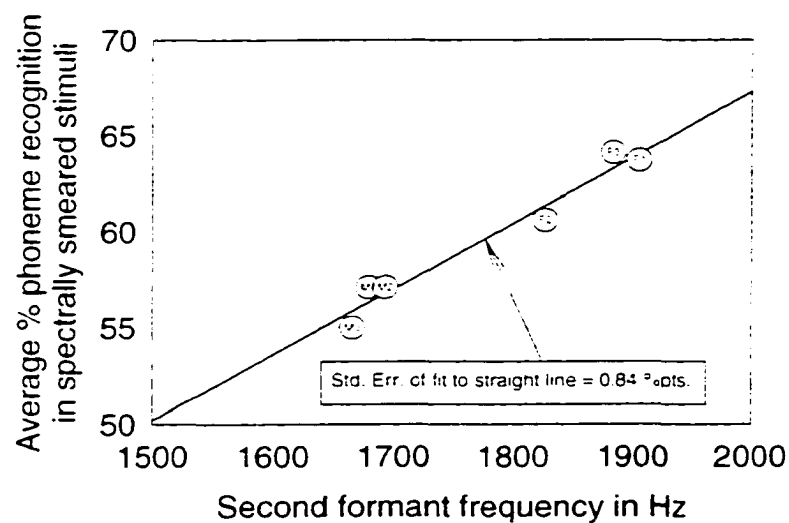


Figure 5.3. Phoneme recognition in smeared speech as a function of the frequency of the second formant, for 3 female and 3 male talkers. Also shown is the least-squares fit to a linear relationship.

#### 5.4 Discussion

The formant data are in keeping with the basic assumptions on which this study was based - namely, that the average formant frequencies and, therefore, the average formant spacing would be higher in female than in male talkers. Moreover, the mean frequencies for formants 1 through 3 are similar to those reported in the literature.

In the correlation analysis, the mean formant spacing measured between formants 1 and 2 was a good predictor of smeared phoneme recognition - supporting the original hypothesis that spectral smearing interferes with speech perception because it obscures formant patterns. The best single predictor of smeared phoneme recognition was not a measure of spacing but the average frequency of F2. With so few data points, however, it would be a mistake to attach too much importance to this observation unless it can be replicated in more detailed studies. Note, also, that the range of F2 was greater than the range of the F2/F1 spacing, that errors in the measurement of F2 are less likely than errors in the measurement of F1, and that any errors of measurement affect the accuracy of a difference more than the accuracy of each measure.

#### 5.5. Summary

These data confirm that formant frequencies and formant spacing were higher in this sample of female talkers than in this sample of male talkers. Of the measures of mean formant spacing, the difference between F2 and F1 was the best predictor of phoneme recognition in the smeared speech. The standard error of fit to a linear relationship was

1.6 percentage points. In these data, the frequency of F2 was a somewhat better predictor of phoneme recognition than was the spacing of F1 and F2 but because of the small sample size, possible errors in the measurement of formant frequencies, and range effects, generalization is not warranted at this stage..

## CHAPTER 6 DISCUSSION AND CONCLUSIONS

### 6.1. Phoneme recognition data

#### 6.1.1. General effect of smearing

The general effects of spectral smearing on phoneme recognition observed in the present study are in keeping with those reported by Boothroyd et al. (1996), using the same smearing method and the same test materials. For example, extrapolation of the group mean phoneme recognition scores, as shown in Figure 4.4, suggests that phoneme recognition begins to fall when the smeared bandwidth is in the region of 300Hz and approaches zero for a smeared bandwidth of 20,000 Hz. In other words, uniform spectral smearing, as used in the present study, begins to have an effect on phoneme recognition when the smeared bandwidth approximates the typical formant bandwidth but is not fully effective until all spectral information is removed.

#### 6.1.2. Smearing thresholds

The smeared bandwidth giving 50% phoneme recognition for the average male talker in the present study was 1800 Hz. This value is somewhat higher than the value of 1400 Hz reported for the single male talker in the Boothroyd et al. study. It will be recalled, however, that a strong learning effect was observed in the present study, amounting to around 15 percentage points between the first and last session. No such effect was observed in the Boothroyd et al. study - presumably because subjects listened to only one talker, with very little repetition of test lists. Subtracting 7.5 percentage points from the average scores in the present study provides an estimate of performance at the first test

session. The slope of the performance versus smearing function was 17.8 percentage points per octave. Taken together, these two values predict that the smearing threshold should be 0.42 octaves lower than 1800 Hz in the absence of a learning effect. The resulting value is 1345 Hz, which is in much closer agreement with the Boothroyd et al. data.

### 6.1.3. Slope of the smearing function

In the present study, mean phoneme recognition fell by 17.8 percentage points per octave between smeared bandwidths of 1000 and 2000 Hz. The corresponding value from the Boothroyd et al. study is 20 percentage points per octave. The difference cannot be attributed to the learning effect involved in the present study because there was no evidence of an interaction between smearing and replication. It is, however, a relatively small difference and may be attributable to differences of talkers and listeners in the two studies. In general, it can be concluded that the present findings relating phoneme recognition to the amount of smearing are in keeping with comparable data from the literature.

## 6.2. Replication

### 6.2.1. Causes of replication effect

The presence of a strong replication effect in the present study is of concern. Because subjects listened to 6 talkers, under 3 smearing conditions, over 4 replications, this design required the presentation of 72 lists. Because only 12 lists were used for testing, any single word was heard, on average, 6 times during the study. Even allowing for the

fact that some of the presentations were under conditions of fairly severe distortion, there is a strong possibility that part of the replication effect was the result of increasing familiarity with the test words. An additional factor could be the considerable experience provided to these subjects in listening to smeared speech. As Howard-Jones and Summers (1992) pointed out when discussing the results of their study, "Subjects' difficulties may be partly due to dealing with an unfamiliar form of speech. Hence, these measurements probably give an underestimate of the information content of the test material." The present findings support this suggestion.

#### 6.2.2. Implications of replication effect

Fortunately, there was no evidence in the present study that replication had a differential effect on the various smearing levels, the two gender groups, or on the interaction between gender and the amount of smearing. The replication effect does not, therefore, compromise the principal purpose of this study, which was to determine the effect of talker gender on resistance to spectral smearing. When using spectral smearing to simulate the psychoacoustic effects of sensorineural hearing loss, however, the differences between normally hearing subjects listening to degraded speech for the first time and hearing-impaired subjects with long-term experience and acclimatization must be taken into consideration.

### 6.3. Phoneme differences

#### 6.3.1. Vowels versus consonants

Vowel recognition was more affected by smearing than was consonant recognition. In

addition, vowel recognition decreased more rapidly with increased smearing than did consonant recognition. Both findings are in keeping with the data reported by Boothroyd et al., using the same smearing method and test materials. Two explanations can be proposed for the vowel-consonant difference. First, spectral cues play the major role in vowel identification, whereas both spectral and temporal cues are of importance in consonant identification. It follows that spectral smearing should be more damaging to vowel recognition. Second, cues to vowel identity are contained mainly in the lower frequencies whereas cues to consonant identity often spread into the higher frequencies. The present study employed uniform smearing - that is, smeared bandwidth in Hz was the same at all frequencies. When expressed in terms of percent or octaves, therefore, the amount of smearing decreased with increasing frequency. Frequency resolution of the normal ear is relatively independent of frequency when expressed in percent or octaves. It is probable, therefore, that the effective smearing in the present study decreased with increasing frequency - in which case, one would expect consonant recognition to be less affected than vowel recognition. This last explanation, however, is not supported by the results of ter Keurs et al. (1992) who also observed the vowel-consonant difference when using logarithmic smearing (i.e., smeared bandwidth was proportional to frequency). It seems probable, therefore, that the greater resilience of consonants to spectral smearing reflects the greater role of temporal cues in consonant identification.

### 6.3.2. Initial versus final consonants

The present study did not reveal a difference between initial and final consonants in terms of resistance to smearing. In this respect, the findings differ from those of

Boothroyd et al. (1996) in which final consonants were found to be more resistant to smearing than were initial consonants. It should be noted, however, that the Boothroyd et al. study used only a single male talker. It is possible that there are individual talker differences in the interaction between smearing and consonant position<sup>1</sup>. The present findings point out the danger of generalizing speech perception findings that are based on the speech of a single talker and leave open the question of the differential effects of smearing on initial and final consonants in consonant-vowel-consonant words.

#### 6.4. The gender effect

##### 6.4.1. Existence of a gender effect

The main purpose of the present study was to test the prediction that the speech of female talkers would be more resistant to uniform spectral smearing than the speech of male talkers. This prediction was clearly supported. The main effect of gender was significant at the 1% level. Moreover, the fact that this effect could be demonstrated with only 3 subjects in each gender group implies that the effect is very robust.

##### 6.4.2. Magnitude of the gender effect

In terms of phoneme recognition, the mean difference amounted to only 6.5 percentage points. Because of the slope of the function relating phoneme recognition to smeared bandwidth, this difference amounts to a change of 28% or 0.35 octaves in smearing threshold (the smeared bandwidth giving 50% phoneme recognition). It should be noted, however, that the confidence limits of the male-female difference are quite large. In the

---

<sup>1</sup> In fact, a simple 4-way analysis of variance in the raw data, in which talker is treated as a simple factor at 6 levels shows a significant interaction between talker and phoneme position ( $F[5,60]$ ,  $p=0.0002$ ).

analysis of variance of Table 4.1. the error term for the gender effect, reflecting the variability within gender groups, is  $93\% ^2$  (percentage points to the power 2). The mean for each gender group is based on 108 observations - giving a standard error for the difference between two means of 1.3 percentage points. Multiplying this standard error by the value of  $t$  for 4 degrees of freedom (2.78) provides 95% confidence limits of  $\pm 3.6$  percentage points. In other words, it can be stated with confidence that the average male-female difference in the population represented by this sample of talkers lies between 2.9 and 10.1 percentage points. The corresponding confidence limits for smearing threshold are 0.16 to 0.57 octaves (or 12 to 49%).

#### 6.4.3. Gender effect and phoneme position

Although the gender effect was clearly demonstrated in the present study, more detailed analysis showed an interaction with phoneme position. In post hoc analysis, the gender effect was only convincingly demonstrated for the vowel and the final consonant in CVC words. The absence of a significant gender effect for initial consonants is interesting and may be related to a difference in the balance of temporal versus spectral cues to the identity of pre- and post-vocalic consonants. Further investigation of this possibility is warranted.

#### 6.5. Cause of the gender effect

##### 6.5.1. Initial hypothesis

The prediction of a gender difference was based on the hypothesis that spectral smearing degrades phoneme recognition by obscuring the formant patterns in speech. Because the

average formant spacing is some 20% higher in the speech of women than in the speech of men. It was predicted that the smearing threshold would be some 20% higher for female talkers than for male talkers. To the extent that the gender effect was found, and that its confidence limits include the predicted magnitude, the present findings offer support for the initial hypothesis.

#### 6.5.2. Support from absolute formant data

In fact, the average formant difference between the female and male talkers used in the present study was estimated to be 16% or 0.22 octaves. This value is somewhat lower than the average female-male difference in smearing threshold. It should be noted, however, that formant estimates were based only on the temporal centers of a limited selection of vowels. Note, also, that cues to vowel and consonant identity are also present in the formant patterns within consonantal segments and in the transitional regions of vocalic segments. It is inappropriate, therefore, to expect an exact quantitative agreement between the behavioral data and the formant data as measured in the present study.

#### 6.5.3. Support from correlational data

The fact that the frequency spacing between the first and second formants was a good predictor of phoneme recognition in smeared speech offers support for the underlying hypothesis that the effects of spectral smearing are caused by the loss of access to formant patterns. It was found, however, that the average frequency of an individual's second formant was at least as good as, and possibly better than, any estimate of formant spacing as a predictor of resistance to smearing. This finding may reflect the fact that the

frequency of the second formant, and its pattern of change over time, is the single most informative acoustic cue in the speech signal. This cue provides the listener with information about the position of the tongue in relation to the front and back of the mouth and, therefore, about place of articulation. Numerous studies have shown that the range of frequencies most important in phoneme recognition corresponds with the range of the second formant. As discussed in Chapter 5, however, it would be a mistake to attach too much significance to the finding that the absolute frequency of F2 was a better predictor of smeared phoneme recognition than the relative values of F1 and F2. The subject sample in this study was very small and both variables were, as expected, bimodally distributed. There are also procedural and statistical factors that could account for the observed difference. Further work on this issue is indicated.

#### 6.5.4. Perceptual spectral smearing

As shown in appendix D, one effect of spectral smearing is to create a single spectral peak from two adjacent spectral peaks – i.e., a to merge formants. Exactly how a listener might interpret the results is not clear, but there is evidence in the literature that such merging already occurs at the perceptual level. (See Strange, 1999, for a brief review of this work.) When asked to match natural vowels to synthetic vowels containing only one or two formants, listeners appear to average formants 1 and 2 in back vowels, where the F1/F2 spacing is low, and to average the second and third formants in front vowels where the F2/F3 spacing is low. These findings may help explain why studies of spectral smearing have consistently found that quite large amounts of smearing are needed to have a major effect on phoneme recognition. They do not, however, undermine the initial

hypothesis that the primary effect of spectral smearing is to reduce access to the formant patterns of speech.

#### 6.5.5. Other gender differences

The prediction that female speech would be more resistant to the effects of spectral smearing than male speech was based on the theory that the negative effects of the smearing arise from the loss of access to the formant patterns in the acoustic speech signal. To the extent that the prediction was realized, the findings support the theory. To put it another way, the data are consistent with the conclusion that female speech is more resistant to spectral smearing than male speech because the formant frequencies in female speech are more widely spaced. At this stage of the research, however, it is not possible to accept this conclusion with confidence. In the previous paragraph, for example, it was suggested that female speech might be more resistant to the effects of smearing than male speech because, a) the second formant frequency in female speech is higher than in male speech, and b) the particular algorithm used in this study produced less effective smearing at higher frequencies. One must also recognize that there are other differences between male and female speech than the average values and spacing of the vocal tract formants. An obvious difference is that of voice fundamental frequency. Other known differences include the source spectrum (Monsen and Engebretson, 1977; Klatt and Klatt, 1990), the long-term average spectrum (Boothroyd, Erickson, and Medwetsky, 1994; Levitt and Webster, 1991), and the spectral content of the consonant /s/ (Boothroyd and Medwetsky, 1992). Clearly, further work will be required to determine exactly what characteristics of the acoustic speech signal are responsible for

individual differences in resistance to the effects of spectral smearing.

## 6.6. Further research

### 6.6.1. Smearing algorithms

The present study should be replicated with logarithmic smearing, in which the smeared bandwidth in Hz is proportional to frequency. The finding that the gender effect is still present would support the original hypothesis regarding the importance of formant spacing. If, however, the gender effect is eliminated, this would lend support to the importance of second formant frequency. Independently of any gender effect, it would be valuable to measure frequency resolution as a function of frequency in hearing-impaired subjects and to replicate those functions artificially so as to determine their effects on speech perception in normally hearing subjects. The purpose of these studies would be to determine the role played by the specific smearing algorithm when measuring the effects of smearing

### 6.6.2. Speech processing

In the present study, the gender effect was examined because it provided a natural difference of average formant spacing. As mentioned earlier, however, several other acoustic effects accompany the gender difference. In a continued search for the acoustic correlates of resistance to spectral smearing, it will be important to introduce changes in the speech of individual talkers. With modern digital processing techniques it is possible to modify fundamental frequency and formant frequencies independently. Moreover, one can shift formant frequencies linearly or logarithmically. It is also possible to create

replicas of the speech signal in which formant bandwidths are dramatically reduced - potentially increasing resistance to spectral smearing. The purpose of these studies would be to establish which aspects of the acoustic speech signal are most responsible for resistance to the effects of spectral smearing and, by implication, to the effects of sensorineural hearing loss.

#### 6.6.3. Studies with hearing-impaired subjects

If it can be shown that there are certain manipulations of the acoustic speech signal that can increase resistance to spectral smearing, without compromising overall intelligibility, it will be important to test them on subjects with sensorineural hearing loss. As pointed out in the introduction, frequency resolution is only one psychoacoustic ability that is affected by sensorineural hearing loss. Other known abilities include temporal resolution, loudness perception, intrastimulus masking, and forward masking. The purpose of these studies will be to determine the relative importance of loss of spectral resolution and other psychoacoustic consequences of sensorineural hearing loss.

#### 6.6.4. Intra-talker effects

As a follow-up to the present gender-based study the effect of formant spacing on resistance to smearing should be tested within the speech of individual talkers. One would predict, for example, that front vowels, with widely spaced first and second formants would be more resistant to smearing than back vowels. Such a finding is not, however, a foregone conclusion because of the previously mentioned perceptual spectral smearing that appears to be a part of normal vowel perception (Strange, 1999).

## 6.7. Clinical implications

### 6.7.1. Speech perception testing

The present findings add to a growing body of data that point to the dangers of using a single talker when measuring speech perception abilities in hearing-impaired subjects. Differences in the acoustic properties of the talker's speech may well interact with the psychoacoustic deficits of the hearing-impaired listener. In fact, it is a common complaint of hearing-impaired adults that they can understand certain types of talkers better than others.

### 6.7.2. Psychoacoustic evaluation

If we can establish clear relationships between the acoustic properties of the talker and the psychoacoustic abilities of the listener, there may be justification for expanding the scope of clinical evaluation to include such things as frequency resolution. Clearly, the practice of audiology has not suffered by limiting itself mainly to measures of pure-tone threshold, speech recognition, and middle ear function, sometimes supplemented by loudness measures. But, as hearing aids become increasingly sophisticated, it is probable that equally sophisticated advances will be needed in clinical practice.

### 6.7.3. Teachers of hearing-impaired children

To the extent that the smearing algorithm used in this study provides a reasonable simulation of the more important effects of sensorineural hearing loss, the finding that female speech is more resistant to spectral degradation than is male speech is

encouraging - most of the teachers of hearing-impaired children are women. It is clear, however, that a lot more work is needed before we can say exactly what aspects of the acoustic speech signal are most desirable for the educational environment.

#### 6.7.4. Amplification

Taken at face value, the present findings support the idea that speech intelligibility could be improved for hearing-impaired listeners by artificially modifying the acoustic speech signal in ways that extend beyond the current options for frequency-dependent gain (selective amplification) and input-dependent gain (compression amplification).

#### 6.8. Conclusions

1. The speech of female talkers is more resistant to the effects of uniform spectral smearing than is the speech of male talkers.
2. It can be concluded with 95% confidence that the smearing threshold (here defined as the smeared bandwidth giving 50% phoneme recognition in consonant-vowel-consonant words) is between 0.16 and 0.57 octaves higher in female talkers than in male talkers. A more precise estimate will require data from a larger sample of talkers.
3. The data support the hypothesis that female speech is more resistant to the effects of spectral smearing than is male speech because the formants in female speech are more widely spaced.
4. Further research will be needed: a) to establish whether the present findings are specific to the uniform smearing algorithm used in this study, and n) to determine exactly which properties of the acoustic speech signal are most responsible for

resistance to spectral smearing.

APPENDIX A  
SUBJECT CONSENT FORM

AUDITORY SPEECH PERCEPTION MEASUREMENT  
PRINCIPAL INVESTIGATOR: ANGELA CHUNG-I LI  
FACULTY ADVISOR: ARTHUR BOOTHROYD, PH.D.

The present experiment is designed to measure speech perception in normally hearing individual using acoustically processed words. During the experiment you will be asked to listen to sets of recorded monosyllabic words, sometimes the words are easy to hear, and other times the words are difficult to hear. In each stimulus presentation, you will be asked to repeat what you hear and write down each word on the response form. There will be two test sessions and each will last approximately 1 1/2 hours.

While the results of this study may be reported at professional meetings and/or may be published, subject anonymity will be protected.

Please note that the Graduate School has a committee that safeguards your interests. This committee requires us to point out that the research done here is of an experimental nature and is done in a university and not a clinical setting.

If you have any questions regarding your participation in this project, please feel free to ask me or to contact my advisor, professor Arthur Boothroyd, Ph.D. (Tel: (212) 642-2352).

-----

I agree to participate in the experimental investigation described above. I understand that my participation is voluntary and that I may withdraw at any time without penalty.

PRINT NAME: \_\_\_\_\_

SIGNATURE: \_\_\_\_\_

DATE: \_\_\_\_\_

**APPENDIX B**  
**INSTRUCTIONS TO SUBJECTS**

1. The experiment will be divided into two sessions, each session will take about one and half hours.
2. The second session will take place at least one week after the completion of the first session.
3. At the beginning of the first session, you will be given a hearing test for both ears. You will need to raise your hand when you hear a beep (even it is very soft).
4. You will hear a list of words. Some words are easy to hear, and others are difficulty to understand. You are asked to repeat and write down each word you hear. Please do not hesitate to guess if you are not sure of the word. No word will repeat twice.
5. When you finish responding to each word, the next word will be presented to you.
6. You will be given answer sheets to record your responses.
7. The experimenter will not provide you with any feedback on your performance.

APPENDEX C  
AB ISOPHONEMIC WORD LISTS

LIST 1	LIST 2	LIST 3
FISH	FUN	FILL
DUCK	WILL	CATCH
PATH	VAT	THUMB
CHEESE	SHAPE	HEAP
RACE	WREATH	WISE
HIVE	HIDE	RAVE
BONE	GUESS	GOT
WEDGE	COMB	SHOWN
LOG	CHOOSE	BED
TOMB	JOB	JUICE
LIST 4	LIST 5	LIST 6
BADGE	BATH	HUSH
HUTCH	HUM	GAS
KILL	DIG	THIN
THIGHS	FIVE	FAKE
WAVE	WAYS	CHIME
REAP	REACH	WEAVE
FOAM	JOKE	JET
GOOSE	NOOSE	ROB
NOT	POT	DOPE
SHED	SHELL	LOSE

## APPENDIX C (contd.)

LIST 7	LIST	LIST 9
MATH	HUG	FIB
HIP	DISH	THATCH
GUN	BAN	SUM
RIDE	RAGE	HEEL
SIEGE	CHIEF	WIDE
VEIL	PIES	RAKE
CHOSE	WET	GOES
SHOOT	COVE	SHOP
WEB	LOOSE	VET
COUGH	MOTH	JUNE
LIST 10	LIST 11	LIST 12
JUG	KISS	SHIP
LATCH	BUZZ	RUG
WICK	HASH	FAN
FAITH	THIEVE	CHEEK
SIGN	GATE	HAZE
BEEP	WIFE	DICE
HEM	POLE	BOTH
ROD	WRETCH	WELL
VOTE	DODGE	JOT
SHOES	MOON	MOVE

## APPENDIX C (Contd.)

LIST 13	LIST 14	LIST 15
WISH	THUG	HAVE
DUTCH	WITCH	WIG
JAM	TEAK	BUFF
HEATH	WRAP	MICE
LAZE	VICE	TEETH
BIKE	JAIL	JAYS
ROVE	HEN	POACH
PET	SHOWS	RULE
FOG	FOOD	DEN
SOON	BOMB	SHOCK

APPENDIX D  
SPECTROGRAPHIC EXAMPLES OF SMEARING

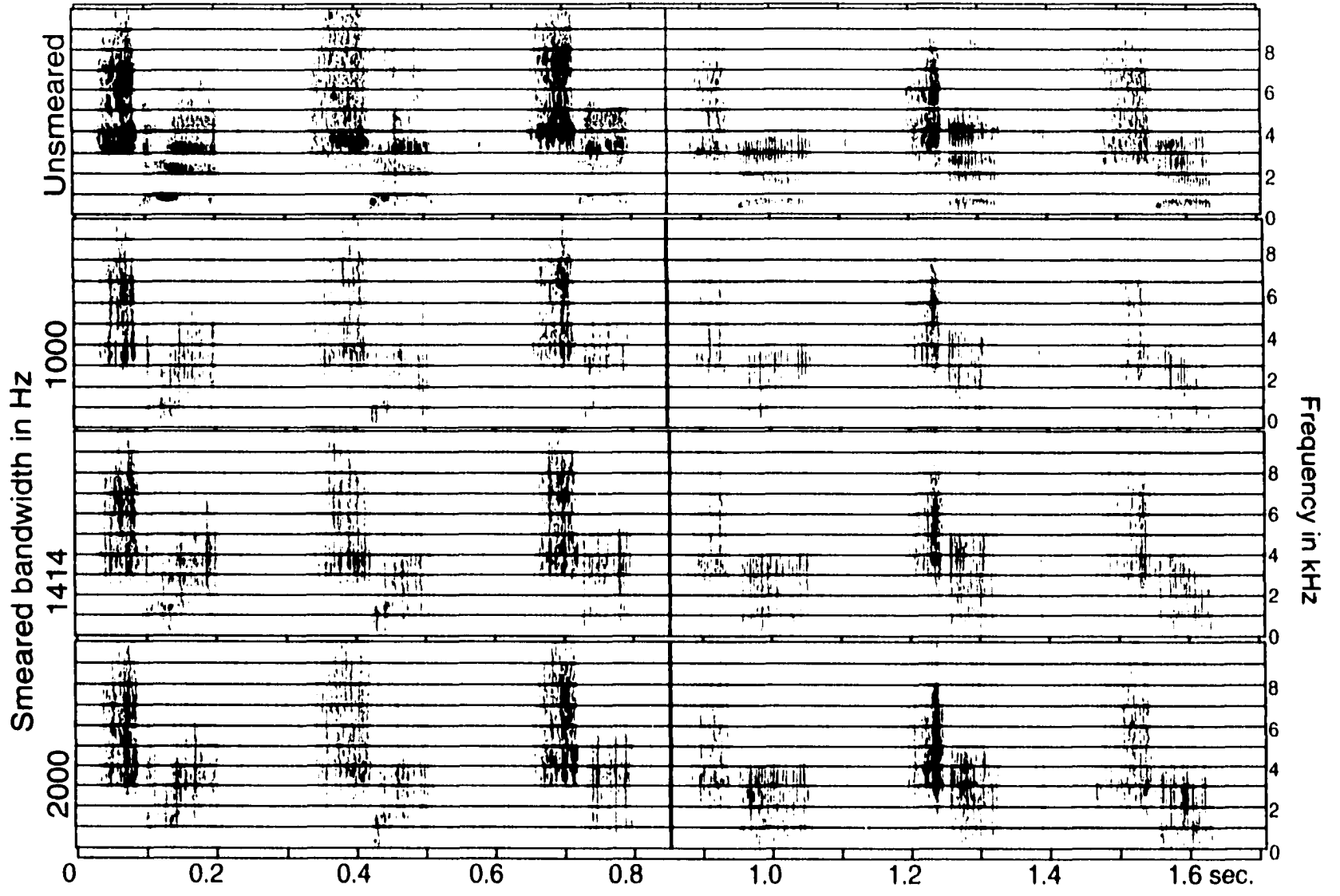


Figure D1. Spectrograms of the word "shed", spoken by 3 female (left) and 3 male (right) talkers, unprocessed (top), and processed with smeared bandwidths of 1000, 1414, and 2000 Hz.

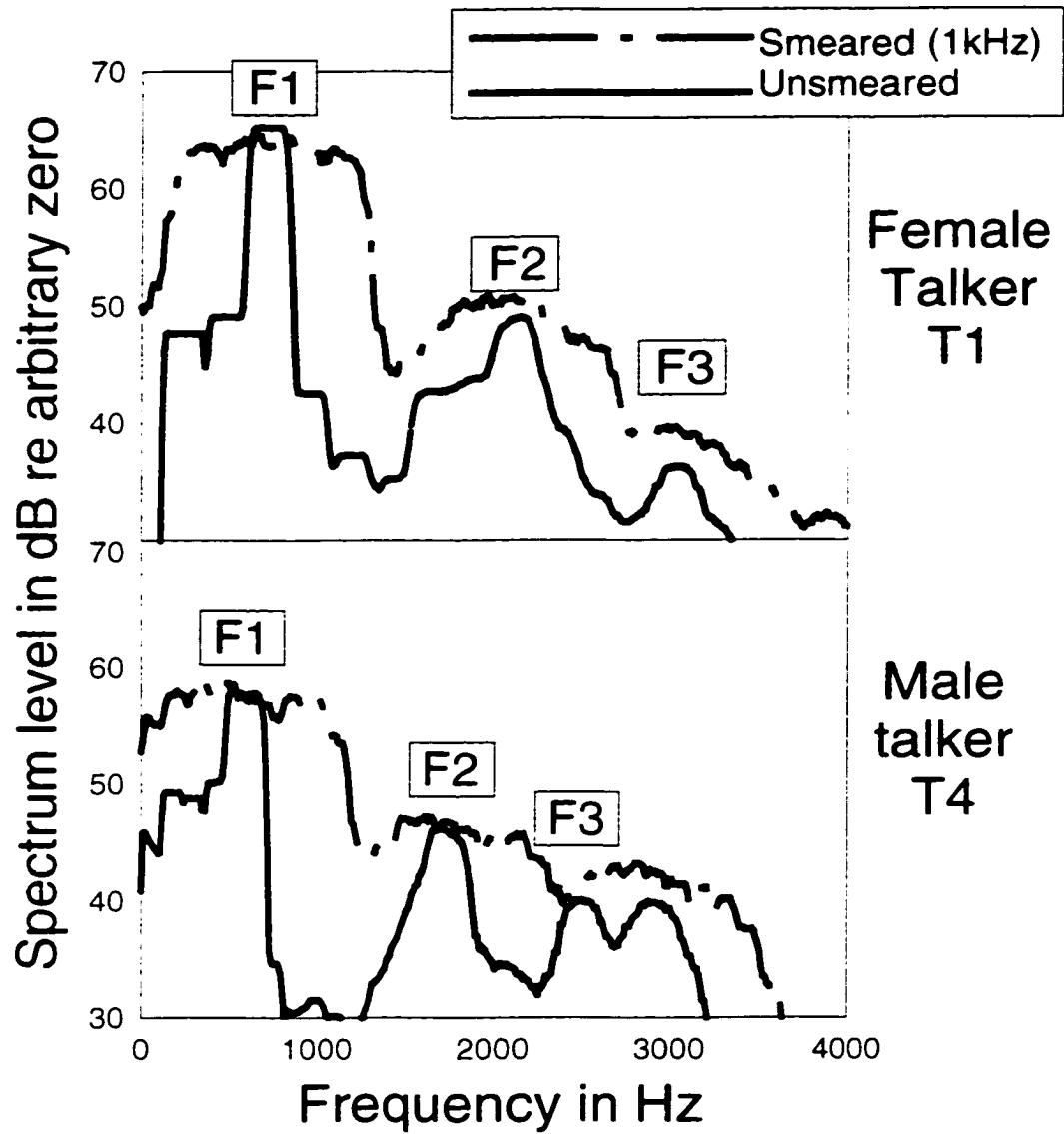


Figure D2. Examples of the effect of smearing, with a smeared bandwidth of 1000Hz, on the spectrum of the vowel / $\epsilon$ / taken from recordings of the word "shed" spoken by a female and a male talker.

## APPENDIX E DADISP WORKSHEET

The commands in the Dadisp worksheet used for smearing the sound files are described as follows:

Name of worksheet: Smearing

**Window 1: ReadB(A:\THIGHS.wav",3)|setdeltax(1/22050)**

Remarks:

Read in the unsmeared sound file from A drive as a signed integer file. then set the sampling rate to 22050 points per second.

**Window 2: Highpass(22050.0.150.0,3.0.30.0,1.0)**

Remarks:

Create a highpass filter, with sampling rate of 22050 points per second, and given pass band edge = 150 Hz

stop band edge = 1 Hz

attenuation in stop band = 30 dB

ripple in pass band = 3 dB

The purpose is to remove any low frequency noise taken on the original recordings.

**Window 3: Filter(extract(W1.23,length(W1)-22),W2)|setxoffset(0)**

Remarks:

Remove 22 points of header from file in Window 1. then filter with Window 2. Renumber the points, starting with 0.

**Window 4: Ravel(W3,500,1,250)**

Remarks:

Divide the file in Window 3 into sections of 500 points long (23 msec), starting with point 1, and each section overlaps the previous one by 250 points.

The purpose is to rearrange the single column of points. Instead of having a single column of numbers, the file now has many columns, each column is 500 points long. For example, if a word was 1 second long with 22050 points, now it has roughly 88 columns, each has 500 points long.

**Window 5: Extract(interpolate(transpose(sqrt(colmean(W4^2))),250),-250,length(W3))|setdeltax(1/22050)|setxoffset(0)|fmax**

Remarks:

This command works inside out of each bracket, beginning with  $W4^2$ : take square values in the table of  $W4$ .

COLMEAN: find the mean of each column.

SQRT: find square root of each mean. The result is R.M.S. value of each column which has 500 points per section. A table with only 1 row is obtained, consisting of 88 means.

TRANSPPOSE: put the 88 values into a single column.

INTERPOLATE(---, 250): replace each of the 88 values with 250 of the original points. To provide the same number of points as the original wave.

EXTRACT (---,-250.LENGTH(W3)): shift the starting point by 250 points to line up with the original wave same as W3.

Then re-set the time interval from one sample to the next (sampling rate of 22050), renumber the points starting at 0, find the max value (peak) and put an invisible cursor at it.

**Window 6: (Clip(W5,max(W5)/2,max(W5)/2+1)-max(W5)/2)/fmax**

Remarks:

Create a R.M.S. envelope which is between 0 and 1. Check W5 for any value below 50% (= 6 dB below peak to peak R.M.S. value) of maximum is set to 0, and any value above is set to 1. Put an invisible cursor at the beginning of the square envelope.

**Window 7: Deriv(W6)/fmin**

Remarks:

Identify the start and end of the square envelope in W6, and set invisible cursor to the minimum(i.e., the end).

**Window 8: Extract(W3,curpos(W6),curpos(W7)-curpos(W6))**

Remarks:

Extract that portion of the waveform from W3 for which the R.M.S. amplitude is within 6 dB of the peak value.

**Window 9:**

**W3/stdev(W8)\*7000loverplot(gline(length(W3),1/22050,0.2^15))loverplot(gline(length(W3),1/22050,0.-(2^15))loverplot(W6\*5000)**

Remarks:

Set the R.M.S. value of the section defined in W8 to 7000.

Overplot W3 between  $2^{15}$  and  $-2^{15}$  to show that digital limits are not exceeded.

Overplot clipped amplitude envelope in W6 for visual confirmation of its appropriate location.

**Window 10: Lpnoise.1.354**

Remarks:

Read in low-pass filtered noise of 354 Hz which was previously saved as a series in a dataset).

**Window 11: Lpnoise.1.500**

Remarks:

Read in low-pass filtered noise of 500 Hz.

**Window 12: Lpnoise.1.707**

Remarks:

Read in low-pass filtered noise of 707 Hz.

**Window 13: Lpnoise.1.1000**

Remarks:

Read in low-pass filtered noise of 1000 Hz.

**Window 14: W9\*Extract(W10,1,length(W3))\*2**

Remarks:

Multiply waveform from W9 (which has vowel R.M.S. = 7000) by 354 Hz LP noise (which is cut to the same length of the word), double its amplitude.

**Window 15: W9\*Extract(W11,1,length(W3))\*2**

Remarks:

Multiply waveform from W9 by 500 Hz LP noise, double its amplitude to balance R.M.S. level after smearing.

**Window 16: W9\*Extract(W12,1,length(W3))\*1.6**

Remarks:

Multiply waveform from W9 by 707 Hz LP noise, multiply its amplitude by 1.6.

**Window 17: W9\*Extract(W13,1,length(W3))\*1.4**

Remarks:

Multiply waveform from W9 by 1000 Hz LP noise, multiply its amplitude by 1.4.

**Window 18:**

```
Ravel(concat(gseries(0,max(W9),min(W9),max(W14),min(W14),max(W15),min(W15),max(W16),min(W16),max(W17),min(W17),0),gline(12,1,0,2^15),gline(12,1,0,-(2^15))),12)
```

Remarks:

Generate a zig-zag line from maxima to minima of the smeared and unsmeared files to confirm that digital limits of  $\pm 2^{15}$  have not been exceeded.

Finally, a MACRO called "save" saves the smeared (W14, 15, 16, 17) and unsmeared (W9) files as signed integer.

For example:

S35 writeb("word.35".3,W14) --- word smeared at LP noise 354 Hz has been saved as signed integer digital file in Dadisp program.

S50 writeb("word.50".3,W15) --- word smeared at LP noise 500 Hz has been saved as signed integer file.

S70 writeb("word.70".3,W16) --- word smeared at LP noise 707 Hz has been saved as signed integer file.

S1K writeb("word.1K".3,W17) --- word smeared at LP noise 1K Hz has been saved as signed integer file.

S00 writeb("word.00".3,W9) --- unsmeared word has been saved as signed integer file.

All files with extension of smearing values were renamed in DOS. For example, thighsX.35, where thighs was the word, X was the talker number, and 35 was the smearing value for 354 LP noise. Individual directory was created for each talker and smearing value, using talkers' initials + smearing value, such as: md JO35 ( to make a directory for JO with smearing condition of 354 LP noise), md JO50 ( to make a directory for JO with smearing condition of 500 LP noise),

md JO70 ( to make a directory for JO with smearing condition of 707 LP noise),

md JO1K ( to make a directory for JO with smearing condition of 1K LP noise),

md JO00 ( to make a directory for JO with non-smearing condition).

Repeated the same procedure for the other 6 talkers NG35, NG50, NG70, NG1K, NG00; BM35, BM50, BM70, BM1K, BM00; TC35, TC50, TC70, TC1K, TC00; CO35, CO50, CO70, CO1K, CO00; FS35, FS50, FS70, FS1K, FS00; DR35, DR50, DR70, DR1K, DR00.

All files were moved into an appropriate directory by separately moving 3-letter words, 4-letter words, 5-letter words, and 6-letter words of one talker and one smearing value. Each directory should contain 150 word files. In each directory, 150 files were renamed from \*.35 to \*. (to remove extension of smearing value).

These renamed files were then converted to sound files using a program "Addhead.exe" written by Eddy Yeung. Headers were added to 150 files in each directory by running "Addhead \*. 22050 1". Repeated the same procedure for individual directory.

**APPENDIX F  
COUNTERBALANCING TEST SEQUENCE**

LISTENER	TALKER	SMEARING LEVELS (+/-) IN HZ											
		500	707	1000	500	707	1000	500	707	1000	500	707	1000
L1	T1	13	8	7	10	4	15	9	12	5	11	14	6
	T2	11	6	5	8	14	13	7	10	15	9	12	4
	T3	8	15	14	5	11	10	4	7	12	6	9	13
	T4	14	9	8	11	5	4	10	13	6	12	15	7
	T5	5	12	11	14	8	7	13	4	9	15	6	10
	T6	15	10	9	12	6	5	11	14	7	13	4	8
L2	T2	12	7	6	9	15	14	8	11	4	10	13	5
	T3	9	4	15	6	12	11	5	8	13	7	10	14
	T4	7	14	13	4	10	9	15	6	11	5	8	12
	T5	6	13	12	15	9	8	14	5	10	4	7	11
	T6	10	5	4	7	13	12	6	9	14	8	11	15
	T1	4	11	10	13	7	6	12	15	8	14	5	9
L3	T4	10	11	7	9	14	8	12	13	6	15	4	5
	T5	7	8	4	6	11	5	9	10	15	12	13	14
	T6	6	7	15	5	10	4	8	9	14	11	12	13
	T1	8	9	5	7	12	6	10	11	4	13	14	15
	T2	15	4	12	14	7	13	5	6	11	8	9	10
	T3	11	12	8	10	15	9	13	14	7	4	5	6
L4	T3	9	10	6	8	13	7	11	12	5	14	15	4
	T2	13	14	10	12	5	11	15	4	9	6	7	8
	T1	12	13	9	11	4	10	14	15	8	5	6	7
	T6	4	5	13	15	8	14	6	7	12	9	10	11
	T5	5	6	14	4	9	15	7	8	13	10	11	12
	T4	14	15	11	13	6	12	4	5	10	7	8	9

## APPENDIX F (Continued)

L5	T1	11	7	10	4	13	12	6	9	8	15	14	5
	T3	15	11	14	8	5	4	10	13	12	7	6	9
	T5	13	9	12	6	15	14	8	11	10	5	4	7
	T2	14	10	13	7	4	15	9	12	11	6	5	8
	T4	5	13	4	10	7	6	12	15	14	9	8	11
	T6	9	5	8	14	11	10	4	7	6	13	12	15
L6	T6	6	14	5	11	8	7	13	4	15	10	9	12
	T4	4	12	15	9	6	5	11	14	13	8	7	10
	T2	8	4	7	13	10	9	15	6	5	12	11	14
	T5	7	15	6	12	9	8	14	5	4	11	10	13
	T3	10	6	9	15	12	11	5	8	7	14	13	4
	T1	12	8	11	5	14	13	7	10	9	4	15	6
L7	T2	4	13	12	14	7	15	9	6	5	8	11	10
	T4	12	9	8	10	15	11	5	14	13	4	7	6
	T6	6	15	14	4	9	5	11	8	7	10	13	12
	T1	8	5	4	6	11	7	13	10	9	12	15	14
	T3	14	11	10	12	5	13	7	4	15	6	9	8
	T5	11	8	7	9	14	10	4	13	12	15	6	5
L8	T5	9	6	5	7	12	8	14	11	10	13	4	15
	T3	13	10	9	11	4	12	6	15	14	5	8	7
	T1	7	4	15	5	10	6	12	9	8	11	14	13
	T6	5	14	13	15	8	4	10	7	6	9	12	11
	T4	15	12	11	13	6	14	8	5	4	7	10	9
	T2	10	7	6	8	13	9	15	12	11	14	5	4

## APPENDIX G SUBJECT ANSWER SHEET

Subject:

Age:

Date:

Condition:			
Word	Response	Stimulus	n
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
Total			
Percent			

Condition:			
Word	Response	Stimulus	n
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
Total			
Percent			

Condition:			
Word	Response	Stimulus	n
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
Total			
Percent			

Condition:			
Word	Response	Stimulus	n
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
Total			
Percent			

## APPENDIX H LINEAR PREDICTIVE CODING ANALYSIS

The following analysis parameters were used for spectrogram display:

Spectrogram frame length (bandwidth) --- 323 Hz

Pre-emphasis --- No

Window weighting --- Blackman

Darkness scale --- 18.0

Gain adjustment --- 0.0

Display range --- Percent

Log or Linear display --- linear

Size of cell reduction --- ( x grid : 8 )

For LPC display:

Frame length --- 20 ms

Filter order --- 16

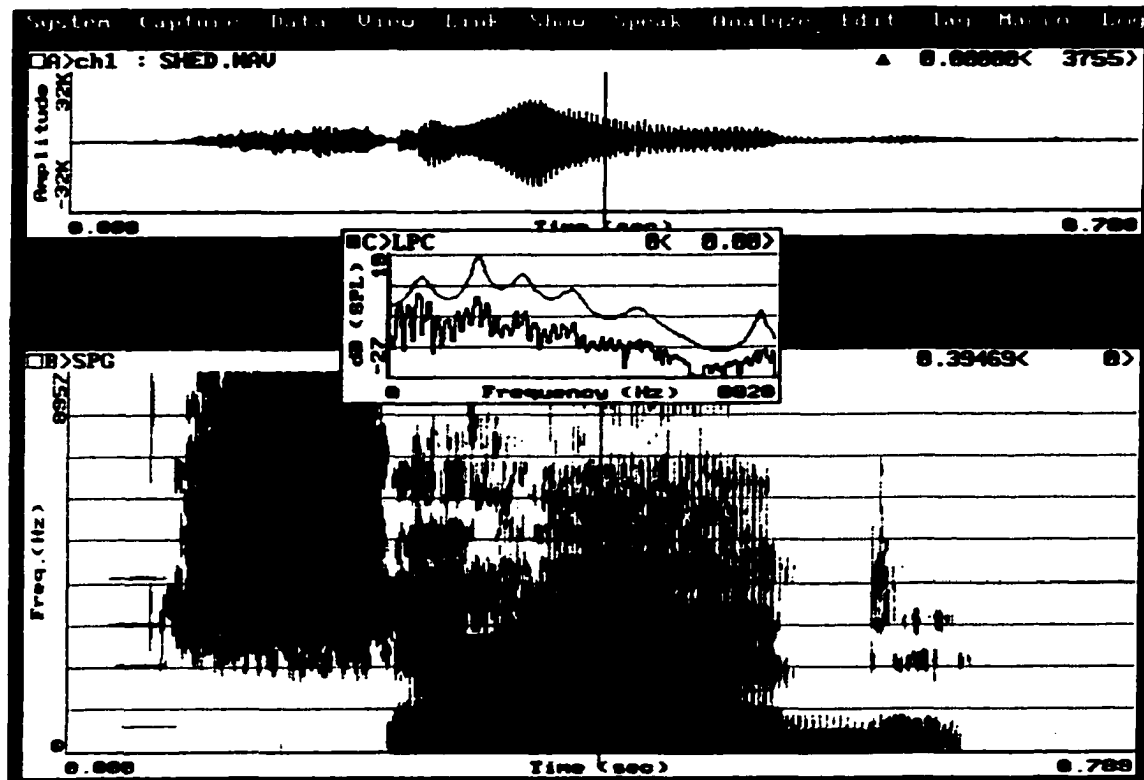
Response pre-emphasis --- yes

Window weighting --- yes

Analysis method --- autocorrelation

Display frequency --- frequency

## Linear Predictive Coding Analysis for Talker 1(Contd.)



RESULTS: LPC FREQUENCY RESPONSE  
 SOURCE: A:\SHED.WAV  
 FRAME SIZE: 20.0 msec  
 PRE-EMPHASIS: 0.900  
 WINDOW WEIGHTING: ON  
 ANALYSIS TYPE: AUTOCORRELATION  
 ANALYSIS START: 0.395 sec  
 COMMENTS:

FORMANT (Hz)	B/WIDTH (Hz)
725.49	272.39
2029.22	149.88
3045.69	341.47
4185.21	416.35
5710.87	579.08
6610.37	1310.04
8506.50	194.83
9850.91	666.61

**APPENDIX I**  
**RAW DATA OF SPECTRAL SMEARING RESULTS (II-PHONEME SCORES)**

FEMALE			LISTENER								AV	MALE			LISTENER								AV
TKR	SMR	REP	L1	L2	L3	L4	L5	L6	L7	L8		TKR	SMR	REP	L1	L2	L3	L4	L5	L6	L7	L8	
T1	1000	1	63	47	77	43	87	63	53	83	64.5	T4	1000	1	60	43	63	50	60	60	60	70	58.3
		2	70	70	63	70	67	47	60	87	66.8			2	57	80	67	57	77	50	60	60	63.5
		3	83	70	73	63	73	83	57	77	72.4			3	73	77	63	80	87	67	60	80	73.4
		4	80	87	70	83	77	87	60	93	79.6			4	70	60	73	63	87	67	73	77	71.3
	AV	74	69	71	65	76	70	58	85	70.8	AV	65	65	67	63	78	61	63	72	66.6			
1414	1000	1	80	50	40	67	73	43	70	70	61.6	1414	1000	1	57	43	57	53	53	53	50	80	55.8
		2	60	53	43	73	83	37	73	63	60.6			2	63	47	40	60	80	43	53	63	56.1
		3	67	60	60	63	77	77	67	67	67.3			3	60	63	33	73	73	50	53	47	56.5
		4	73	83	67	60	90	63	70	90	74.5			4	77	47	60	83	77	53	57	73	65.9
	AV	70	62	53	66	81	55	70	73	66.0	AV	64	50	48	67	71	50	53	66	58.6			
2000	1000	1	63	30	60	30	43	33	40	43	42.8	2000	1000	1	40	53	23	53	30	37	30	47	39.1
		2	53	40	57	53	60	50	63	63	54.9			2	47	47	37	50	40	37	37	67	45.3
		3	63	47	63	60	63	53	60	57	58.3			3	67	43	40	50	60	30	27	60	47.1
		4	57	70	40	63	77	63	50	70	61.3			4	70	43	43	50	63	60	50	47	53.3
	AV	59	47	55	52	61	50	53	58	54.3	AV	56	47	36	51	48	41	36	55	46.2			
T2	1000	1	63	57	70	70	73	57	80	73	67.9	T5	1000	1	63	47	63	53	73	57	40	53	56.1
		2	73	57	67	63	73	47	70	90	67.5			2	60	60	53	63	77	47	53	73	60.8
		3	70	73	63	63	80	63	67	77	69.5			3	77	70	70	63	67	67	53	70	67.1
		4	60	80	87	67	77	83	70	90	76.8			4	53	60	67	70	87	70	63	80	68.8
	AV	67	67	72	66	76	63	72	83	70.4	AV	63	59	63	62	76	60	52	69	63.2			
1414	1000	1	60	40	37	67	57	37	57	67	52.8	1414	1000	1	63	40	50	60	57	60	43	63	54.5
		2	47	60	47	80	57	47	57	73	58.5			2	67	43	50	67	73	40	67	60	58.4
		3	73	70	53	57	60	43	50	60	58.3			3	70	67	50	57	63	43	37	60	55.9
		4	60	67	67	73	83	50	57	73	66.3			4	67	43	43	57	80	63	50	80	60.4
	AV	60	59	51	69	64	44	55	68	58.9	AV	67	48	48	60	68	52	49	66	57.3			
2000	1000	1	50	40	50	33	53	33	47	50	44.5	2000	1000	1	53	27	40	47	43	43	33	40	40.8
		2	43	53	40	47	57	47	50	57	49.3			2	50	33	53	53	50	40	33	50	45.3
		3	60	50	40	57	57	47	53	67	53.9			3	73	73	57	53	73	37	33	57	57.0
		4	80	60	60	50	63	53	73	60	62.4			4	57	57	50	67	77	50	57	67	60.3
	AV	58	51	48	47	58	45	56	59	52.5	AV	58	48	50	55	61	43	39	54	50.8			
T3	1000	1	70	60	63	57	73	53	73	80	66.1	T6	1000	1	40	47	60	63	53	57	57	60	54.6
		2	70	73	80	87	87	70	77	80	78.0			2	63	60	67	43	87	53	63	60	62.0
		3	73	83	80	73	80	60	73	90	76.5			3	80	63	57	60	77	90	57	73	69.6
		4	80	77	77	70	77	83	63	83	76.3			4	87	63	63	77	77	80	57	77	72.6
	AV	73	73	75	72	79	67	72	83	74.2	AV	68	58	62	61	74	70	59	68	64.7			
1414	1000	1	47	47	57	43	67	47	47	60	51.9	1414	1000	1	50	47	40	53	37	27	50	60	45.5
		2	60	67	63	43	73	63	57	70	62.0			2	57	63	50	50	70	47	37	50	53.0
		3	63	77	77	67	80	57	60	77	69.8			3	53	43	53	70	60	53	47	73	56.5
		4	67	80	63	70	80	70	67	77	71.8			4	63	63	67	67	73	67	63	73	67.0
	AV	59	68	65	56	75	59	58	71	63.8	AV	56	54	53	60	60	49	49	64	55.5			
2000	1000	1	43	47	47	50	43	27	47	67	46.4	2000	1000	1	43	33	30	47	20	27	37	30	33.4
		2	43	53	57	63	67	33	60	47	52.9			2	47	40	40	43	53	40	40	37	42.5
		3	57	47	30	50	60	43	63	77	53.4			3	50	50	50	60	47	40	50	53	50.0
		4	73	63	57	53	70	67	63	70	64.5			4	47	60	53	43	60	50	40	73	53.3
	AV	54	53	48	54	60	43	58	65	54.3	AV	47	46	43	48	45	39	42	48	44.8			

(TKR=Talker, SMR=Smeared bandwidth in Hz, REP=Replication, AV=Average)

**APPENDIX I**  
**RAW DATA OF SPECTRAL SMEARING RESULTS (I2-INITIAL CONSONANT SCORES)**

TKR	SMR	REP	LISTENER								AV	TKR	SMR	REP	LISTENER								AV
			L1	L2	L3	L4	L5	L6	L7	L8					L1	L2	L3	L4	L5	L6	L7	L8	
T1	1000	1	40	40	80	50	90	70	60	90	65.0	T4	1000	1	70	70	70	70	60	70	60	80	68.8
		2	70	80	60	80	80	40	60	90	70.0			2	50	90	70	90	90	70	50	70	72.5
		3	**	80	70	60	80	90	50	80	76.3			3	80	90	50	80	90	70	60	80	75.0
		4	90	90	80	70	80	90	50	90	80.0			4	90	90	80	70	**	70	80	70	81.3
		AV	75	73	73	65	83	73	55	88	72.8			AV	73	85	68	78	85	70	63	75	74.4
1414	1	1	80	50	40	70	70	50	80	80	65.0	1414	1	1	50	50	80	50	50	70	50	90	61.3
		2	60	80	40	70	90	70	80	70	70.0			2	90	60	60	70	70	70	80	71.3	
		3	70	50	60	60	70	80	80	60	66.3			3	80	50	50	70	70	70	80	65.0	
		4	70	70	70	50	90	70	60	80	70.0			4	80	50	70	80	60	80	70	80	71.3
		AV	70	63	53	63	80	68	75	73	67.8			AV	75	53	65	68	63	73	68	75	67.2
2000	1	1	70	50	80	30	50	60	30	60	53.8	2000	1	1	60	60	20	60	30	30	30	40	41.3
		2	70	50	50	60	60	70	80	80	65.0			2	40	60	40	60	50	50	50	80	53.8
		3	60	70	70	50	50	70	70	50	61.3			3	70	60	30	50	50	50	20	70	50.0
		4	70	80	40	70	80	60	60	70	66.3			4	80	60	60	60	70	70	60	70	66.3
		AV	68	63	60	53	60	65	60	65	61.6			AV	63	60	38	58	50	50	40	65	52.8
T2	1000	1	50	60	80	70	90	30	80	70	66.3	T5	1000	1	70	60	60	60	80	70	50	60	63.8
		2	50	60	60	50	60	40	80	90	61.3			2	60	80	60	70	80	50	60	80	67.5
		3	70	70	70	70	70	80	60	80	71.3			3	80	90	80	80	80	80	70	70	78.8
		4	80	70	90	70	50	80	60	80	72.5			4	60	70	60	70	80	70	70	70	68.8
		AV	63	65	75	65	68	58	70	80	67.8			AV	68	75	65	70	80	68	63	70	69.7
1414	1	1	60	50	40	80	70	50	60	80	61.3	1414	1	1	60	60	60	60	60	60	50	80	61.3
		2	60	70	50	90	50	50	60	80	63.8			2	70	70	70	60	80	60	80	70	70.0
		3	80	80	60	40	50	70	40	60	60.0			3	70	80	50	60	60	50	50	70	61.3
		4	70	60	60	80	80	50	70	80	68.8			4	70	70	70	70	70	70	60	90	71.3
		AV	68	65	53	73	63	55	58	75	63.4			AV	68	70	63	63	68	60	60	78	65.9
2000	1	1	50	50	50	20	50	40	60	50	46.3	2000	1	1	70	40	70	40	50	50	40	40	50.0
		2	40	50	50	50	50	50	70	70	53.8			2	40	50	50	70	50	60	40	60	52.5
		3	70	80	50	70	60	80	50	70	66.3			3	80	60	60	60	80	40	40	50	58.8
		4	70	80	70	50	60	60	80	50	65.0			4	50	60	60	60	70	60	50	60	58.8
		AV	58	65	55	48	55	58	65	60	57.8			AV	60	53	60	58	63	53	43	53	55.0
T3	1000	1	70	50	60	40	90	50	70	80	63.8	T6	1000	1	50	70	70	80	50	80	60	80	67.5
		2	70	80	60	80	80	70	70	90	75.0			2	60	60	80	30	**	70	60	80	67.5
		3	80	**	80	90	80	60	70	90	81.3			3	80	80	60	80	70	90	50	80	73.8
		4	90	90	80	70	90	90	70	90	83.8			4	90	50	60	70	80	90	70	70	72.5
		AV	78	80	70	70	85	68	70	88	75.9			AV	70	65	68	65	75	83	60	78	70.3
1414	1	1	40	50	70	60	50	70	60	60	57.5	1414	1	1	70	50	50	80	30	50	60	70	57.5
		2	50	90	90	40	70	60	60	50	63.8			2	60	80	50	60	70	40	60	60	60.0
		3	30	90	**	60	80	80	40	70	68.8			3	60	40	80	70	70	70	50	70	63.8
		4	60	80	70	70	70	80	70	90	73.8			4	80	80	60	70	80	70	70	70	72.5
		AV	45	78	83	58	68	73	58	68	65.9			AV	68	63	60	70	63	58	60	68	63.4
2000	1	1	60	50	60	60	40	10	80	60	52.5	2000	1	1	40	30	40	60	30	50	30	40	40.0
		2	30	60	50	70	70	50	60	60	56.3			2	70	60	60	60	60	60	50	40	57.5
		3	60	60	40	40	50	40	60	90	55.0			3	50	60	70	80	50	50	60	40	57.5
		4	90	70	70	60	80	80	60	70	72.5			4	50	80	70	50	80	60	50	70	63.8
		AV	60	60	55	58	60	45	65	70	59.1			AV	53	58	60	63	55	55	48	48	54.7

(TKR=Talker, SMR=Smearred bandwidth in Hz, REP=Replication, AV=Average, \*\*=1000)

**APPENDIX I**  
**RAW DATA OF SPECTRAL SMEARING RESULTS (I3-VOWEL SCORES)**

TKR	SMR	REP	LISTENER								AV	TKR	SMR	REP	LISTENER								AV
			L1	L2	L3	L4	L5	L6	L7	L8					L1	L2	L3	L4	L5	L6	L7	L8	
T1	1000	1	90	30	80	30	80	60	50	80	62.5	T4	1000	1	40	10	40	40	60	50	60	50	43.8
		2	70	60	50	60	60	20	50	80	56.3			2	40	70	50	30	80	40	50	50	51.3
		3	80	50	70	60	60	80	50	60	63.8			3	60	60	40	60	90	70	60	80	65.0
		4	80	80	70	90	80	80	60	90	78.8			4	50	30	50	50	90	60	70	90	61.3
		AV	80	55	68	60	70	60	53	78	65.3			AV	48	43	45	45	80	55	60	68	55.3
1414	1	1	70	30	20	60	60	20	70	60	48.8	1414	1	1	50	30	40	30	40	30	40	70	41.3
		2	60	30	40	60	70	10	60	50	47.5			2	50	20	10	40	80	20	30	40	36.3
		3	70	50	40	50	70	80	50	70	60.0			3	40	50	20	60	60	30	30	30	40.0
		4	70	90	50	60	90	50	80	90	72.5			4	70	30	40	80	80	30	50	60	55.0
		AV	68	50	38	58	73	40	65	68	57.2			AV	53	33	28	53	65	28	38	50	43.1
2000	1	1	70	30	50	20	30	10	30	30	33.8	2000	1	1	20	40	10	40	10	10	10	30	21.3
		2	30	20	50	50	50	30	50	40	40.0			2	40	30	30	40	20	20	20	50	31.3
		3	60	10	50	60	80	30	30	50	46.3			3	60	20	20	30	60	10	30	40	33.8
		4	40	60	20	50	70	50	30	70	48.8			4	60	30	20	40	50	30	30	40	37.5
		AV	50	30	43	45	58	30	35	48	42.2			AV	45	30	20	38	35	18	23	40	30.9
T2	1000	1	50	50	70	70	60	50	80	80	63.8	T5	1000	1	60	20	50	40	70	30	20	40	41.3
		2	80	50	60	70	80	40	70	**	68.8			2	70	60	60	60	80	30	40	60	57.5
		3	80	80	50	50	**	50	80	70	70.0			3	70	50	60	50	50	60	40	70	56.3
		4	60	80	90	70	90	80	80	**	81.3			4	30	60	80	60	80	60	60	70	62.5
		AV	68	65	68	65	83	55	78	88	70.9			AV	58	48	63	53	70	45	40	60	54.4
1414	1	1	60	30	20	70	50	30	60	60	47.5	1414	1	1	60	30	40	60	40	30	30	40	41.3
		2	40	50	40	80	50	30	30	70	48.8			2	60	30	30	60	80	20	50	30	45.0
		3	80	70	40	40	60	20	50	60	52.5			3	70	60	30	50	60	40	30	40	47.5
		4	50	50	60	60	80	20	60	70	56.3			4	50	20	20	40	70	60	30	60	43.8
		AV	58	50	40	63	60	25	50	65	51.3			AV	60	35	30	53	63	38	35	43	44.4
2000	1	1	30	20	50	20	60	10	40	40	33.8	2000	1	1	20	10	50	40	40	20	20	20	27.5
		2	30	30	20	30	60	30	40	20	32.5			2	50	20	40	20	50	10	20	40	31.3
		3	40	20	10	40	30	30	40	50	32.5			3	60	90	40	40	70	20	30	60	51.3
		4	80	50	40	50	50	50	70	60	56.3			4	70	60	50	60	80	20	60	70	58.8
		AV	45	30	30	35	50	30	48	43	38.8			AV	50	45	45	40	60	18	33	48	42.2
T3	1000	1	70	70	60	50	50	40	60	70	58.8	T6	1000	1	40	40	50	50	50	40	60	50	47.5
		2	70	70	80	90	**	70	60	70	76.3			2	80	50	40	50	70	40	60	50	55.0
		3	60	70	80	70	80	60	70	90	72.5			3	70	40	50	50	70	**	50	70	62.5
		4	70	60	70	70	70	80	70	70	70.0			4	80	60	70	80	70	80	40	70	68.8
		AV	68	68	73	70	75	63	65	75	69.4			AV	68	48	53	58	65	65	53	60	58.4
1414	1	1	40	30	50	30	70	20	50	40	41.3	1414	1	1	50	40	10	40	30	10	60	40	35.0
		2	50	50	50	50	80	60	60	70	58.8			2	60	60	50	30	90	30	20	30	46.3
		3	80	60	50	50	90	40	70	70	63.8			3	40	40	30	60	60	30	50	80	48.8
		4	80	90	60	70	80	60	70	60	71.3			4	70	50	60	60	70	40	60	60	58.8
		AV	63	58	53	50	80	45	63	60	58.8			AV	55	48	38	48	63	28	48	53	47.2
2000	1	1	30	30	20	20	40	20	20	70	31.3	2000	1	1	40	20	10	30	20	20	10	20	21.3
		2	40	40	40	70	60	20	40	30	42.5			2	30	20	30	30	50	30	30	30	31.3
		3	60	30	0	40	50	10	60	60	38.8			3	60	50	40	40	30	20	50	50	42.5
		4	50	40	40	50	50	60	60	50	50.0			4	40	50	30	30	40	30	20	60	37.5
		AV	45	35	25	45	50	28	45	53	40.6			AV	43	35	28	33	35	25	28	40	33.1

(TKR=Talker, SMR=Smeared bandwidth in Hz, REP=Replication, AV=Average; \*\*=100)

**APPENDIX I**  
**RAW DATA OF SPECTRAL SMEARING RESULTS (I4-FINAL CONSONANT SCORES)**

TKR	SMR	REP	LISTENER								AV	TKR	SMR	REP	LISTENER								AV	
			L1	L2	L3	L4	L5	L6	L7	L8					L1	L2	L3	L4	L5	L6	L7	L8		
T1	1000	1	60	70	70	50	90	60	50	80	66.3	T4	1000	1	70	50	80	40	60	60	60	80	62.5	
		2	70	70	80	70	60	80	70	90	73.8			2	80	80	80	50	60	40	80	60	66.3	
		3	70	80	80	70	80	80	70	90	77.5			3	80	80	**	**	80	60	60	80	80.0	
		4	70	90	60	90	70	90	70	**	80.0			4	70	60	90	70	70	70	70	70	70	71.3
		AV	68	78	73	70	75	78	65	90	74.4			AV	75	68	88	65	68	58	68	73	70.0	
1414	1	1	90	70	60	70	90	60	60	70	71.3	1414	1	1	70	50	50	80	70	60	60	80	65.0	
		2	60	50	50	90	90	30	80	70	65.0			2	50	60	50	70	90	40	60	70	61.3	
		3	60	80	80	80	90	70	70	70	75.0			3	60	90	30	90	90	50	50	60	65.0	
		4	80	90	80	70	90	70	70	**	81.3			4	80	60	70	90	90	50	50	80	71.3	
		AV	73	73	68	78	90	58	70	78	73.1			AV	65	65	50	83	85	50	55	73	65.6	
2000	1	1	50	10	50	40	50	30	60	40	41.3	2000	1	1	40	60	40	60	50	70	50	70	55.0	
		2	60	50	70	50	70	50	60	70	60.0			2	60	50	40	50	50	40	40	70	50.0	
		3	70	60	70	70	60	60	80	70	67.5			3	70	50	70	70	70	30	30	70	57.5	
		4	60	70	60	70	80	80	60	70	68.8			4	70	40	50	50	70	80	60	30	56.3	
		AV	60	48	63	58	65	55	65	63	59.4			AV	60	50	50	58	60	55	45	60	54.7	
T2	1000	1	80	60	60	70	70	90	80	70	72.5	T5	1000	1	60	60	50	60	70	70	50	60	60.0	
		2	90	60	80	70	80	60	60	90	73.8			2	50	40	70	60	70	60	60	80	61.3	
		3	60	70	70	70	70	60	60	80	67.5			3	80	70	70	60	70	60	50	70	66.3	
		4	40	90	80	60	90	90	70	90	76.3			4	70	50	60	80	*	80	70	**	76.3	
		AV	68	70	73	68	78	75	68	83	72.5			AV	65	55	63	65	78	68	58	78	65.9	
1414	1	1	60	40	50	50	50	30	50	60	48.8	1414	1	1	70	30	50	60	70	90	50	70	61.3	
		2	40	60	50	70	70	60	80	70	62.5			2	70	30	50	80	60	40	70	80	60.0	
		3	60	60	60	90	70	40	70	60	63.8			3	70	60	70	60	70	40	30	70	58.8	
		4	60	90	80	80	90	80	40	70	73.8			4	80	40	40	60	**	60	60	90	66.3	
		AV	55	63	60	73	70	53	60	65	62.2			AV	73	40	53	65	75	58	53	78	61.6	
2000	1	1	70	50	60	60	50	50	40	60	55.0	2000	1	1	70	30	40	60	40	60	40	60	50.0	
		2	60	80	50	60	60	60	40	80	61.3			2	60	30	30	70	50	50	40	50	47.5	
		3	70	50	60	60	80	30	70	80	62.5			3	80	70	70	60	70	50	40	60	62.5	
		4	90	50	70	50	80	50	70	70	66.3			4	50	50	40	80	80	70	60	70	62.5	
		AV	73	58	60	58	68	48	55	73	61.3			AV	65	45	45	68	60	58	45	60	55.6	
T3	1000	1	70	60	70	80	80	70	90	90	76.3	T6	1000	1	30	30	60	60	60	50	50	50	48.8	
		2	70	70	**	90	80	70	**	80	82.5			2	50	70	80	50	90	50	70	50	63.8	
		3	80	80	80	60	80	60	80	90	76.3			3	90	70	60	50	90	80	70	70	72.5	
		4	80	80	80	70	70	80	50	90	75.0			4	90	80	60	80	80	70	60	90	76.3	
		AV	75	73	83	75	78	70	80	88	77.5			AV	65	63	65	60	80	63	63	65	65.3	
1414	1	1	60	60	50	40	80	50	30	80	56.3	1414	1	1	30	50	60	40	50	20	30	70	43.8	
		2	80	60	50	40	70	70	50	90	63.8			2	50	50	50	60	50	70	30	60	52.5	
		3	80	80	80	90	70	50	70	90	76.3			3	60	50	50	80	50	60	40	70	57.5	
		4	60	70	60	70	90	70	60	80	70.0			4	40	60	80	70	70	90	60	90	70.0	
		AV	70	68	60	60	78	60	53	85	66.6			AV	45	53	60	63	55	60	40	73	55.9	
2000	1	1	40	60	60	70	50	50	40	70	55.0	2000	1	1	50	50	40	50	10	10	70	30	38.8	
		2	70	60	80	50	70	30	80	50	61.3			2	40	40	30	40	50	30	40	40	38.8	
		3	50	50	60	70	80	80	70	80	67.5			3	40	40	50	60	60	50	40	70	51.3	
		4	80	80	60	50	80	60	70	90	71.3			4	50	50	60	50	60	60	50	90	58.8	
		AV	60	63	65	60	70	55	65	73	63.8			AV	45	45	45	50	45	38	50	58	46.9	

APPENDIX J  
RAW DATA OF FORMANT RESULTS (J1)

Formant frequency F1, F2, F3, F4 for Talker 1 in 7 monophthongs  
of AB word list 1,2,3

	Phonetic symbols							Mean
	i	ɪ	ɛ	æ	ɑ	u	ʌ	
List 1	327	586	708	974	749	510	796	664
List 2	380	544	809	880	500	765	800	668
List 3	358	574	757	1207	777	409	791	696
F1 Mean	355	568	758	1020	675	561	796	676
List 1	2902	2112	2252	1796	1301	1100	1604	1867
List 2	2818	2050	2061	1852	1592	1445	1500	1903
List 3	3048	2071	2217	1859	1295	1635	1501	1947
F2 Mean	2923	2078	2177	1836	1396	1393	1535	1905
List 1	3386	2738	3163	2937	2911	3000	2922	3008
List 2	3279	3104	3089	2728	3000	2731	2000	2847
List 3	3480	2961	3111	3043	2810	3752	2991	3164
F3 Mean	3382	2934	3121	2903	2907	3161	2638	3006
List 1	4471	4119	4101	4073	3713	4000	4130	4087
List 2	4256	4130	4059	4000	4200	3569	4000	4031
List 3	4352	3973	4154	4279	3670	4566	3500	4071
F4 Mean	4360	4074	4105	4117	3861	4045	3877	4063

APPENDIX J  
RAW DATA OF FORMANT RESULTS (J2)

Formant frequency F1, F2, F3, F4 for Talker 2 in 7 monophthongs  
of AB word list1,2,3

	Phonetic symbols							Mean
	i	ɪ	ɛ	æ	ɑ	u	ʌ	
List 1	446	529	583	650	654	480	734	582
List 2	400	505	620	707	900	370	590	585
List 3	336	500	650	815	624	389	760	582
F1 Mean	394	511	618	724	726	413	695	583
List 1	2855	2273	2200	1150	1113	1300	1518	1773
List 2	2700	1950	2130	1993	1200	1368	1415	1822
List 3	2928	2057	2200	1890	1244	1507	1350	1882
F2 Mean	2828	2093	2177	1678	1186	1392	1428	1826
List 1	3271	3169	3012	1700	3196	2900	2958	2887
List 2	3300	3184	3233	3054	3000	2700	3400	3124
List 3	3600	3107	3200	3017	2991	2758	3190	3123
F3 Mean	3390	3153	3148	2590	3062	2786	3183	3045
List 1	4338	4421	4315	3200	4100	4000	4300	4096
List 2	4000	3800	4500	4300	3600	3900	4369	4067
List 3	4760	4350	4400	4230	3900	3800	4100	4220
F4 Mean	4366	4190	4405	3910	3867	3900	4256	4128

APPENDIX J  
RAW DATA OF FORMANT RESULTS (J3)

Formant frequency F1, F2, F3, F4 for Talker 3 in 7 monophthongs of AB word list 1,2,3

	Phonetic symbols							Mean
	i	I	U	æ	A	u	ʌ	
List 1	277	504	729	962	704	408	788	625
List 2	362	420	734	938	810	340	640	606
List 3	304	590	603	800	893	303	550	578
F1 Mean	314	505	689	900	802	350	659	603
List 1	2910	2405	2470	1575	1216	1547	1480	1943
List 2	2663	2219	1722	1722	1366	1627	1318	1805
List 3	3100	1866	1696	2266	1410	1700	1275	1902
F2 Mean	2891	2163	1963	1854	1331	1625	1358	1883
List 1	3845	3097	2822	2733	3261	3307	3116	3169
List 2	3300	2974	2681	2748	2600	3132	3369	2972
List 3	3600	3154	2500	3000	2861	2800	3447	3052
F3 Mean	3582	3075	2668	2827	2907	3080	3311	3064
List 1	4233	4222	4302	4292	4058	4047	4027	4169
List 2	4300	4250	4170	4455	4068	4362	4357	4280
List 3	4600	4332	3300	4302	4028	3800	4270	4090
F4 Mean	4378	4268	3924	4350	4051	4070	4218	4180

APPENDIX J  
RAW DATA OF FORMANT RESULTS (J4)

Formant frequency F1, F2, F3, F4 for Talker 4 in 7 monophthongs of AB word list1,2,3

	Phonetic symbols							Mean
	i	ɪ	ɛ	æ	ɑ	u	ʌ	
List 1	248	477	492	618	800	397	657	527
List 2	289	568	562	656	638	369	529	516
List 3	315	568	543	533	498	393	495	478
F1 Mean	284	538	532	602	645	386	560	507
List 1	2200	1937	1894	1690	1150	1210	1410	1642
List 2	2271	1681	1802	1769	1456	1537	1268	1683
List 3	2598	1472	1759	1792	1371	1755	1247	1713
F2 Mean	2356	1697	1818	1750	1326	1501	1308	1679
List 1	2800	2800	2500	2702	2800	2100	2750	2636
List 2	3326	2700	2600	2793	2696	2462	2796	2768
List 3	3000	2700	2800	2788	2600	2463	2804	2736
F3 Mean	3042	2733	2633	2761	2699	2342	2783	2713
List 1	3469	3622	3629	3176	3000	3106	3244	3321
List 2	3518	3300	3000	3085	3288	3205	3144	3220
List 3	3600	3200	3600	3350	3100	3133	3300	3326
F4 Mean	3529	3374	3410	3204	3129	3148	3229	3289

APPENDIX J  
RAW DATA OF FORMANT RESULTS (J5)

Formant frequency F1, F2, F3, F4 for Talker 5 in 7 monophthongs of AB word list1,2,3

	Phonetic symbols							Mean
	i	ɪ	ɛ	æ	ɑ	u	ʌ	
List 1	200	337	458	679	785	251	729	491
List 2	218	409	328	767	800	231	590	478
List 3	248	316	414	850	964	200	922	559
F1 Mean	222	354	400	765	850	227	747	509
List 1	2351	2036	2098	1657	1100	1303	1555	1729
List 2	2209	1858	1896	1721	1000	1200	1327	1602
List 3	2290	1858	1982	1703	1400	1600	1400	1748
F2 Mean	2283	1917	1992	1694	1167	1368	1427	1693
List 1	3000	3537	2600	2601	2800	2200	2250	2713
List 2	3379	2500	2652	2722	2500	2346	2830	2704
List 3	3000	2200	2574	2563	2311	2400	2779	2547
F3 Mean	3126	2746	2609	2629	2537	2315	2620	2654
List 1	3813	3956	3800	3833	4000	3500	3565	3781
List 2	3775	3800	3880	3700	3783	3521	3700	3737
List 3	3820	3563	3779	3628	3533	3500	3736	3651
F4 Mean	3803	3773	3820	3720	3772	3507	3667	3723

APPENDIX J  
RAW DATA OF FORMANT RESULTS (J6)

Formant frequency F1, F2, F3, F4 for Talker 6 in 7 monophthongs of AB word list1,2,3

	Phonetic symbols							Mean
	i	ɪ	ɛ	æ	ɑ	u	ʌ	
List 1	220	440	517	705	701	315	652	507
List 2	264	502	543	745	841	338	500	533
List 3	247	406	555	598	858	348	600	516
F1 Mean	244	449	538	683	800	334	584	519
List 1	2220	1831	1924	1806	1000	1327	1462	1653
List 2	2093	1635	1749	1788	1200	1804	1300	1653
List 3	2258	1828	1850	1805	1300	1805	1010	1694
F2 Mean	2190	1765	1841	1800	1167	1645	1257	1666
List 1	3168	2678	2672	2600	1935	2771	2136	2566
List 2	3229	2775	2626	2551	2000	2246	2473	2557
List 3	3284	2694	2635	2648	2182	2390	2742	2654
F3 Mean	3227	2716	2644	2600	2039	2469	2450	2592
List 1	3791	3265	3406	3515	3124	3369	3698	3453
List 2	3713	3695	3696	3762	3274	3251	3296	3527
List 3	3989	3609	4092	3613	3465	3246	3221	3605
F4 Mean	3831	3523	3731	3630	3288	3289	3405	3528

## REFERENCES:

- Baer, T and Moore, B.C.J. (1993). Effects of spectral smearing on the intelligibility of sentences in noise. J Acoust Soc Am, 94:1229-41.
- Bilger, R.C. and Wang, M.D. (1976). Consonant confusions in patients with sensorineural hearing loss. J Speech Hear Res, 19:718-48.
- Bonding, P. (1979). Frequency selectivity and speech discrimination in sensorineural hearing loss. Scan Audiol, 8:205-15.
- Boothroyd, A. (1968). Developments in speech audiometry. Sound, 2:3-10.
- Boothroyd, A. (1984). Auditory perception of speech contrasts by subjects with sensorineural hearing loss. J Speech Hear Res, 27:134-44.
- Boothroyd, A. and Medwetsky, L. (1992). Spectral distribution of /s/ and the frequency response of hearing aids. Ear & Hearing, 13:150-57.
- Boothroyd, A., Erickson, F.N., and Medwetsky, L. (1994). The hearing aid input: a phonemic approach to assessing the spectral distribution of speech. Ear & Hearing, 15:432-42.
- Boothroyd, A., Mulhearn, B., Ostroff, J., and Gong, J. (1996). Effects of spectral smearing on phoneme and word recognition. J Acoust Soc Am, 100:1807-18.
- Borden, G.J. and Harris, K.S. (1984). Speech Science Primer. Baltimore: Williams & Wilkins.
- Carney, A.E. and Nelson, D.A. (1983). An analysis of psychophysical tuning curves in normal and pathological ears. J Acoust Soc Am, 73:268-78.
- Celmer, R.D. and Bienvenue, G.R. (1987). Critical bands in the perception of speech by normal and sensorineural hearing loss listeners. In M.E.H. Schouten.(Ed.). The Psychophysics of Speech Perception. Dordrecht: Martinus Nijhoff Publishers, pp.473-80.
- Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M., and Gerstman, L.J. (1952). Some experiments on the perception of synthetic speech sounds. J Acoust Soc Am, 24:597-606.
- Delattre, P.C., Liberman, A.M., and Cooper, F.S. (1955). Acoustic loci and transitional cues for consonants. J Acoust Soc Am, 27:769-73.
- Dreschler, W.A. and Plomp, R. (1980). Relations between psychophysical data and speech perception for hearing-impaired-subjects, I. J Acoust Soc Am, 68:1608-15.
- Dreschler, W.A. and Plomp, R. (1985). Relations between psychophysical data and speech perception for hearing-impaired-subjects, II. J Acoust Soc Am.

78:1261-70.

- Dubno, J.R., Dirks, D.D., and Ellison, D.E. (1989). Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. I: The contribution of selected frequency regions. J Acoust Soc Am. 85:347-54.
- Dubno, J.R., Dirks, D.D., and Schaefer, A.B. (1987). Effects of hearing loss on utilization of short-duration spectral cues in stop consonant recognition. J Acoust Soc Am. 81:1940-47.
- Evans, E.F. (1978). Peripheral auditory processing in normal and abnormal ears: Physiological considerations for attempts to compensate for auditory deficits by acoustic and electrical prostheses. Scand Audiol (Suppl). 6:9-47.
- Festen, J.M. and Plomp, R. (1983). Relations between auditory functions in impaired hearing. J Acoust Soc Am. 76:652-62.
- Fitzgibbons, P.J. and Wightman, F. L. (1982). Gap detection in normal and hearing-impaired listeners. J Acoust Soc Am. 72:761-65.
- Fletcher, H. (1953). Speech and Hearing in Communication. New York: D. van Nostrand Co.
- Florentine, M. and Buus, S. (1984). Temporal gap detection in sensorineural and simulated hearing-impairments. J Speech Hearing Res. 27:449-55.
- Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (1980). Frequency selectivity in normal-hearing and hearing-impaired listeners. J Speech Hearing Res. 23:643-69.
- Fox, R.A. (1989). Dynamic information in the identification and discrimination of vowels. Phonetica. 46:97-116.
- French, N.R. and Steinberg, J.C. (1947). Factors governing the intelligibility of speech sounds. J Acoust Soc Am. 19:90-119.
- Fry, D.B. (1978). The role and primacy of the auditory channel in speech and language development. In M. Ross, and T.G. Giolas, (Eds.). Auditory Management of Hearing Impaired Children. Baltimore: University Park Press, pp. 15-43.
- Gagné, J.P. and Erber, N.P. (1987). Simulation of sensorineural hearing impairment. Ear and Hearing. 8:232-43.
- Glasberg, B.R. and Moore, B.C.J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J Acoust Soc Am. 79:1020-33.
- Hack, Z.C. and Erber, N.P. (1982). Auditory, visual, and auditory-visual perception of vowels by hearing-impaired children. J Speech Hearing Res.

25:100-107.

- Howard-Jones, P.A. and Summers, I.R. (1992). Temporal features in spectrally degraded speech. Acoustic Letters, 15:159-63.
- Kewley-Port, D. (1983). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. J Acoust Soc Am, 72:379-89.
- Klatt, D.H. and Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. J Acoust Soc Am, 87:820-57.
- Levitt, H. and Webster, J.C. (1991). Effects of noise and reverberation on speech. In C.Harris.(Ed.), Handbook of Acoustic Measurement and Noise Control (3<sup>rd</sup> Ed.). New York: McGraw Hill, pp. 16.1-16.20.
- Liberman, A.M., Delattre, P.C., Gerstman, L.J., and Cooper, F.S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. J Exp Psychol, 52:127-37.
- Miller, G.A. and Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. J Acoust Soc Am, 27:338-52.
- Monsen, R.B. and Engebretson, A.M. (1977). Study of variations in the male and female glottal wave. J Acoust Soc Am, 62:981-993.
- Moore, B.C.J. and Glasberg, B.R. (1987). Relationship between psychophysical abilities and speech perception for subjects with unilateral and bilateral cochlear hearing impairments. In M.E.H.Schouten (Ed.), The Psychophysics of Speech Perception. Dordrecht: Martinus Nijhoff Publishers, pp.449-60.
- Nearey, T.M. (1989). Static, dynamic, and relational properties in vowel perception. J Acoust Soc Am, 85:2088-113.
- Patterson, R.D., Nimmo-Smith, I., Weber, D.L., and Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram and speech threshold. J Acoust Soc Am, 72:1788-1803.
- Peterson, G.E. and Barney, H.L. (1952). Control methods used in a study of the vowels. J Acoust Soc Am, 24:175-84.
- Pick, G.F., Evans, E.F., and Wilson, J.P. (1977). Frequency resolution in patients with hearing loss of cochlear origin. In: E.F. Evans and J.P. Wilson (Eds.), Psychophysics and Physiology of Hearing. New York: Academic Press, pp.273-81.
- Plomp, R. (1964). The ear as a frequency analyzer. J Acoust Soc Am, 36:1628-36.
- Rakerd, B. and Verbrugge, R.R. (1985). Linguistic and acoustic correlates of the

- perceptual structure found in an individual differences scaling study of vowels. J Acoust Soc Am. 77:296-301.
- Raphael, L. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. J Acoust Soc Am. 51:1296-1303.
- Raphael, L. (1981). Durations and contexts as cues to word-final cognate opposition in English. Phonetica. 38:126-47.
- Rosen, S. and Fourcin, A. (1986). Frequency selectivity and the perception of speech. In: B.C.J. Moore (Ed.), Frequency Selectivity in Hearing. London: Academic Press. pp.373-487.
- Ross, M. (1982). Hard of Hearing Children in Regular Schools. Englewood Cliffs, NJ: Prentice-Hall.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science. 270:303-04.
- Simon, G.R. (1963). The critical bandwidth in recruiting ears and its relation to temporal summation. J Aud Res. 3:109-19.
- Stelmachowicz, P.G., Jesteadt, W., Gorga, M.P., and Mott, J. (1985). Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. J Acoust Soc Am. 77:620-27.
- Strange, W., Verbrugge, R., Shankweiler, D., and Edman, T. (1976). Consonantal environment specifies vowel identity. J Acoust Soc Am. 60:213-24.
- Strange, W. (1989). Evolving theories of vowel perception. J Acoust Soc Am. 85:2081-87.
- Strange, W. (1999). Perception of vowels: Dynamic constancy. In J.M. Pickett, The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology. Needham Heights, MA: Allyn & Bacon, pp. 153-165.
- Studebaker, G.A. and Hochberg, I. (1993). Acoustical Factors Affecting Hearing Aid Performance. Boston, MA: Allyn and Bacon.
- Studebaker, G.A. and Bess, F.H. (Eds.) (1982). The Vanderbilt Hearing Aid Report. Upper Darby, PA: Monographs in Contemporary Audiology.
- Summers, I.R., and Al-Dabbagh, A.D. (1985). Electronically simulated loss of frequency resolution. Br J Audiol. 19:49-52.
- Summers, I.R., Al-Dabbagh, A.D., and Garnham, C.W. (1986). The electronic simulation of sensorineural hearing loss. Proc. IEEE/8<sup>th</sup> Annual Conference of the Engineering in Medicine and Biology Society. 1868-71.

- Summers, I.R. (1991). Electronically simulated hearing loss and the perception of degraded speech. In: D.L. Wise(Ed.), Bioinstrumentation and Biosensors. New York: Marcel Dekker. pp. 589-610.
- Ter Keurs, M., Festen, J.M., and Plomp, R. (1992). Effects of spectral smearing on speech perception. J Acoust Soc Am, 91:2872-80.
- Turner, C.W. and Henn, C.C. (1989). The relation between vowel recognition and measures of frequency resolution. J Speech Hear Res, 32:49-58.
- Tyler, R.S., Summerfield, Q., Wood, E.J., and Fernandes, M.A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. J Acoust Soc Am, 72:740-52.
- Tyler, R.S. (1986). Frequency resolution in hearing-impaired listeners. In: B.C.J. Moore(Ed.), Frequency Selectivity in Hearing. London: Academic Press. pp. 309-71.
- Vernon, M. and Andrews, J.F. (1990). The Psychology of Deafness: Understanding deaf and hard-of-hearing people. White Plains, NY: Longman.
- Villchur, E. (1974). Simulation of the effect of recruitment on loudness relationships in speech. J Acoust Soc Am, 56:1601-11.
- Villchur, E. (1977). Electronic models to simulate the effect of sensory distortions on speech perception by the deaf. J Acoust Soc Am, 62:665-74.
- Zwicker, E. and Schorn, K. (1978). Psychoacoustical tuning curves in audiology. Audiology, 17:120-40.