INFORMATION TO USERS

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

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

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

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

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

UMI

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

CONSONANT RECOGNITION IN ON-HARMONIC VERSUS BETWEEN-HARMONIC NOISE

by

JANET REATH SCHOEPFLIN

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

1997

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 Philosphy.

9-15-97
Date

Chair of Examining Committee

9-18-97
Date

Executive Officer

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

CONSONANT RECOGNITION IN ON-HARMONIC
VERSUS BETWEEN-HARMONIC NOISE

by

Janet Reath Schoepflin

Adviser: Professor Harry Levitt

The consonant recognition performance of ten normally hearing adults was measured for twenty synthesized vowel-consonant nonsense syllables in speech-spectrum shaped noise, comb-filtered on-harmonic noise, and comb-filtered between-harmonic noise at four signal-to-noise ratios under full-band, low-pass, and high-pass filtering conditions to test the hypothesis that only those portions of the noise spectrum that lie in the same critical bands as the speech spectrum are effective in masking the speech.

Since the width of the critical band increases logarithmically with frequency, it was anticipated that speech recognition scores for the low-pass filter condition would be higher for between-harmonic noise than for on-harmonic noise because the speech harmonics and between-harmonic noise occupy different critical bands, but co-exist in the same critical bands for the on-harmonic noise. For the high-pass filter condition, no difference in masking effect between the on-harmonic and between-harmonic noise was

expected since both of the noises occupy the same critical bands as the speech. For the full-band condition, speech recognition scores were again expected to be higher for the between-harmonic noise than for the on-harmonic noise because the between-harmonic noise only occupies the same critical bands as the speech in the high-frequency region of the spectrum.

Results supported the hypothesis in that speech recognition scores for the between-harmonic noise were significantly higher than those for the on-harmonic noise at the poorer signal-to-noise ratios in the full-band and low-pass filter conditions, and that there were no significant differences among the noise types at any of the signal-to-noise ratios tested in the high-pass filter condition. The implications of the findings with regard to current and future noise reduction strategies are discussed.

ACKNOWLEDGEMENTS

I am indebted to many individuals for their guidance and encouragement in the completion of this paper. I gratefully acknowledge the support of my colleagues and friends at Temple University who began the journey with me, and those at Hunter College of The City University of New York who helped me conclude it. I also acknowledge with appreciation the students at those two institutions who always inspired and uplifted me with their enthusiasm and good cheer. I owe a particular debt of gratitude to those students who served as subjects at various stages of the project.

I am also grateful to the entire faculty of the Ph.D. Program in Speech and Hearing Sciences at The Graduate School of The City University of New York for their helpful suggestions as the study was being developed. My deepest appreciation goes to the members of my supervisory committee: Professor Irving Hochberg, whose counsel throughout the project was so beneficial and whose editorial skill was invaluable; Professor Mark Weiss, whose patience and gentility in the face of my unending questions were exceeded only by his vast knowledge of acoustics and his wizardry at the computer; and my mentor and chair, Distinguished Professor Harry Levitt, whose brilliance as a scholar and researcher was the ideal that always motivated me and whose ever-positive attitude continued to encourage me. I learned a great deal more than

scientific inquiry from each of them.

Finally, I acknowledge the love and support given to me by my mother, Margaret, my late father, Raymond, and especially my beloved husband, Robert, and our three children, Adam, Zachary, and Daniel. Each of them played a significant role in the completion of this project, and each deserves to be commended. Ice cream for everyone!

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background noise adversely affects speech recognition, particularly for those with sensorineural hearing loss (Chung and Mack, 1979; Cohen and Keith, 1976; Cooper and Cutts, 1971; Dirks, Morgan, and Dubno, 1982; Keith and Talis, 1972; Leshowitz, 1977; Martin and Pickett, 1970; Miller, 1947; Niemeyer, 1965; Plomp and Mimpen, 1979; Ross et al., 1965; Suter, 1985). Since most daily conversations occur within noisy settings (Pearsons, Bennett, and Fidell, 1976; Plomp, 1977), auditory research efforts have been directed toward developing methods of enhancing the speech signal or to eliminating or reducing the background noise in order to improve communication.

Both single-microphone and multi-microphone noise reduction methods have been developed, but the primary emphasis has been on single-microphone techniques, due to practical considerations such as wearability and increased cost associated with multi-microphone techniques. While single-microphone methods have achieved reductions in background noise levels, sometimes substantially, speech recognition performance has generally not improved significantly (Lim, 1982; Lim and Oppenheim, 1979; Schafer, 1982).

According to the articulation index (French and Steinberg, 1947), speech recognition performance in noise can be predicted by determining the proportion of

audible speech within 20 contiguous frequency bands, each of which makes an equal and independent contribution to recognition. These bands correspond closely to the filter network of the basilar membrane referred to as the critical bands.

The critical bands account for a variety of psychoacoustic phenomena, including masking. Masking is defined by the American National Standards Institute (1989) as "The process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound" (p. 3). Masking of speech cues can occur directly, when the speech and noise occupy identical frequency regions, or can occur indirectly, when frequency components of the speech and noise occupy separate, but adjacent, regions within the same critical band. Summation of the speech and noise within each critical band will result in the noise masking the speech signal, either directly or indirectly. As a consequence, speech recognition performance will be decreased.

Because the harmonic structure of the voice is important to the identification of both the prosodic and phonemic features of speech, direct or indirect masking of the vocal harmonics is likely to reduce speech recognition ability. If a particular noise reduction strategy does not remove the masking effect of the noise from the intelligibility-bearing harmonics, speech recognition performance is not expected to improve, even if the overall noise level is decreased.

The purpose of this study is to examine the extent to which direct and indirect masking of the harmonics of the speech signal reduces speech recognition ability and the effect of the critical band on this masking. It is hypothesized that since the vocal

harmonics are crucial information-bearing components of voiced speech, noise that occupies the identical frequency regions as the harmonics of the speech signal (on-harmonic noise) will be more damaging to speech recognition performance than noise that occupies the frequency regions between the harmonics (between-harmonic noise). It is further hypothesized that when the between-harmonic noise is within the same critical band as an harmonic, the masking effect of the noise will be the same as on-harmonic noise since all the energy within the critical band is summed; alternatively, between-harmonic noise that occupies a separate critical band than the speech harmonic(s) will have no effect on speech recognition performance. Results of the study may contribute to an understanding of why single-microphone noise reduction methods have not been successful in improving speech recognition performance, and may have implications for the design of future noise reduction and speech-processing devices, such as hearing aids and other auditory sensory aids.

CHAPTER 2

REVIEW OF THE LITERATURE

Speech intelligibility depends on both acoustic and non-acoustic factors. The acoustic factors include the spectral and temporal characteristics of the speech signal, and the spectral and temporal characteristics of the competing signal, including reverberation. The non-acoustic factors include such aspects as message predictability, listener familiarity with the speaker and/or the material, listener vigilance or attention, and the availability of visual cues. The spectro-temporal factors relating to the speech signal and the competing signal are of primary interest in this discussion.

Speech Acoustics

The long-term average speech spectrum contains energy over a wide frequency range. The spectral energy important for intelligibility lies within the range between 100 Hz and 10,000 Hz (French and Steinberg, 1947). Most of the energy is below 1000 Hz, with the peak spectral density in the region of 500 Hz. Intelligibility of speech, however, is contained within the higher frequency components.

Figure 1 (adapted from Levitt, 1986) shows speech spectrum curves for 1/3 octave band pressure levels for a normally speaking average male and average female talker at a distance of 1 meter from the talker's lips. Within each band, the peaks of the

speech signal are roughly 12 dB above the root-mean-square (rms) level. Also shown are typical frequency ranges containing important cues for intelligibility for the various sounds of speech. Although these frequency and intensity ranges represent typical values, it must be remembered that there may be significant differences in spectral shape and level depending upon vocal effort, dialect, context, individual talker variation, and distance from and orientation to the speaker.

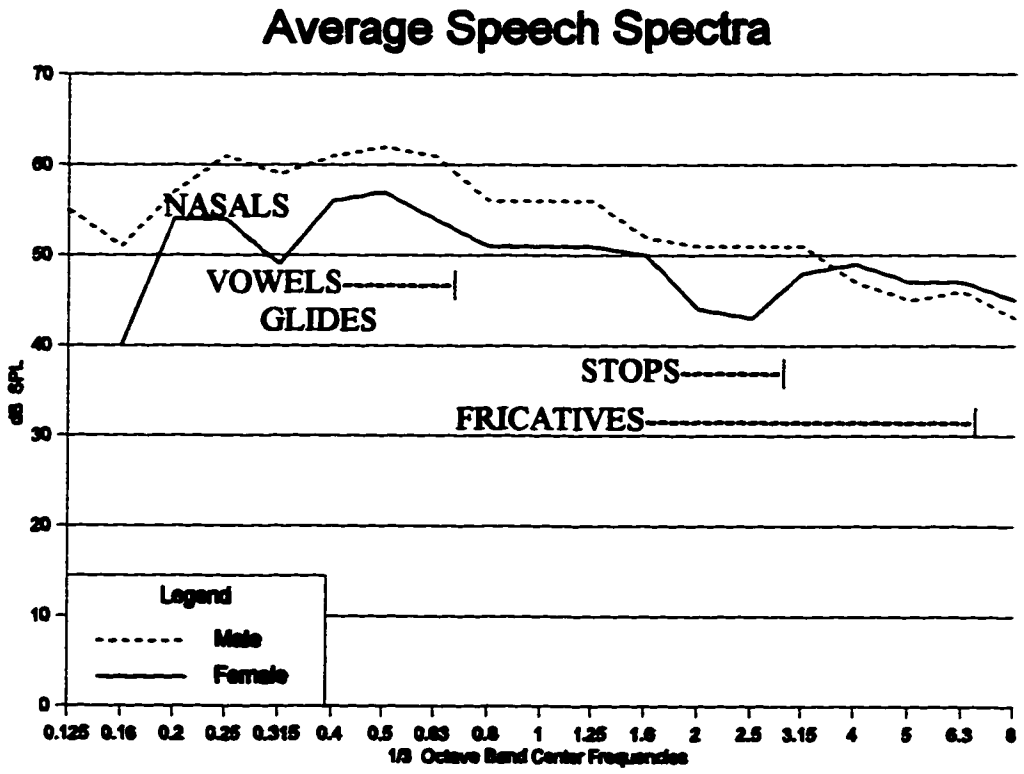


Figure 1. Average speech spectra for male and female talkers (adapted from Levitt, 1986).

The acoustic characteristics of speech sounds are a consequence of the way they are produced. Three sound sources can be identified within the vocal tract. One source is the glottal pulse train, created as the vocal folds repeatedly open and close. The other sources are turbulence of the airstream causing audible frication, and the sudden release of air pressure causing a transient plosive sound.

The repetition rate of the glottal pulses determines the fundamental frequency and harmonic structure of the voice. When the glottal source passes through the vocal tract, its spectrum (fundamental frequency and harmonics) is modified by the resonances of the vocal tract as shaped by the articulators. Vowels are produced as the glottal source passes through a relatively open vocal tract, defined primarily by the place and degree of tongue constriction and the amount of lip rounding. On average, vowels are of greater power, longer duration, and lower frequency than consonants. This is true because consonants are generally produced with more constriction in the vocal tract than are vowels and because the sound source for consonants is not always at the glottis.

A common means of classifying consonants is to describe them according to three features: voicing, manner of articulation, and place of articulation. The voicing feature is expressed as voiced or voiceless, referring to the presence or absence of vocal fold vibration during production of the consonant. The manner feature describes the way in which the consonant is produced, and is often characterized using the categories glide (to include the semivowels and liquids), nasal, stop, fricative, and affricate. As its name indicates, the place feature describes the location in the vocal tract at which the major

articulatory constriction occurs and specifies the articulators involved in the production of the sound. Table 1 summarizes the voicing, manner, and place classification of the 25 commonly recognized consonants of English.

Table 1. -- Classification of English Consonants

PLACE	MANNER (Voiceless/Voiced)				
	Glide	Nasal	Stop	Fricative	Affricate
Bilabial	w, ʍ	m	p/b		
Labiodental				f/v	
Linguadental				θ/ð	
Alveolar	j, l	n	t/d	s/z	ʧ/ʤ
Palatal	r				ʃ/ʒ
Velar/Pharyngeal		ŋ	k/g	h	

Although produced with more constriction, the glides are fairly similar in production to the vowels and, as can be seen in figure 1, their acoustic characteristics are correspondingly similar.

Nasal consonants are produced by the occlusion of the oral tract by the lips or

tongue and by the coupling of the oral and nasal cavities by the lowering of the velum. The glottal pulses are passed through the relatively large nasal cavity and out through the nostrils, resulting in a strong low frequency energy, referred to as nasal murmur, as well as antiresonances in the low frequency region.

The stop consonants are produced by complete closure of the vocal tract followed by a rapid release from closure, resulting in a period of silence (or very low level sound) during the closure, followed by a brief, relatively low intensity, high frequency noise burst and rapidly varying resonances known as formant transitions. The voiced stop consonants combine the transient source with the glottal source.

The fricatives are produced as the breath stream is forced through a narrow constriction in the vocal tract, resulting in turbulence. Sounds produced in this way have a broad frequency spectrum. Affricates are produced as combined stop-fricatives, with the acoustic characteristics reflecting both the stop and fricative components. As with the voiced stops, voiced fricatives and affricates are produced with two sound sources.

In addition to contributing to the identification of the segmentals, the glottal sound source provides important information concerning prosodics, including linguistically significant stress patterns (Fry, 1958; Ladefoged, 1963) and intonation contours (Lieberman, 1967), and the mood (Uldall, 1960) or emotional intent of the speaker (Lieberman and Michaels, 1962).

As mentioned previously, the relative intensities of the sounds of speech vary over a wide range. As shown in figure 1, for speech produced at a fixed vocal effort, the

range between the most intense speech sounds (the vowels) and the least intense speech sounds (the weak voiceless fricatives) is approximately 30 dB. For speech that varies significantly in vocal effort, as from a heavily emphasized syllable in a loud voice to an unstressed syllable produced at a low vocal effort, the dynamic range of the speech signal can exceed 60 dB.

The spectrum and level of the speech signal at the listener's ear can be varied over a wide range before there is a breakdown in intelligibility. The classic research done by French and Steinberg (1947) demonstrated that either a high-pass or low-pass filter with a cut-off frequency of about 1800 Hz produced an intelligibility score of approximately 70% for nonsense syllables, yet almost perfect recognition for sentences; similarly, the work of Hirsh, Reynolds and Joseph (1954) demonstrated that word intelligibility in quiet was relatively unaffected when frequencies below about 1700 Hz were retained. Also, the findings of Miller, Heise and Lichten (1951) showed that word recognition scores were reduced to about 50% at a signal-to-noise ratio of 0 dB, while intelligibility of more conversational materials at the same signal-to-noise ratio approached 100%.

Although audibility of the entire speech spectrum is not required for good speech recognition performance, the proportion of the speech spectrum available to the listener is a fairly good predictor of relative intelligibility. One predictive method of speech recognition performance that has achieved widespread use is the Articulation Index (French & Steinberg, 1947).

Basically, the Articulation Index (AI) divides the speech spectrum into 20 contiguous frequency bands, each of which is assumed to make an equal and independent contribution to intelligibility. These are shown in table 2. The speech peaks are roughly 12 dB above the rms speech levels in each band. The speech peak-to-rms-noise ratio (or speech peak-to-threshold ratio, whichever is lower) in each band determines the contribution of that band to speech recognition performance. The maximum contribution for each band is 30 dB. The minimum contribution for a band is 0 dB. These contributions are summed together and then normalized to give a single number between 0.0 and 1.0 representing the proportion of speech audible to the listener. The calculated AI is then converted into a speech recognition score.

Table 2. --Bands of Equal Importance to Speech Recognition
(from French & Steinberg, 1947)

Band	Bandwidth	Lower cutoff (Hz)	Upper cutoff (Hz)
1	125	250	375
2	130	375	505
3	140	505	645
4	150	645	795
5	160	795	955
6	175	955	1130
7	185	1130	1315
8	200	1315	1515
9	205	1515	1720
10	210	1720	1930
11	210	1930	2140
12	215	2140	2355
13	245	2355	2600
14	300	2600	2900
15	355	2900	3255
16	425	3255	3680
17	520	3680	4200
18	660	4200	4860
19	860	4860	5720
20	1280	5720	7000

The AI, as originally proposed by French and Steinberg, assumed an equal weight for each band when summing bands. Subsequent researchers have shown that, depending on the speech materials used, different weights need to be assigned to each band (Pavlovic, 1987; Studebaker, Pavlovic, and Sherbecoe, 1987; Studebaker and Sherbecoe, 1993). These weights are referred to as the importance function. The traditional AI was based on a nonsense syllable importance function. The importance functions for different types of speech material have been obtained empirically by various researchers (Duggirala et al., 1988; Pavlovic, 1987; Pavlovic, 1989; Studebaker, Pavlovic, and Sherbecoe, 1987; Studebaker and Sherbecoe, 1989; Studebaker and Sherbecoe, 1991).

Speech recognition ability is also influenced by the non-acoustic factors mentioned above regarding the predictability of the message, the familiarity of the listener with the speaker and/or the material, the attention and motivation of the listener, and the availability of visual cues. Through the development of importance functions for different types of speech materials, particularly that for "average speech" (Pavlovic, 1987), many of these non-visual factors have been incorporated into the AI, and its predictive power has been improved. Relatively good predictions of speech recognition performance for both normal-hearing and hearing-impaired subjects have been achieved using the AI with the speech level distributions reflected in figure 1 (Dubno, Dirks, and Schaefer, 1989; Humes et al., 1986; Kamm, Dirks, and Bell, 1985; Pavlovic, Studebaker, and Sherbecoe, 1986; Skinner and Miller, 1983).

Competition/Noise

Speech intelligibility is adversely affected by noise because the information-bearing speech cues are masked. The reduction in intelligibility depends upon the spectral and temporal characteristics of the speech and noise, as well as the non-acoustic factors noted previously. Masking is most effective when the frequency components of the masker and the signal are similar (Wegel and Lane, 1924) and when the masker and signal co-exist in time (Elliott, 1962).

The classic experiment by Fletcher (1940), relating the absolute threshold of a sinusoid to the bandwidth of a narrow band noise masker, revealed that increases in the bandwidth of the noise (with a constant spectrum level) increased the pure tone threshold up to a particular bandwidth, beyond which increasing the bandwidth had essentially no effect on the signal threshold. Fletcher called these passbands the "critical bandwidths" and likened them to rectangular filters corresponding to distances on the basilar membrane.

Since Fletcher's time, numerous researchers have examined the critical band concept with regard to various auditory phenomena, including threshold sensitivity (Gässler, 1954; Morton and Carpenter, 1963; Zwicker, 1954), loudness perception (Scharf, 1961; Scharf, 1970; Zwicker, Flottorp, and Stevens, 1957), and frequency resolution (Plomp, 1964), all with similar findings concerning the width and shape of the auditory filter. Gässler's (1954) study devoted to threshold sensitivity study and Morton and Carpenter's (1963) detection study are of particular interest here since the findings

suggest that all the energy within a critical band is summed.

A currently popular model views the critical bandwidths as a series of overlapping bandpass filters, the widths of which broaden as the center frequency increases. At moderate intensity levels, the filters are roughly symmetric, becoming less sharply tuned on the low frequency side as intensity increases (Moore and Glasberg, 1987), thereby making possible an "upward spread" of masking. The equivalent rectangular bandwidth (ERB) of the filters is believed to reflect a uniform distance along the basilar membrane.

Summarizing research findings, Scharf (1970) has published the estimated bandwidths for 24 representative contiguous auditory filters. These are seen in table 3. It is interesting to note the striking similarity between Scharf's middle 20 bands (from band 3 to band 22), and those 20 bandwidths shown in table 2, described by French and Steinberg (1947) as being equally important to speech recognition.

Table 3. --Critical Bandwidths (from Scharf, 1970)

Band #	Center Frequency (Hz)	Critical Band (Hz)	Lower cutoff (Hz)	Upper cutoff (Hz)
1	50			100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1000	160	920	1080
10	1170	190	1080	1270
11	1370	210	1270	1480
12	1600	240	1480	1720
13	1850	280	1720	2000
14	2150	320	2000	2320
15	2500	380	2320	2700
16	2900	450	2700	3150
17	3400	550	3150	3700
18	4000	700	3700	4400
19	4800	900	4400	5300
20	5800	1100	5300	6400
21	7000	1300	6400	7700
22	8500	1800	7700	9500
23	10,500	2500	9500	12,000
24	13,500	3500	12,000	15,500

Each critical band serves to filter those noise components outside its passband. Only those noise components within the critical band are effective in masking the signal, and the signal threshold is determined by the signal-to-noise ratio for that filter. As table 3 shows, the critical bandwidth at a center frequency of 7000 Hz is roughly ten times that of the bandwidth at 700 Hz (i.e., the bandwidth increases logarithmically with center frequency, except for frequencies below about 500 Hz). This fact, in conjunction with the increasing asymmetry of the filters at increasing intensity, gives rise to the finding that the low intensity, high frequency consonants, are, in general, more easily masked than the more intense, low and middle frequency vowels and voiced continuants.

Noise Reduction

Both analog and digital technologies have been employed to reduce noise, thereby improving the signal-to-noise ratio, with an ultimate goal of improving speech recognition performance in noise. Noise reduction strategies involve separating the signal and noise sources in terms of their spatial, spectral, or temporal characteristics.

Noise reduction in the spatial domain is accomplished through the use of directional microphones provided the speech and noise arrive from different directions. In the absence of high reverberation (which effectively blurs directional cues), binaural amplification with directional microphones provides an improvement in signal-to-noise ratio with corresponding improvements in intelligibility over monaural listening with omnidirectional microphones (Hawkins and Yacullo, 1984). Similar results have also been obtained using advanced signal processing techniques, such as adaptive beam

forming (Peterson et al., 1987) and adaptive noise cancellation (see below).

Noise reduction in the time domain is accomplished by reducing those intense signals that are of brief duration and unlikely to be components of the speech signal (e.g., a door slam), or by constantly comparing the incoming signal to the average speech signal; if the low frequency energy is greater than what is expected, it is assumed that the energy is noise, and the filter automatically adjusts to minimize the interference. The methods used include peak clipping, multiband amplitude compression, and time-domain adaptive filtering.

Peak clipping reduces noise by eliminating strong peaks in the noise signal. This technique is also a relatively simple method of limiting the output of a hearing aid. It is, however, of little or no value in improving speech intelligibility.

While not designed specifically for noise-reduction, multichannel amplitude compression (and related forms of amplitude compression) provides differential amplification characteristics depending upon the level and spectral characteristics of the signal being amplified, theoretically resulting in an improved signal-to-noise ratio if the background noise is primarily low-frequency, and the gain in the low-frequency channel is reduced while the high-frequency gain is unchanged. Although Moore (1987) reported improved speech intelligibility in noise with a two-channel compression system, improvements in speech recognition performance have not generally been noted (Abramovitz, 1980; Lippman, Braida, and Durlach, 1981; Nabelek, 1983; Walker, Byrne, and Dillon, 1984), and it has been suggested that the method may actually

introduce distortions that can reduce the signal-to-noise ratio and decrease intelligibility (Haggard et al., 1987; Plomp, 1988).

Time-domain adaptive filtering provides for an improved signal-to-noise ratio through the use of an algorithm which "subtracts" a reference noise signal from the primary speech-plus-noise signal. These techniques often employ a Least Mean Square (LMS) adaptive filter (Widrow, 1966) which continuously adjusts itself to achieve the minimum error, i.e., the optimal noise reduction. Spatial separation between the microphones enhances the noise-cancellation effect. Experiments employing time-domain adaptive filtering methods have shown improved signal-to-noise ratios with modest improvements in intelligibility in controlled test environments. (Brey et al., 1984; 1987; Chabries et al., 1982; Harris et al., 1988; Schwander and Levitt, 1987). Single-channel time-domain adaptive filters have also been described (Graupe and Causey, 1977; Sambur, 1978); while improvements in signal-to-noise ratio were reported and initial findings suggested improved intelligibility scores (Graupe, Grosspietsch, and Basseas, 1987; Stein and Dempsey-Hart, 1984; Wolinsky, 1986), later studies found no significant improvements in speech recognition performance with this method (Schum, 1990; Tyler and Kuk, 1989; Van Tasell, Larsen, and Fabry, 1988). The filter used by Graupe and Causey (1977) actually combined both time-domain and frequency-domain filtering (see below).

One comb-filtering technique (Frazier et al., 1976), exploiting the periodicity of the speech time waveform against random wide-band noise, was able to reduce the noise,

but resulted in decreased intelligibility at several signal-to-noise ratios (Lim, Oppenheim, and Braida, 1978; Lim and Oppenheim, 1979). Presumably, the noise that remained was concentrated around the harmonics, the center frequencies of the passbands of the filters.

A related single-microphone technique that has been developed is sinusoidal modeling (McAulay and Quatieri, 1986) in which the spectral peaks of speech-in-noise are replaced with sinusoids which match the frequency, amplitude, and phase of the peaks. Reproducing only the most intense peaks produces an increased signal-to-noise ratio because the lower intensity portions of the spectrum, assumed to be noise, are suppressed while the more intense portions, assumed to be speech, are reproduced. The study by Kates (1994), however, demonstrated that while decreasing the number of sinusoids used to reproduce the speech improved the signal-to-noise ratio, speech recognition performance and perceived intelligibility were reduced for normal listeners in both quiet and noise.

Noise reduction in the frequency domain is achieved through editing the frequency spectrum of the speech-plus-noise. Many of these methods involve making an estimate of the spectrum of the noise during a pause in the speech, and then using this information to reduce the noise in frequency regions where components of the speech signal are not present. This method can be effective if the noise occupies a narrow frequency range and can be filtered out. It is considerably less effective in those situations where the noise spectrum is similar to that of the speech, since the attenuation will apply to both spectra. Although listeners in some studies using these approaches

show small increases in speech recognition ability, many listeners show no improvement or even reduced intelligibility (Lim, 1983; Sigelman and Preves, 1987; Stach, Speerschneider, and Jerger, 1987). Frequency-domain adaptive filters have also been used experimentally (Christiansen, Chabries, and Lynn, 1982; Dentino, McCool, and Widrow, 1978; Ferrara, 1980; 1985) to reduce the computational time associated with time-domain adaptive filters. The end result of the two methods is not significantly different.

While most of the described single-microphone noise reduction methods improve the signal-to-noise ratio, experimental evaluations have not shown corresponding improvements in speech recognition performance. Whereas these strategies can and do reduce overall noise levels, thereby improving the speech-to-noise ratio, the remaining noise continues to mask the crucial information-bearing components of the speech signal.

In order to explain this effect, it is hypothesized that if the interfering noise occupies the same frequency region as the speech signal, then speech recognition will be reduced through direct masking. Further, even if the noise does not co-exist at the identical frequencies as the speech signal, masking will still occur, indirectly, if both the speech components and noise components occupy the same critical bands, as a result of the summation of the noise and the speech signal, or if there is substantial spread of masking across critical bands. In this investigation, the components of particular interest in the speech signal are the vocal harmonics.

One way of testing this hypothesis is to measure speech recognition performance

using a speech signal with a non-varying harmonic structure, masked by comb-filtered noise in which the peaks of the comb-filtered noise either fall on or between the harmonics of the speech signal, and to investigate by means of low-pass and high-pass filtering, the effect of critical bandwidth on speech recognition performance under these masking conditions.

CHAPTER 3

EXPERIMENTAL METHOD

In this experiment, synthetic speech with a fixed harmonic structure was used to produce twenty vowel-consonant (VC) nonsense syllables, ten of which represented a high-pitched female (high f_0) talker and ten of which represented a male (low f_0) talker. The consonant recognition performance of normal hearers was measured for these VCs in three noise types: (1) speech-spectrum shaped noise, (2) comb-filtered on-harmonic noise, and (3) comb-filtered between-harmonic noise. Three filter conditions were used with each of the preceding speech-in-noise conditions: (1) full-band speech and noise (no filter); low-pass speech and noise; and (3) high-pass speech and noise. These eighteen experimental conditions (two talkers in three noise conditions under three filter conditions) allowed for a direct test of the hypothesis that removing between-harmonic noise does not affect speech recognition if the speech and noise occupy separate critical bands.

It was anticipated that the on-harmonic noise would mask the speech signal at both low and high frequencies, but that the between-harmonic noise would mask the speech in the higher frequencies only, since at lower frequencies the noise and speech harmonics occur in separate critical bands. Thus, in the low-pass condition, a large difference was anticipated between the on-harmonic and between-harmonic noise conditions, whereas in the high-pass condition, no difference in speech recognition

performance was anticipated. For the full-band (unfiltered) condition, the anticipated scores were expected to lie between these two extremes. Similarly, performance in the speech-spectrum noise condition was expected to fall between the on-harmonic and between-harmonic noise masking conditions. Because the harmonics of the female talker were more widely spaced than those of the male talker, the separation of noise and speech harmonics into non-overlapping critical bands was greater for the female voice, and the difference between on-harmonic and between-harmonic performance for the female talker was expected to be greater than that for the male talker. Finally, although isolated voiceless consonants do not have an harmonic structure, the vowel-to-consonant formant transitions do. Hence, it was of interest to include both voiceless and voiced consonants in the investigation, with the expectation that the difference between the test scores for on-harmonic and between-harmonic noises on the voiced consonants would be greater than that for their voiceless cognates.

Subjects

Ten adults, four males and six females, served as subjects for this study. Their ages ranged from 22 to 45 years (mean of 31). All subjects were native speakers of English and met the following criteria at the time of their participation: (1) normal hearing sensitivity, defined as pure tone thresholds equal to or better than 20 dB HL for octave frequencies 250 Hz through 8000 Hz, bilaterally; and (2) normal otologic findings, defined as normal (Jerger Type A) tympanograms and normal contralateral acoustic reflex thresholds, bilaterally. Subjects were paid for their participation.

Test Stimuli

Test stimuli consisted of ten (10) synthesized vowel-consonant (VC) syllables in an /ɑC/ format. Syllables were chosen as stimuli since: (1) they are considered by many to constitute the basic unit of spoken language (Fry, 1964; Kozhevnikov and Chistovich, 1965), while remaining essentially free of the linguistic complexity and bias of words; (2) the stop consonants cannot be produced in isolation; and (3) the vowel-consonant transition is an important cue in correct listener identification of consonants (Delattre, Liberman, and Cooper, 1955; Harris, 1958; Liberman, et al., 1954). Synthetic speech was chosen since the fundamental frequency and harmonic structure of the syllables could be fixed precisely. The /ɑC/ context was chosen since /ɑ/ assumes a low-central tongue arch, a neutral lip posture, an open jaw position, and lax tongue and jaw musculature (Lindblom and Sundberg, 1971; Ling, 1976); moreover, the findings of Dubno and Levitt (1981) suggest good consonant recognition when combined with /ɑ/.

Syllables were generated on a DECtalk DCT01 speech synthesizer with a 120 Hz fundamental frequency (0 Hz variation) representing a male talker and with a 280 Hz fundamental frequency (0 Hz variation) representing a high-pitched female talker to yield the following vowel-consonant nonsense syllables:

/ɑp/ /ɑt/ /ɑk/ /ɑs/ /ɑf/ /ɑb/ /ɑd/ /ɑg/ /ɑz/ /ɑv/

Data from Peterson and Barney (1952) indicate the average fundamental frequency of /ɑ/ to be 124 Hz for adult male talkers, 212 Hz for adult female talkers, and 256 Hz for child talkers.

Each token was sampled at a rate of 10,000 samples per second and digitized with 12-bit accuracy. The time-waveforms of the tokens were edited to minimize vowel duration as a cue to final consonant voicing, to yield, for the female voice, an average VC duration of 413 ms (range of 393 ms to 428 ms) and an average vowel duration of 297 ms (range of 277 ms to 320 ms), and, for the male voice, an average VC duration of 397 ms (range of 385 ms to 414 ms) and an average vowel duration of 277 ms (range of 267 ms to 321 ms). This was accomplished by removing an appropriate number of pitch periods from the steady-state portions of the vowels. Each token was also scaled for intensity to yield a peak syllable intensity of 70 dB (re: the digital reference corresponding to 0 dB). Waveform displays and average amplitude spectra for the female and male tokens are shown in Appendix 1. A sampling rate of 20,000 Hz was used in deriving these spectra. A time window of 51.2 msec was used for the short-term spectrum analysis, yielding a frequency resolution of 19.5 Hz.

Two types of noise were used in these tests. One of these was speech-spectrum shaped noise. The other one consisted of a sequence of narrow-band noises spaced uniformly throughout the spectrum. These noises were generated by comb-filtering wide-band speech-spectrum shaped noise. The bandwidth of each narrow band of noise was one-fourth of the fundamental frequency of the speech stimuli, i.e., 30 Hz for the male talker and 70 Hz for the high-pitched female talker. The noise intensity was adjusted so that the root-mean-square (rms) levels of the speech-spectrum shaped noise and the narrow-band on-harmonic and between-harmonic noises were the same. In the on-

harmonic condition, the noise bands were centered over the speech harmonics. In the between-harmonic condition, the noise bands fell mid-way between the harmonics. Amplitude spectra for the noises are shown in Appendix 1.

In the filtered conditions, signals and noises were low-pass and high-pass filtered at 800 Hz for tokens produced by the male speaker and at 1200 Hz for tokens produced by the female speaker. The cut-off frequencies were chosen based on Scharf's data regarding the critical bandwidths. For the male talker, the cut-off of 800 Hz represented the frequency about which the critical band began to encompass one harmonic and two noise bands. For the female talker, the cut-off of 1200 Hz represented the frequency around which the critical band contained one harmonic and one noise band. Preliminary VC recognition tests in which the cut-off frequencies were manipulated from 600 Hz to 1500 Hz were conducted using two listeners. The scores obtained at the chosen cut-off frequencies suggested these to be reasonable for the experiment. Amplitude spectra for the filtered signals and filtered noises are also shown in Appendix 1.

Instrumentation

A schematic diagram of the test instrumentation is shown in figure 2. For all of the tests, the speech and noises were reproduced by computer. For each test stimulus, one of twenty VCs and one of the three corresponding noises were generated. The speech output of the computer was connected to the line input of a Shure M267 mixer and the noise output of the computer was connected to an HP 350D external attenuator (for manipulation by the experimenter to establish the signal-to-noise ratio for the test session)

and then to the input of the mixer. The line output of the mixer was connected to the input of a dual channel Stanford Research Systems SR 650 computer-controllable digital filter that was set by the computer to provide either full-band (10,000 Hz bandwidth), high-pass (800 Hz cut-off frequency for the male tokens and 1200 Hz cut-off frequency for the female tokens) or low-pass (same cut-off frequencies as high-pass) filtering. The output of the filter was connected to a Realistic SA-150 amplifier, which was adjusted by the subject to a comfortable listening level after the headphones (Etymotic Research ER-2) were placed. Insert earphones were chosen due to their superiority over supra-aural earphones in terms of comfort, background noise attenuation, and frequency response (Killion, 1984). Calibration of the levels and spectral characteristics of the speech and noise was determined through computer analyses of digitized recordings of the stimuli. These recordings were generated by use of the research instrumentation described above, coupled to a Bruel and Kjaer sound level meter, the output of which was recorded on a Denon DTR-80P digital tape recorder.

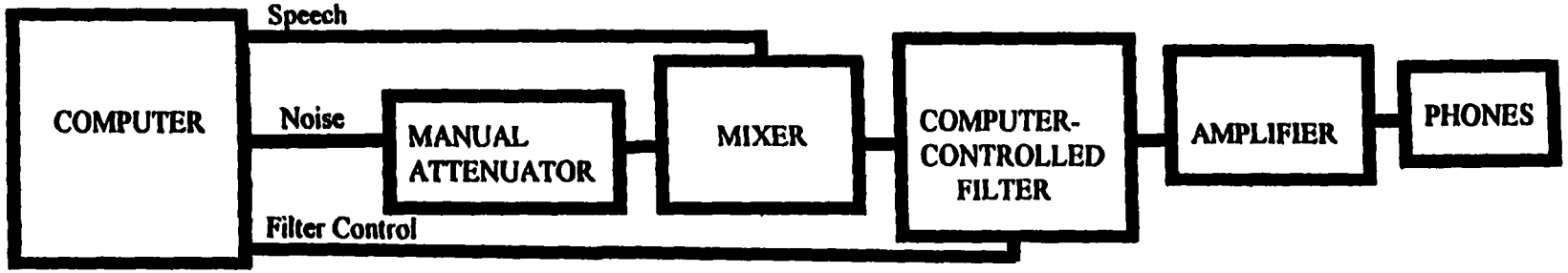


Figure 2. Instrumentation Schematic

Procedure

Subjects were seated in a sound-treated room and given consent forms and printed instructions concerning the task (see Appendix 2). Each subject was given an opportunity to ask questions and familiarize him/herself with the computer keyboard, ten keys of which were clearly labeled with the test syllables. A test session began with a preliminary practice period during which the subject was able to listen to each test syllable (in quiet) at his/her leisure by pressing the appropriate keyboard key. When the subject was sufficiently comfortable with the stimuli, he/she began a formal test session. A formal test session consisted of 1800 tokens (10 syllables x 2 speakers x 3 noises x 3 filter conditions x 10 repetitions); subjects were free to complete a test session over one sitting or more sittings, at their own discretion. Subjects were also free to listen repeatedly to any token as often as they wished.

Presentation of tokens in the formal test session was randomized over: (1) talker, (2) noise-type, (3) filter condition and (4) syllable. When a decision was made as to which test stimulus was presented, the subject pressed the appropriate keyboard key and the next test stimulus item was automatically presented. Four test sessions were conducted with each subject at signal-to-noise ratios of +3, +1, -1, and -5 dB. Each test session was conducted at a constant signal-to-noise ratio. A different signal-to-noise ratio was selected at random for each test session. The choice of signal-to-noise ratios was based on the results of a pilot study in which data were obtained at seven different signal-to-noise ratios to determine the range that would encompass intelligibility scores ranging from at or near maximum (+ 3 dB) down to scores near chance performance (-5 dB).

CHAPTER 4

RESULTS

The percent correct score was computed for each consonant in each test condition for each subject. The data were then subjected to an arcsine-transformation in order to stabilize error variance (Winer, 1971). A limitation of the statistical package used in this analysis was that only 5 factors could be analyzed at a time. Since higher-order interactions with the filter factor were not of interest in this investigation, three separate repeated-measures analyses of variance (ANOVA) were performed for the factors Noise-type (speech-spectrum, narrow-band on-harmonic, narrow-band between-harmonic), Gender (female talker, male talker), Consonant-voicing (voiceless, voiced), and Signal-to-Noise Ratio (+3, +1, -1, -5), repeated over 10 Subjects. The results of the three analyses of variance are shown in table 4.

Although the primary interest concerning the consonants was in comparing performance between voiceless and voiced consonants, it was also possible to analyze the data with respect to the effect on individual consonants. These analyses are shown in Appendix 3. Also shown in the appendices (see Appendix 4) are figures depicting mean performance for each consonant in each noise-type for each talker as a function of signal-to-noise ratio.

Table 4. --ANOVA Results for 3 Filter Conditions for the Factors Noise-type (N), Gender (G), Voicing (V), and Signal-to-Noise Ratio (S/N)

		Full-Band (No Filter)		Low-Pass Filter		High-Pass Filter	
SOURCE	**df**	**F**	**p**	**F**	**p**	**F**	**p**
Noise (N)	2, 18	37.647	**<0.001**	29.350	**<0.001**	8.833	0.002
Gender (G)	1, 9	0.217	0.655	1.489	0.253	0.038	0.844
N x G	2, 18	2.120	0.148	12.110	**<0.001**	2.138	0.145
Cons-Voicing (V)	1, 9	9.121	0.014	9.742	0.012	0.018	0.892
N x V	2, 18	23.188	**<0.001**	16.813	**<0.001**	35.022	**<0.001**
G x V	1, 9	5.192	0.047	19.339	0.002	51.035	**<0.001**
N x G x V	2, 18	2.481	0.110	1.718	0.207	1.470	0.256
S/N	3, 27	97.946	**<0.001**	39.198	**<0.001**	69.030	**<0.001**
N x S/N	6, 54	10.967	**<0.001**	3.361	0.007	1.825	0.111
G x S/N	3, 27	16.492	**<0.001**	2.588	0.073	1.020	0.400
N x G x S/N	6, 54	1.517	0.190	9.191	**<0.001**	1.185	0.328
V x S/N	3, 27	8.398	**<0.001**	5.050	0.007	8.323	**<0.001**
N x V x S/N	6, 54	3.600	0.005	1.451	0.212	3.422	0.006
G x V x S/N	3, 27	3.396	0.032	0.339	0.800	14.058	**<0.001**
N x G x V x S/N	6, 54	2.304	0.047	1.612	0.161	4.705	**<0.001**

Note: Underlining indicates statistical significance. Bold-type plus underlining indicates statistical significance below the .001 level.

A test for homogeneity of the subjects was also carried out. Performance scores were transformed to z-scores and rank-ordered. The z-scores were then plotted against their rankings. If the data are from a normal distribution, the resulting plot should fall on a straight line. Figure 3 shows the data for the ten subjects in this study. As can be seen, the data are approximated by a straight line, indicating that the subjects' scores are approximately normally distributed.

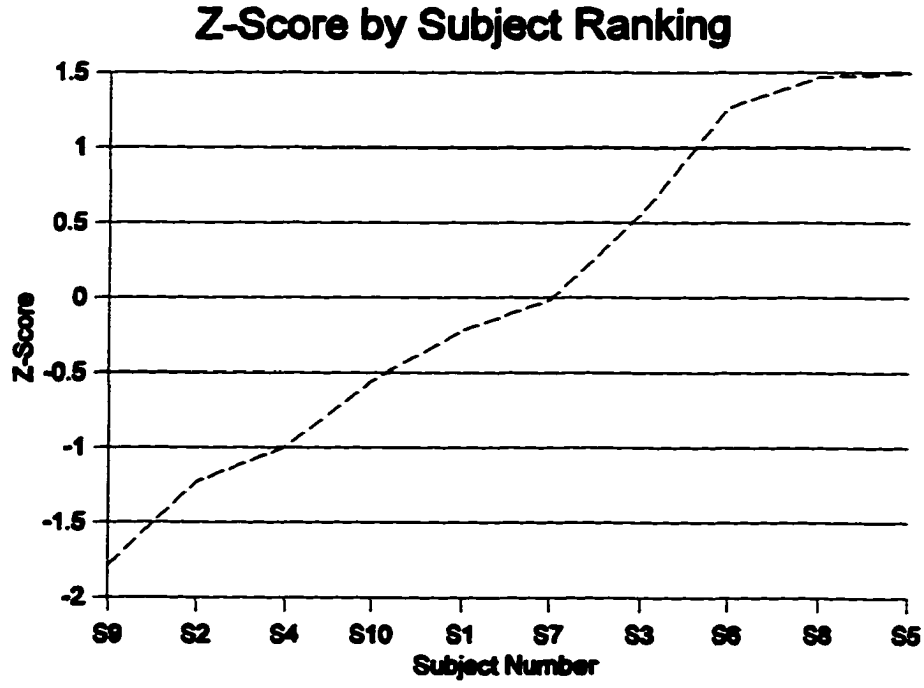


Figure 3. Plot of Subjects' z-scores by rankings.

Main Effects

Significant main effects were observed for Noise-type (N) and Signal-to-Noise Ratio (S/N) in all filter conditions. Significance levels were less than .001 except for Noise-type in the High-Pass filter condition for which a slightly higher significance level was obtained ($p < .004$). Consonant-voicing (V) was significant ($p < 0.015$) for the Full-Band and Low-Pass filter conditions, but not for the High-Pass filter condition. The post-hoc Tukey honestly significant difference (HSD) tests (Tukey, 1953) for Noise-type at the .01 level of significance showed that scores for the between-harmonic (BH) noise were significantly higher than scores for the speech-spectrum (SS) noise and the on-harmonic (OH) noise in the full-band (FB) and low-pass (LP) filter conditions, and that scores for the BH noise did not differ significantly from scores for the SS noise or those for the OH noise in the high-pass (HP) filter condition.

The post-hoc Tukey HSD tests (.01 significance level) for S/N revealed that scores were significantly poorer with each decrease in signal-to-noise ratio in the FB and HP filter conditions and that, in the LP filter condition, the -5 dB S/N was significantly poorer than all of the other levels, and that the -1 dB S/N was significantly poorer than the +3 dB S/N. Scores for the +1 dB S/N did not differ significantly from those in the +3 dB S/N or those in the -1 dB S/N in the LP filter condition.

The post-hoc Tukey HSD tests (.05 significance level) for Consonant-voicing in the FB and LP filter conditions showed that scores on the voiced consonants were significantly better than those on the voiceless consonants. In the HP filter condition, the Consonant-voicing effect was not significant.

Interactions

Significant interactions ($p \leq 0.05$) were obtained for Noise-type x Consonant-voicing, Gender x Consonant-voicing, and Consonant-voicing x S/N in all filter conditions. The interactions Noise-type x S/N, Noise-type x Consonant-voicing x S/N, Gender by Consonant-voicing x S/N, and Noise-type x Gender x Consonant-voicing x S/N were found to be significant in two of the three filter conditions. The interactions Noise-type x Gender and Noise-type x Gender x S/N were significant only in the LP filter condition. The interactions of specific interest in this study were those involving Noise-type, and these are analyzed further in the section that follows.

The post-hoc Tukey HSD tests (.05 significance level) for Noise-type x Consonant-voicing showed significantly higher scores for the voiced consonants in BH noise for the FB and LP filter conditions than for any of the other five interaction terms. In the HP filter condition, significantly higher scores were seen for the voiceless consonants in SS noise than for any of the other five interaction terms. The lowest scores were obtained consistently for the voiceless consonants in OH noise in all filter conditions. Figure 4 shows the histograms for each of the Noise-type by Consonant-voicing interaction terms. The scores are expressed in arcsine units. High bars above a common horizontal line in figure 4 are not significantly different from each other at the .05 level of significance.

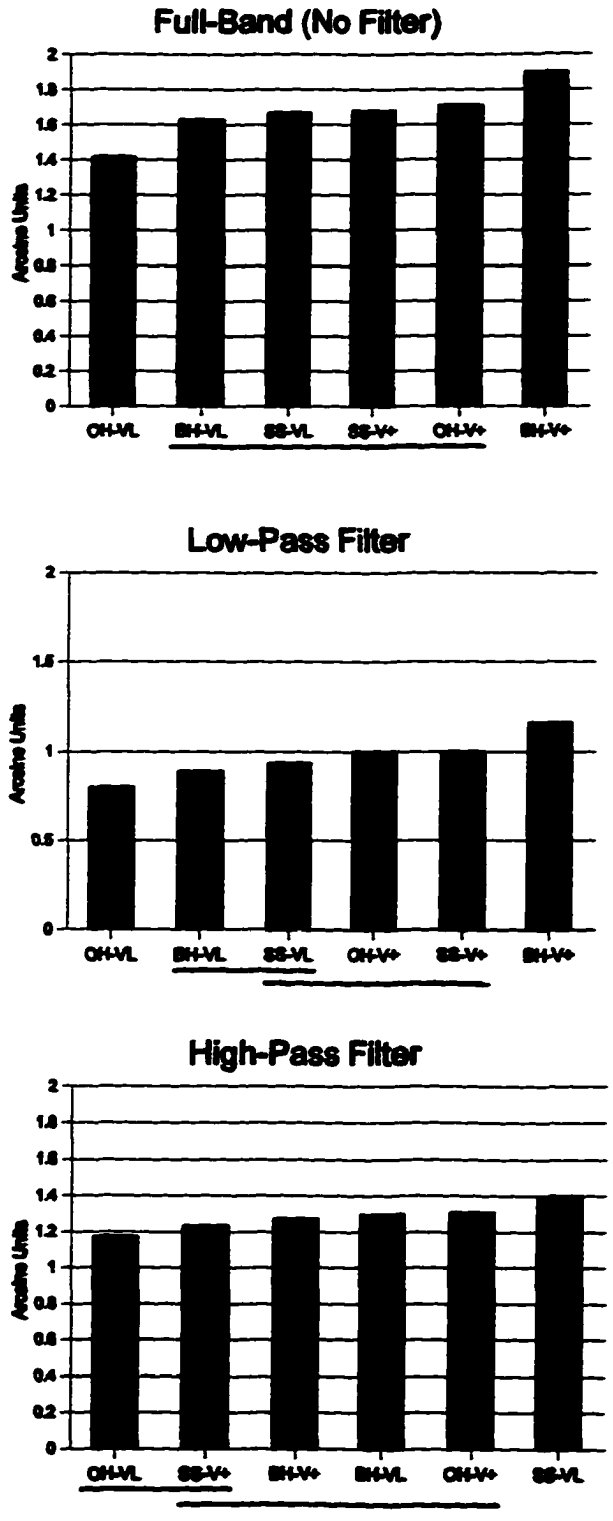
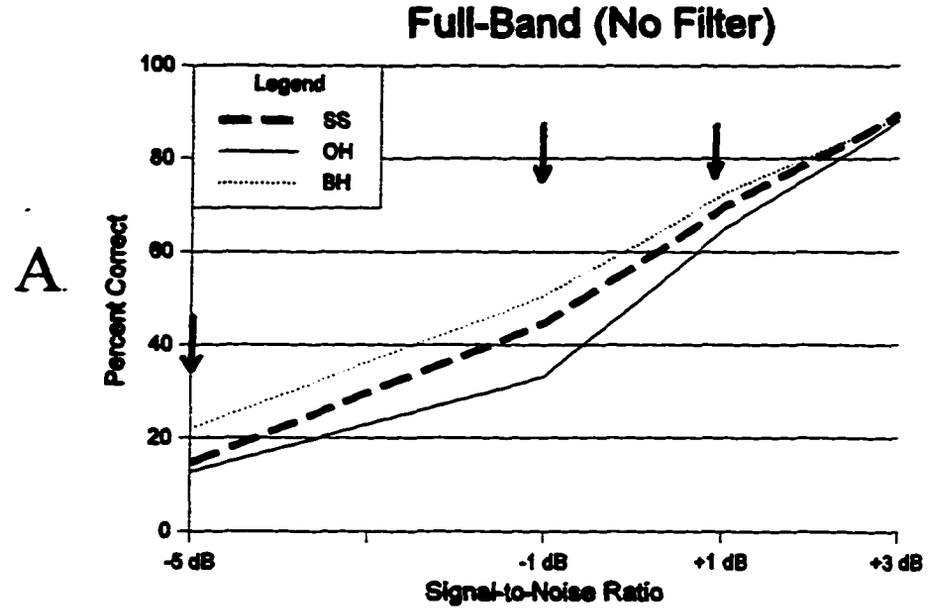
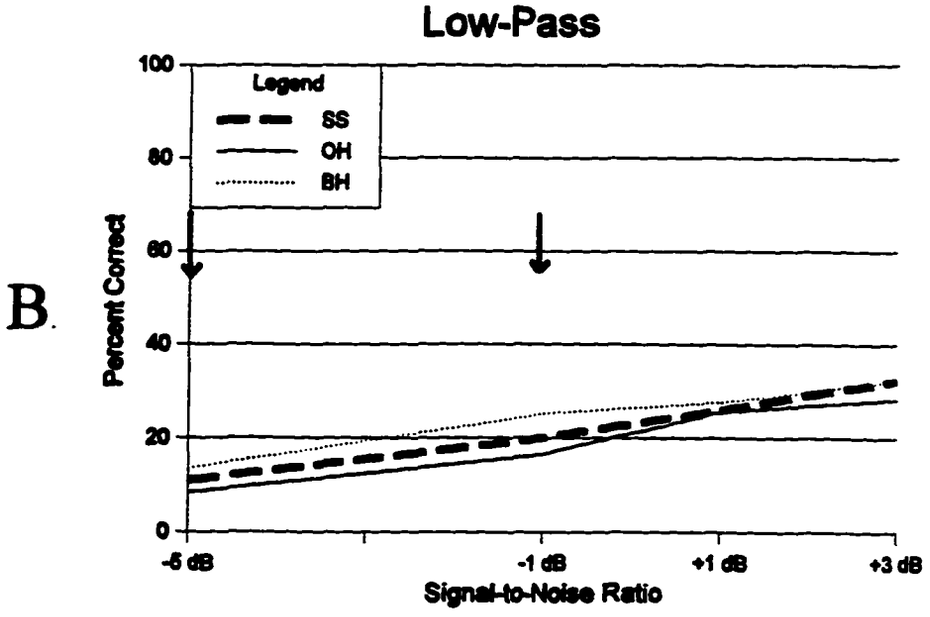


Figure 4. Histograms of Noise-type by Consonant-voicing (voiceless = VL; voiced = V+) interaction (post-hoc test).

The interaction of greatest interest was that of Noise-type x S/N. Figure 5 illustrates this interaction for each filter. As can be seen, scores for the BH noise were consistently higher than those for the OH noise, except in the HP filter condition.





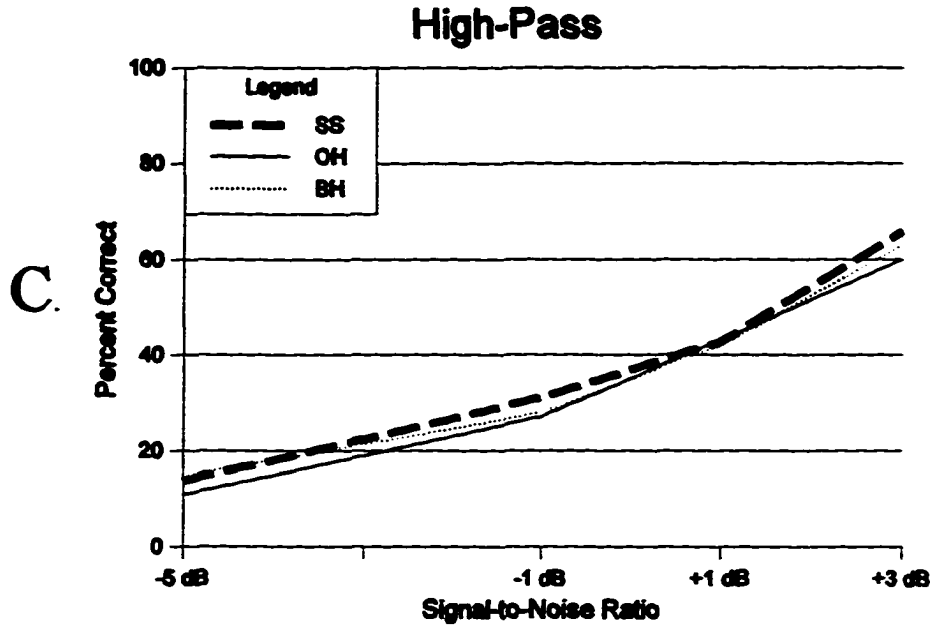


Figure 5. Comparison of performance in speech-spectrum noise, on-harmonic noise and between-harmonic noise at each signal-to-noise ratio for the (A) full-band, (B) low-pass, and (C) high-pass filter conditions.

The post-hoc Tukey test (.01 significance level) for the FB filter condition showed that scores for the BH noise were significantly higher than those for the OH noise except at the highest S/N, at which level there were no significant differences among the noise types. Scores for the OH noise did not differ significantly from those for the SS noise except at the -1 dB S/N where scores for the OH noise were significantly poorer than those for the SS noise.

The post-hoc Tukey test (.01 significance level) for the LP filter condition showed that scores for the BH noise were significantly higher than those for the OH noise at the

two lowest signal-to-noise ratios (-1 dB S/N and -5 dB S/N). At these levels, scores for the OH noise again did not differ significantly from those for the SS noise. For each of the two highest signal-to-noise ratios (+3 dB S/N and +1 dB S/N), the scores for the BH noise were higher than those for the OH noise, but the difference was not statistically significant; indeed, there was no significant difference in scores among the three noise-types (SS, OH, BH) at these levels in the LP filter condition.

In the HP filter condition, there were no significant differences among the three noise types at any of the signal-to-noise ratios tested. In fact, the greatest difference between scores in the HP filter condition at any level was less than 6%.

In summary, scores for the BH noise were significantly higher than those for the OH noise at the three lowest signal-to-noise ratios (+1 dB S/N, -1 dB S/N, and -5 dB S/N) in the full-band condition. In the LP filter condition, scores for the BH noise were significantly higher than those for the OH noise at the two lowest signal-to-noise ratios (-1 dB S/N and -5 dB S/N). The small vertical arrows in figure 5 identify the signal-to-noise ratios at which the significant differences were obtained.

Also of interest was the Noise-type x Gender interaction which was significant only in the LP filter condition. Figure 6 shows the histograms for this interaction according to percent correct scores.

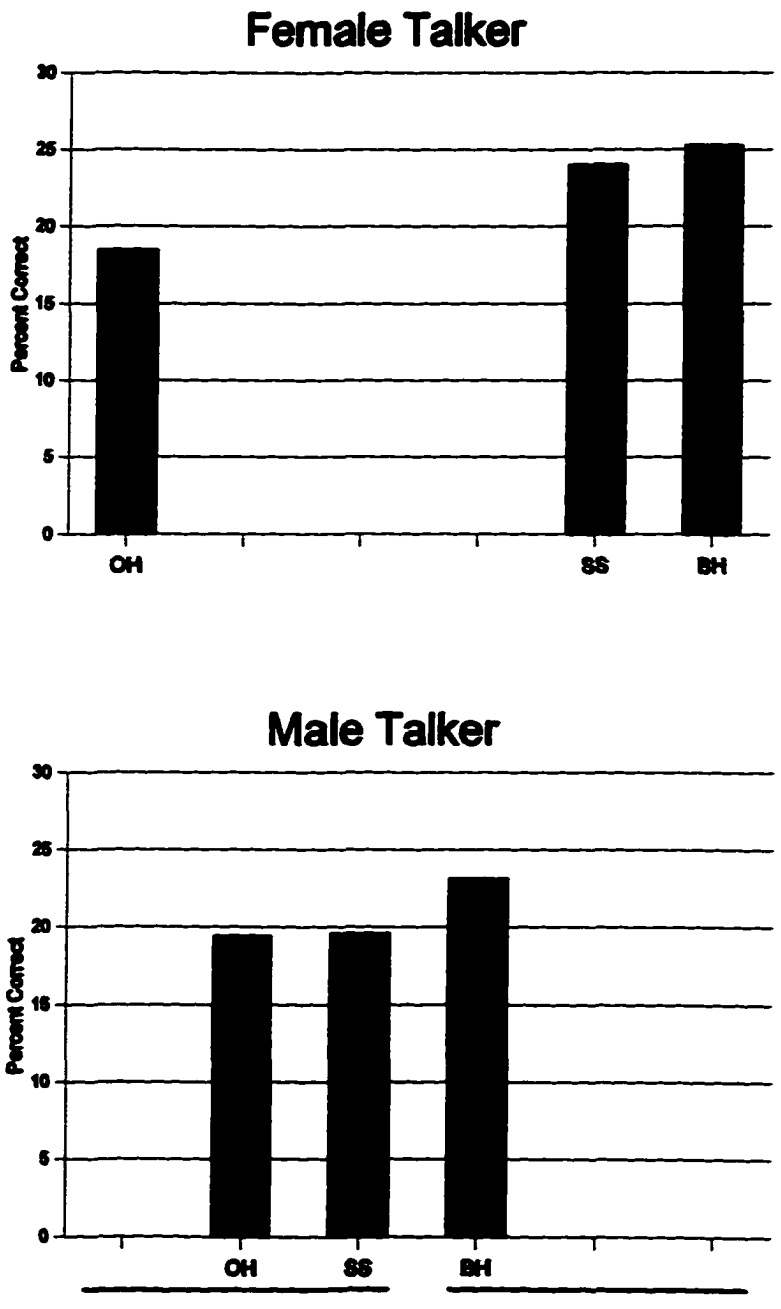


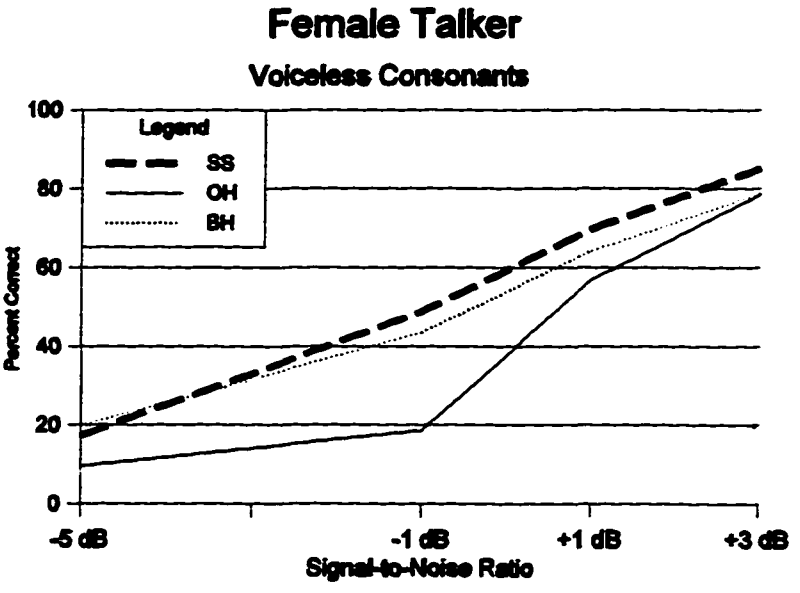
Figure 6. Histograms of percent correct for the Noise-type by Gender interaction for the Low-Pass filter condition (post-hoc Tukey test).

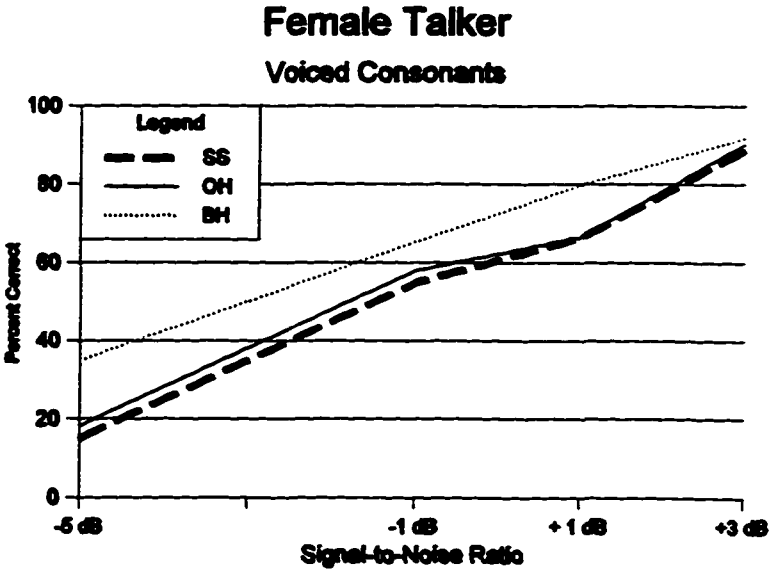
The post-hoc Tukey test (.01 significance level) showed that scores for the SS noise for the female voice and scores for the BH noise for the female and male voices were significantly better than scores for the SS noise for the male voice and scores for the OH noise for the female and male voices. Although scores for tokens produced by the female voice for BH noise were better than those for the male voice, the difference was not significant. The high bars above a common horizontal line in figure 6 are those which are not significantly different from one another.

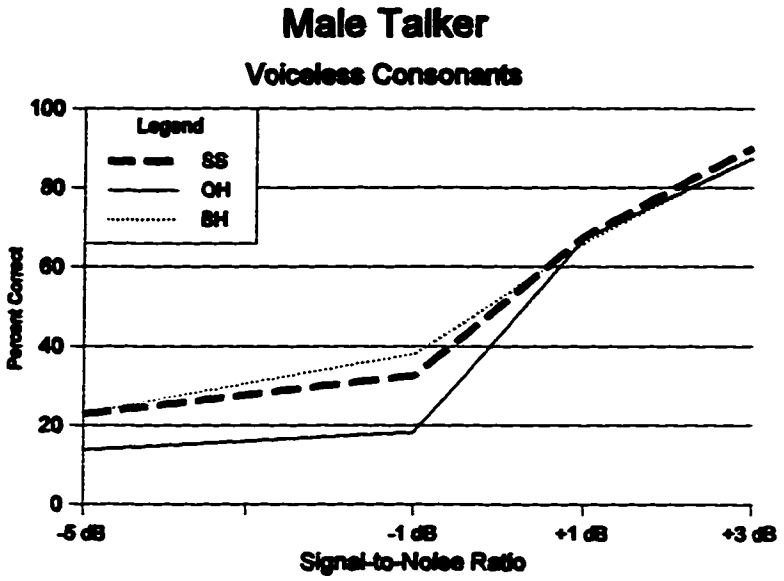
The two most complex interactions were Noise-type x Gender x S/N (significant in the LP filter condition), and Noise-type x Gender x Consonant-voicing x S/N (significant in the FB and HP filter conditions). The nature of these is evident from figure 7. A post-hoc analysis could not be performed on these interactions due to the large number of means for comparison; however, several general observations can be made concerning this figure. In the FB filter condition (figure 7 A), (1) scores for the BH noise were better than those for the OH noise at all signal-to-noise ratios but the highest one; (2) scores for the female talker were better than those for the male talker at poorer signal-to-noise ratios; and (3) scores for the voiced consonants tended to be higher than those for the voiceless consonants. In the LP filter condition (Fig. 7 B), (1) scores for the BH noise were better than those for the OH noise, particularly on the voiced consonants; (2) slightly better performance was seen for the female talker than for the male talker, again more so on the voiced consonants; and (3) performance on the voiced consonants tended to be better than that on the voiceless consonants. In the HP filter condition (Fig. 7 C), (1) scores for the BH noise showed no real advantage over those for the OH noise, except possibly for the

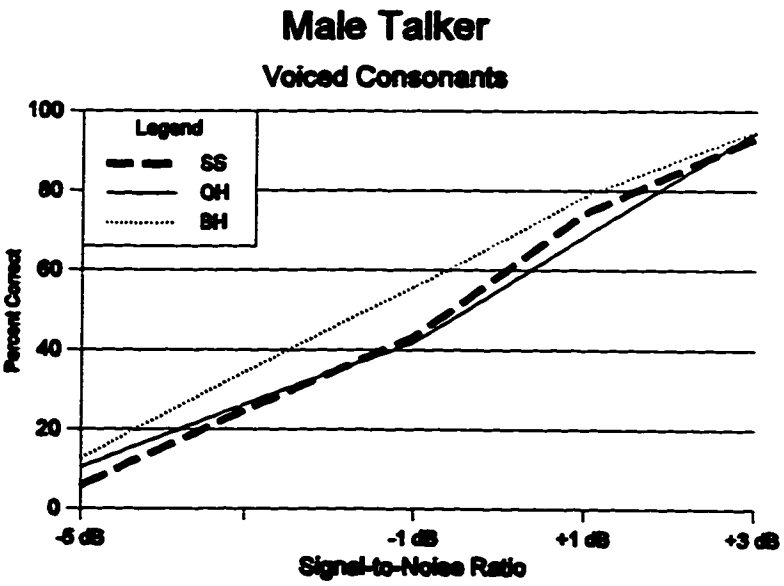
possibly for the male talker at poorer signal-to-noise ratios, and (2) the interaction between talker gender and voicing was complex, but tended to show better performance for the male voiceless tokens than for the female voiceless tokens, and for the female voiced tokens than for the male voiced tokens.

A

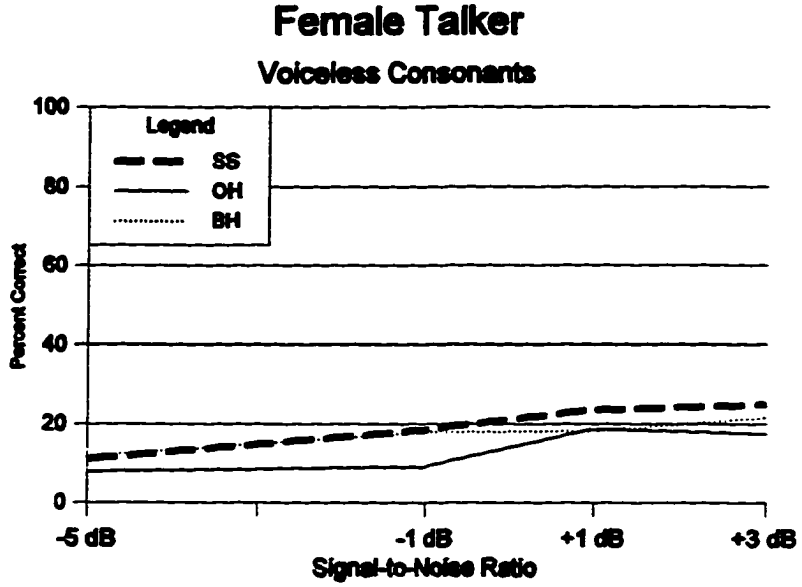


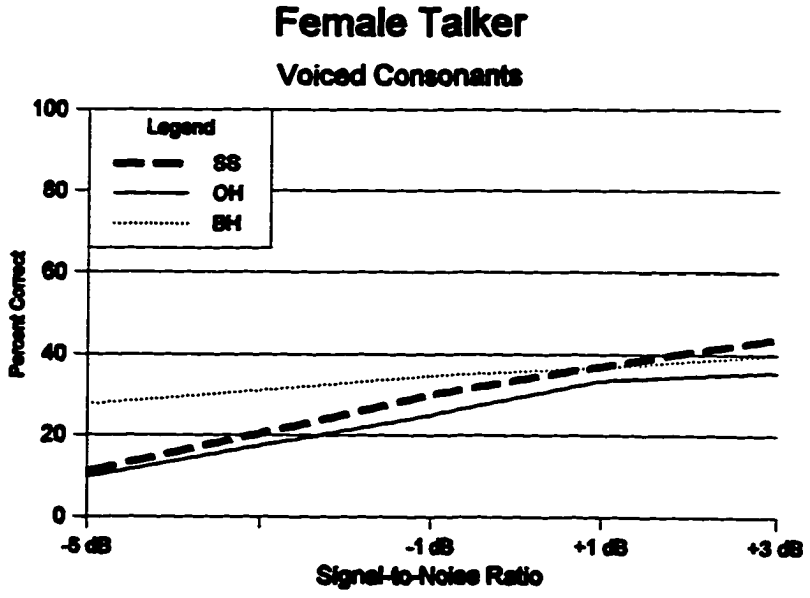


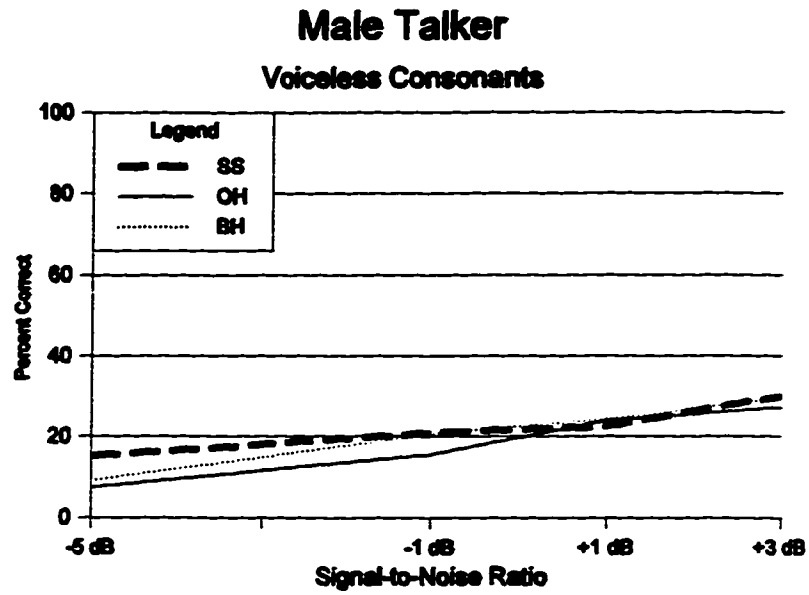


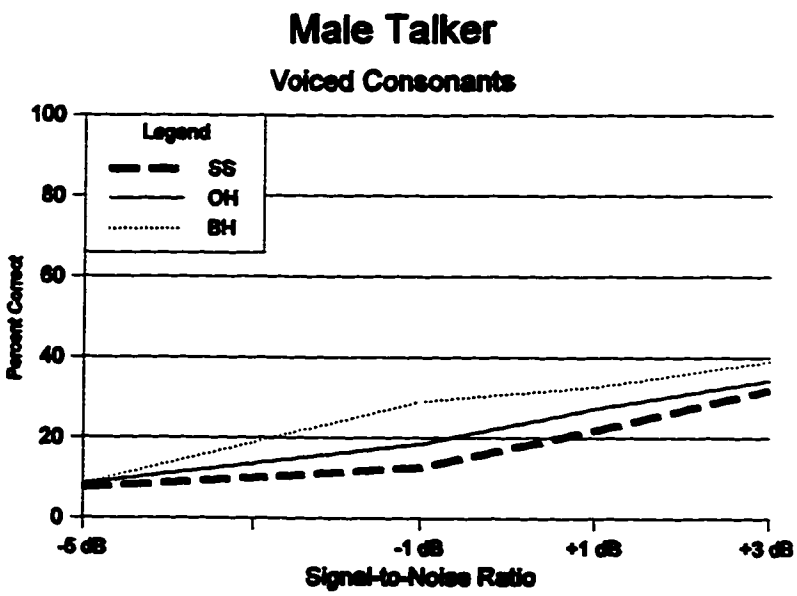


B.

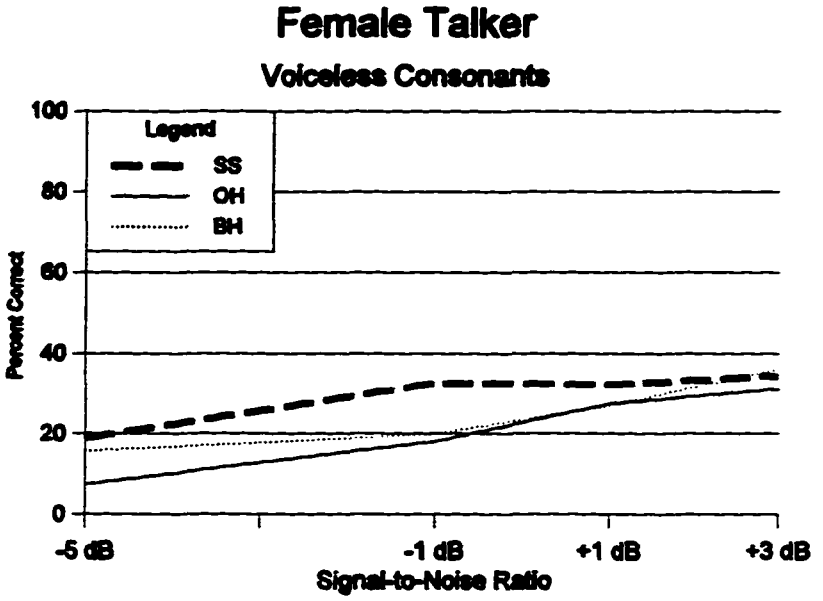


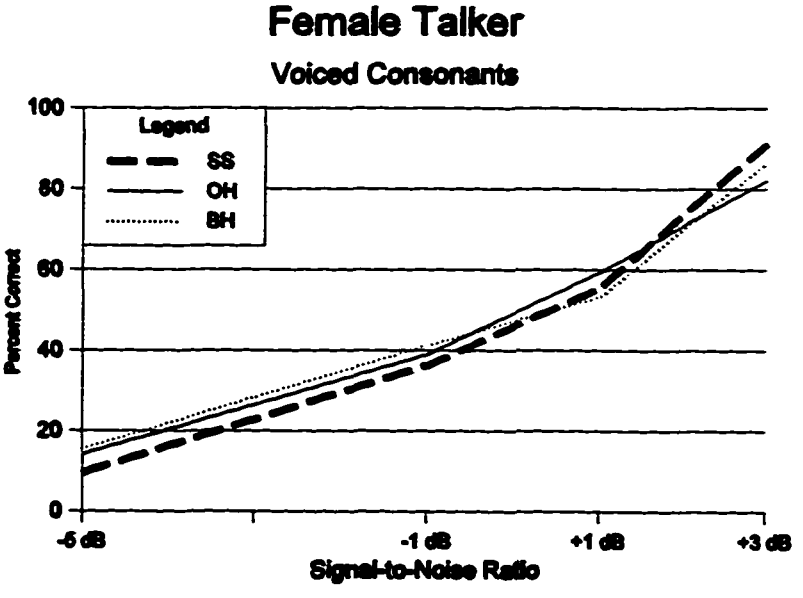






C.





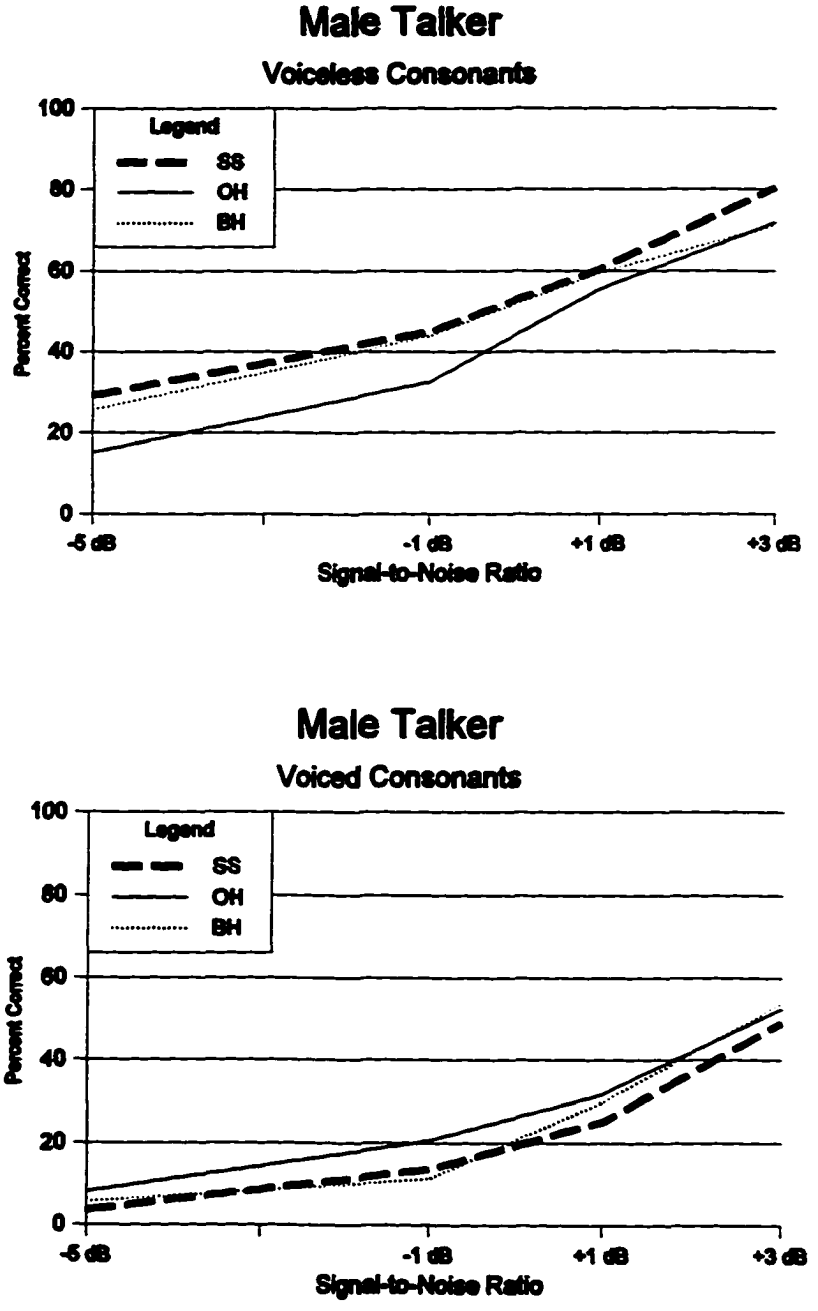


Figure 7. Comparison of performance on the voiced vs. voiceless consonants as produced by the female and male talker in the: (A) Full-Band, (B) Low-Pass, and (C) High-Pass filter conditions.

CHAPTER 5

DISCUSSION

The purpose of this study was to test the hypothesis that between-harmonic noise would have less effect on speech recognition performance than on-harmonic noise, provided the speech and noise occupy separate critical bands. To examine this, speech recognition performance in both on-harmonic and between-harmonic noise was measured. Wide-band speech-spectrum shaped noise was also included as a reference condition. Since the width of the critical band increases logarithmically with frequency, both speech harmonics and comb-filtered noise will co-exist in the same critical bands at higher frequencies. Thus, in order to test the hypothesis directly, the speech-plus-noise was subjected to both low-pass and high-pass filtering. For the low-pass case, the speech harmonics and comb-filtered noise occupy different critical bands for the between-harmonic noise, but co-exist in the same critical bands for on-harmonic noise. For the high-pass case, both on-harmonic and between-harmonic noise occupy the same critical bands as the speech harmonics.

According to the underlying hypothesis, only those portions of the noise spectrum that lie in the same critical bands as the speech spectrum will be effective in masking the speech. Those portions of the noise spectrum lying in separate critical bands will be relatively ineffective in masking the speech. Speech recognition scores for the low-pass

filter condition should, therefore, be significantly higher for between-harmonic noise than for on-harmonic noise. For the high-pass filter condition, however, there should be no difference in the masking effect of the between-harmonic and on-harmonic noise since both of these noises occupy the same critical bands as the speech harmonics and summation will occur.

Predictions for the speech-spectrum noise follow a similar pattern. All three noises were adjusted to have the same overall power. Hence, for the high-pass filter condition, in which both the noise and speech harmonics occupy the same critical bands, speech recognition scores should be the same for all three noises. For the low-pass filter condition, however, the speech-spectrum noise occupies all critical bands and, as a consequence, only serves as an effective masker for those critical bands containing a speech harmonic. Speech recognition performance for speech-spectrum noise should thus be higher than that for on-harmonic noise, which always occupies the same critical bands as the speech harmonics, but poorer than that for between-harmonic noise which does not occupy any critical band containing a speech harmonic within the frequency range of the low-pass filter.

Predictions for the unfiltered, full-band condition follow the same logic. To begin with, speech recognition scores for the full-band condition will be substantially higher than for either the low-pass or high-pass filter conditions. As before, the on-harmonic noise is the most effective masker since it always occupies the same critical bands as the speech harmonics. The between-harmonic noise is less effective as a masker since it only occupies the same critical bands as the speech in the high-frequency region of the

spectrum. The masking effectiveness of the speech-spectrum noise should fall roughly mid-way between that of the on-harmonic and between-harmonic noise since in the low-frequency region in which the speech harmonics occupy roughly every second critical band, the power of the speech spectrum noise is divided almost equally between those critical bands that contain a speech harmonic and those that do not.

The results of the investigation support the above predictions, although the magnitude of the differences in speech recognition scores was smaller than anticipated. As shown in figure 5 B (low-pass filter condition), speech recognition scores for between-harmonic noise were consistently higher than those for on-harmonic noise, but the difference was only statistically significant at the two lowest signal-to-noise ratios. At these two signal-to-noise ratios, speech recognition scores for the speech-spectrum noise were, as predicted, roughly mid-way between the scores for the between-harmonic and on-harmonic noise. At higher signal-to-noise ratios (+1 dB and +3 dB), the performance-intensity functions for all three noises appear to have saturated and no significant differences among the three noise-types were seen.

For the high-pass filter condition, no difference in masking was predicted for the three noise-types and, in accordance with these predictions, the performance-intensity functions were essentially the same for all three noises. As shown in figure 5 C, there were no significant differences among the three noise-types at any of the four speech-to-noise ratios.

The data for the unfiltered, full-band condition (figure 5 A) were similarly consistent with the predictions outlined above. As before, the highest speech recognition

scores were obtained for the between-harmonic noise and the lowest scores were obtained for the on-harmonic noise. The scores were substantially higher than for either the low-pass or high-pass filter conditions with no evidence of saturation in the performance-intensity functions until the highest signal-to-noise ratio (+3 dB), this being the only level at which the difference in speech recognition test scores between the on-harmonic and between-harmonic noise was not statistically significant.

While the largest difference between the on-harmonic and between-harmonic noise scores was seen in the full-band condition, the proportion of change, relative to the maximum achievable score, was greatest in the low-pass filter condition. That is, given the average maximum score of 88% in the full-band condition, a difference score of approximately 9.8% represents a proportion of about 0.11; given the average maximum score of 32% in the low-pass condition, a difference score of approximately 5.3% represents a proportion of about 0.17. So, although the absolute difference between on-harmonic and between-harmonic scores was largest in the full-band condition, the relative difference was largest in the low-pass filter condition, supporting the hypothesis that the difference in performance between on-harmonic and between-harmonic noise would be greatest in the low-pass filter condition because the harmonics and speech occupied separate critical bands. Also contributing to the apparently stronger effect seen in the full-band condition is the lack of distortion introduced by the filter in the low-pass filter condition, and the fact that the low-pass filter produced relatively low scores, notably at the lowest signal-to-noise ratios where increased guessing was particularly evident.

While the results in figure 5 supported the original hypothesis, as mentioned, the

obtained difference scores were not as large as anticipated. Several factors may have contributed to this, the most obvious of which was the use of synthetic speech as stimuli. While using synthetic speech allowed for precise control of the harmonic structure of the voices, synthesized speech lacks the nuances and variations of natural speech, perhaps leading to lower scores and, in some cases, random guessing. Also, including consonants of different manners of production (i.e., glides, nasals, and affricates) might have increased the effect. Another factor is the assumed independence of the critical bands, which is not true, since, in fact, the bands overlap, and complete separation of speech and noise does not occur.

Related to the major hypothesis were the secondary hypotheses that stated that the difference between on-harmonic and between-harmonic performance would be larger for the female voice than for the male voice because the harmonic spacing of the female voice was greater than that of the male voice, and that the difference between on-harmonic and between-harmonic performance would be larger for the voiced consonants than for the voiceless consonants because of the fuller harmonic structure of the voiced consonants.

The results support the hypothesis that the difference between on-harmonic and between-harmonic performance would be larger for the female voice than for the male voice, although not fully. It was anticipated that this effect would be seen in both the full-band and low-pass filter conditions, although it was expected to be more dramatic in the low-pass filter condition. However, a significant Noise-type x Gender interaction was only found in the low-pass filter condition, and as can be seen in figure 6, while the magnitude of difference between on-harmonic scores and between-harmonic scores was

greater for the female talker than for the male talker, scores for tokens produced by the female talker in the between-harmonic noise were not significantly different from scores produced by the female talker in the speech-spectrum noise or from scores produced by the male talker in the between-harmonic noise. Perhaps the difference between talkers in the on-harmonic versus between-harmonic conditions would have been larger if the difference between the fundamental frequencies of the two talkers had been greater (i.e., the male f_0 made lower and the female f_0 made higher, spacing the harmonics for the male more closely and those for the female more widely), or if the cut-off frequency for the low-pass filtering of the female tokens had been increased. Also, the difference may have been greater if natural speech had been used.

With regard to the hypothesis that the difference between on-harmonic and between-harmonic performance would be larger for the voiced consonants than for the voiceless consonants, figure 4 shows that performance for the between-harmonic noise was significantly better than that for the on-harmonic noise (or the speech-spectrum noise) in the full-band and low-pass filter conditions for the voiced consonants. Obviously, recognition for those syllables which maintain voicing from the vowel to the consonant is considerably more affected by noise directly on the harmonics, and performance in that noise-type is significantly poorer than in the between-harmonic noise.

It can also be seen in figure 4 that performance for the on-harmonic noise on voiceless consonants was significantly poorer than that for the other two noise-types in the full-band and low-pass filter conditions. This finding is not easily explained, although it reflects the generally poorer performance seen on the voiceless consonants in all

conditions. Moreover, although the voiceless consonants in isolation do not have harmonic structure, information concerning the consonant is available through the VC transition. It is possible that the on-harmonic noise in the full-band and low-pass filter conditions had an effect on the voiceless consonants at the VC transition that significantly reduced consonant recognition over that for the between-harmonic noise or speech-spectrum shaped noise.

In the high-pass filter condition, performance on the voiceless consonants in the speech-spectrum noise was significantly better than that for the other noise-types for the voiced and voiceless consonants. No significant differences were seen among the remaining interactions, except for performance in the on-harmonic noise for the voiceless consonants, which was significantly poorer than that for all other interactions, except for performance on the voiced consonants in speech-spectrum noise. No significant difference was expected in the high-pass filter condition, since it was hypothesized that indirect masking would affect the recognition of all consonants equally. With the exception of the voiceless consonants in the speech-spectrum shaped noise, this was true. Why the voiceless consonants were more correctly recognized in the speech-spectrum noise in the high-pass filter condition is not entirely clear, but may have to do with the broader spectrum of the speech-shaped noise. A thorough analysis of error patterns through the use of a confusion matrix for each consonant for each condition might help to account for these unexpected findings.

The results of this investigation support the hypothesis that if the interfering noise is either at the same frequencies as those of the speech signal, or within the same critical

bands, then speech recognition will be reduced. However, if the interfering noise falls in critical bands not occupied by the speech signal, then the noise will not reduce speech recognition appreciably. Conversely, removing that noise will not improve speech recognition performance. This finding has implications for the relative lack of improved intelligibility seen in the current single-microphone noise reduction strategies and has implications for future research in noise reduction.

CHAPTER 6

SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

The results of this study demonstrate that the spectrum of the noise which co-exists in time with speech has an influence on the amount of masking that noise will have on the speech, and therefore, on the intelligibility of the speech. Specifically, noise that occupies identical frequency regions as the crucial components of speech will directly mask the speech, and intelligibility will be decreased. Moreover, noise that occupies non-overlapping frequency regions as the speech components, but that is contained within the same critical bands as the speech components, will also mask the speech, this time indirectly, through summation within a critical band, and intelligibility will likewise be reduced. Consequently, removing noise from the speech will improve intelligibility if the frequency regions of the speech components and the noise components occupy the same critical bands; if the speech components and noise components occupy different critical bands, little or no improvement in speech recognition will result.

This result is consistent with the Articulation Index (AI). According to the AI, speech recognition will be reduced as more of the speech spectrum is masked by an interfering noise. If the interfering noise occurs in the same AI bands as the speech, intelligibility will be reduced. If the harmonics of the speech signal occupy separate AI bands from those occupied by the noise, then reducing the noise should not reduce the AI,

and hence, should not reduce intelligibility. In this context, the bands of equal importance to speech recognition are treated in the same way as the critical bands. That is, if the speech and noise occupy different AI bands, then reducing the noise will not show a reduction in the AI.

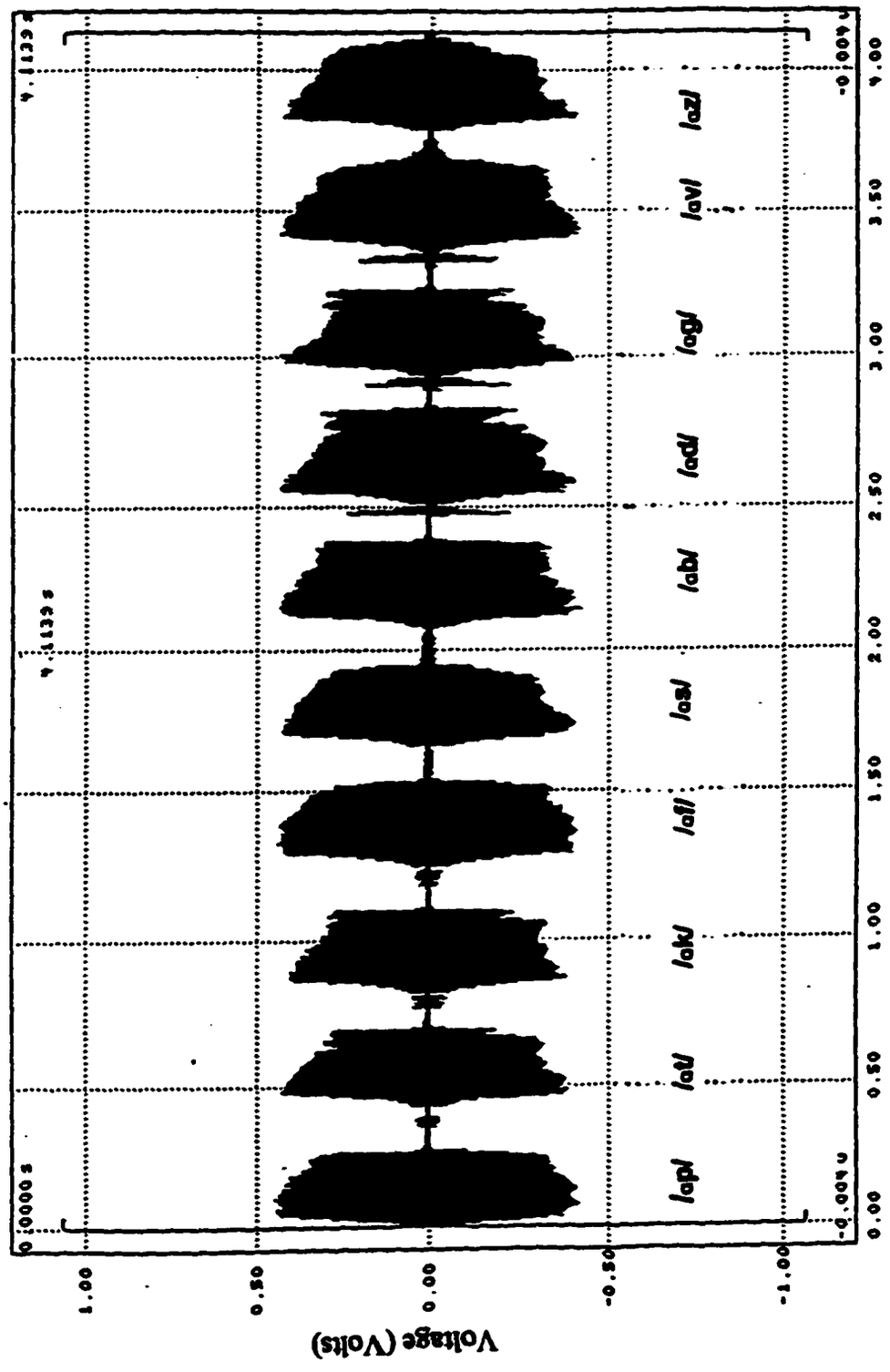
As mentioned in the discussion, the results supported the hypothesis, but the difference score between the on-harmonic and between-harmonic noises was not as large as expected. Among the possible contributors to this finding were the use of synthetic speech, the consonants selected, the choice of fundamental frequencies, and the low-pass filter cut-offs selected. Although the effect may not be as great as anticipated, additional studies are warranted to investigate the influence of each of these factors.

The results of the investigation help to explain why noise reduction strategies that remove noise from spectral regions in which speech is very weak or not present, do not produce an improvement in speech intelligibility, even when there is an improvement in the speech-to-noise ratio. This interpretation would apply to the single-microphone noise reduction methods described in Chapter 2, as well as to the sinusoidal modeling speech-enhancement technique. Of particular relevance is the comb-filtering technique developed by Frazier et al. (1976) in which the "teeth" of the comb-filter corresponded to the peaks in the speech spectrum. Although the technique resulted in an increased signal-to-noise ratio and a subjective preference in quality, intelligibility was decreased at several signal-to noise ratios. A possible explanation for this result is that the speech components and the remaining noise co-existed in the same critical band. With regard to other techniques, certainly in principle, filtering methods which reduce the amount of intense low-frequency

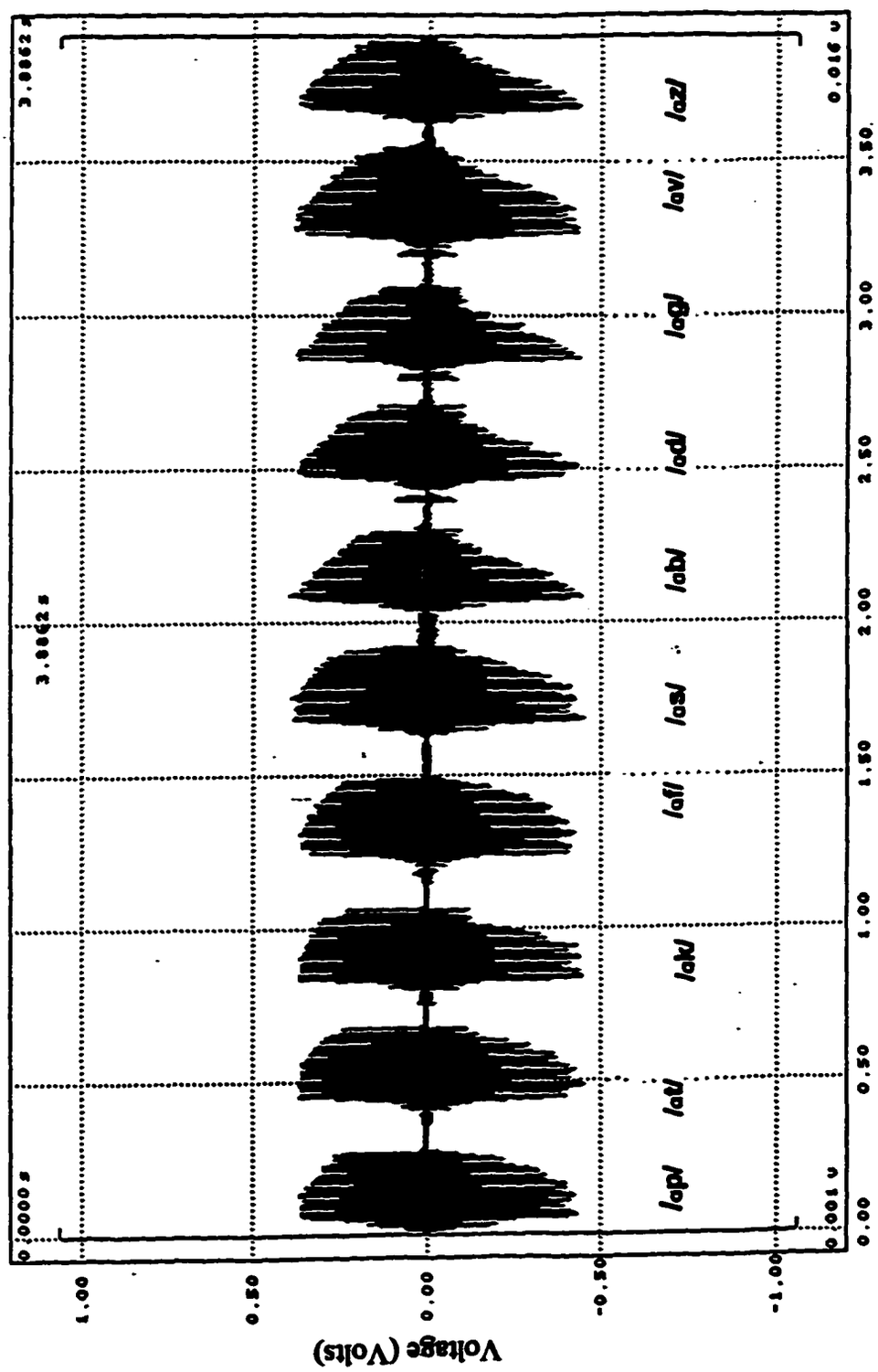
noise which produces an upward spread of masking should improve intelligibility. These methods have demonstrated small improvements in intelligibility under laboratory conditions, but not in the everyday use of hearing aids.

The results of this investigation also have implications for future noise reduction schemes, suggesting that, at high frequencies, speech recognition performance might be improved if between-harmonic noise were removed from a wide-band noise, since at these frequencies the critical bandwidths are relatively large and between-harmonic noise is likely to occur in the same critical bands as the speech harmonics. In this case, the total noise power in the critical band would be reduced, increasing the signal-to-noise ratio in the band, and increasing the AI. Similarly, improvements might also be expected for hearing-impaired individuals whose critical bands are broader than those of normal hearers. Additional research is needed to investigate these issues.

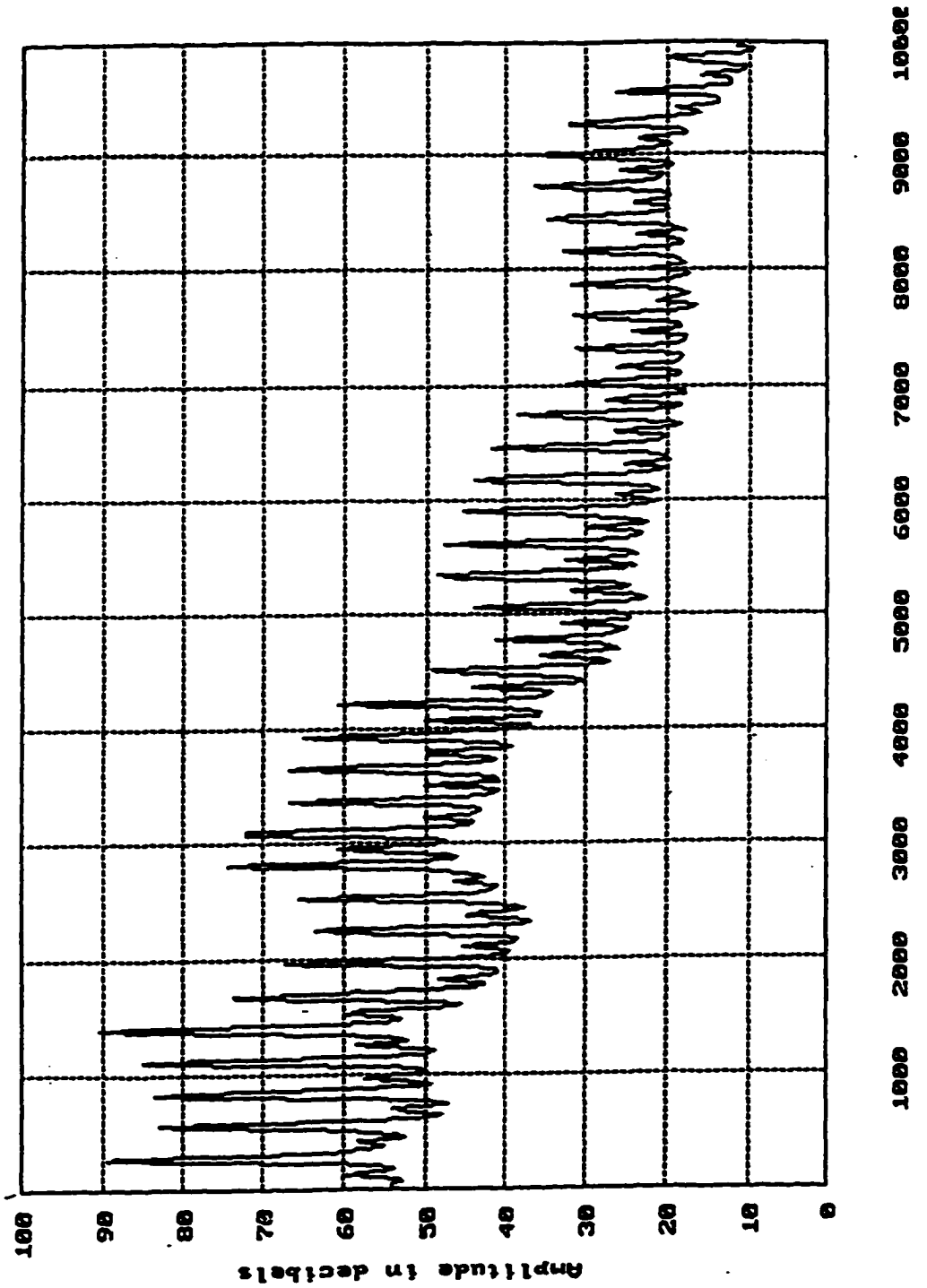
APPENDIX 1



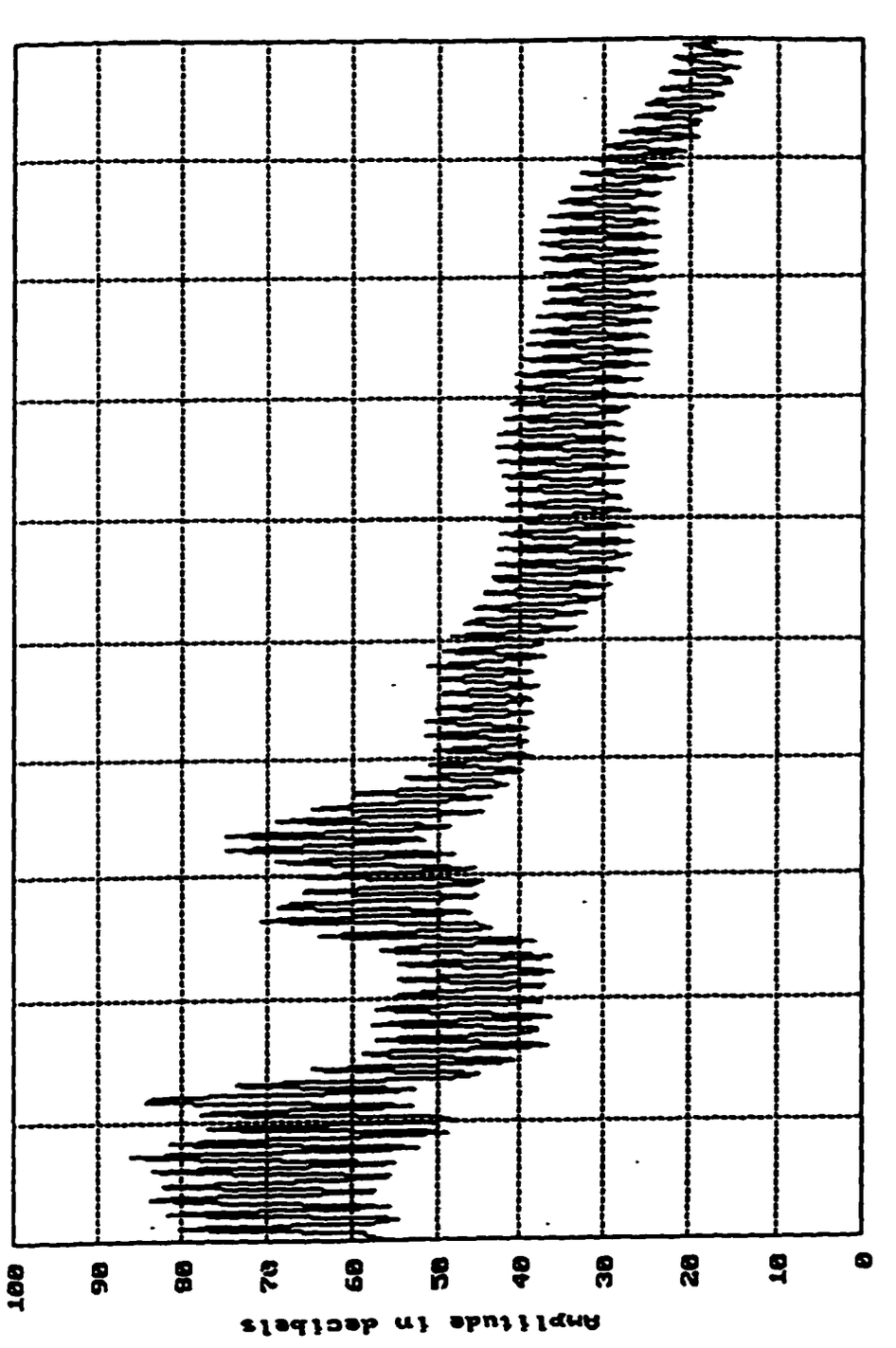
Waveform displays of female /aC/ tokens



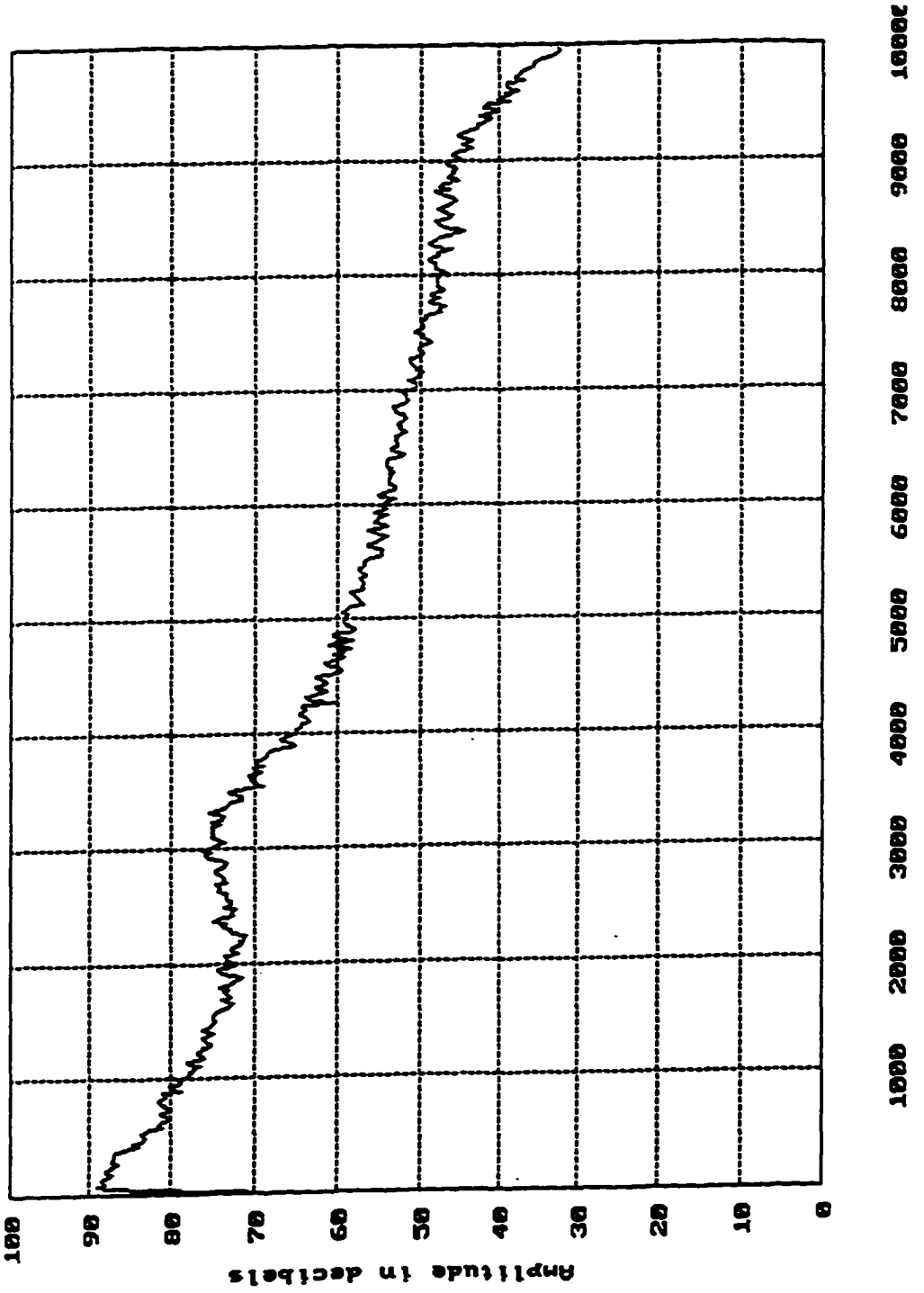
Waveform displays of male /ɑC/ tokens



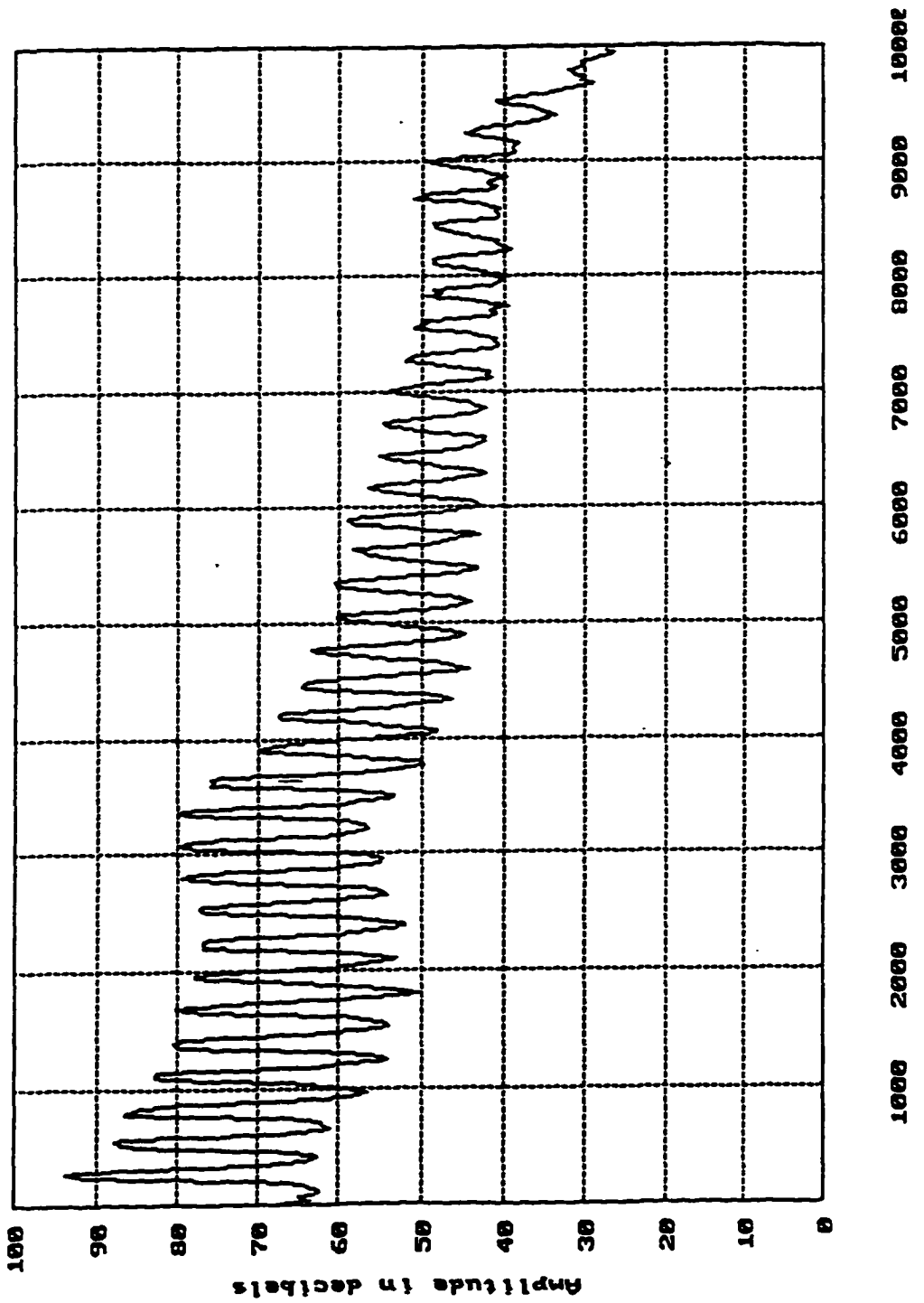
Average amplitude spectrum for female /aC/ tokens



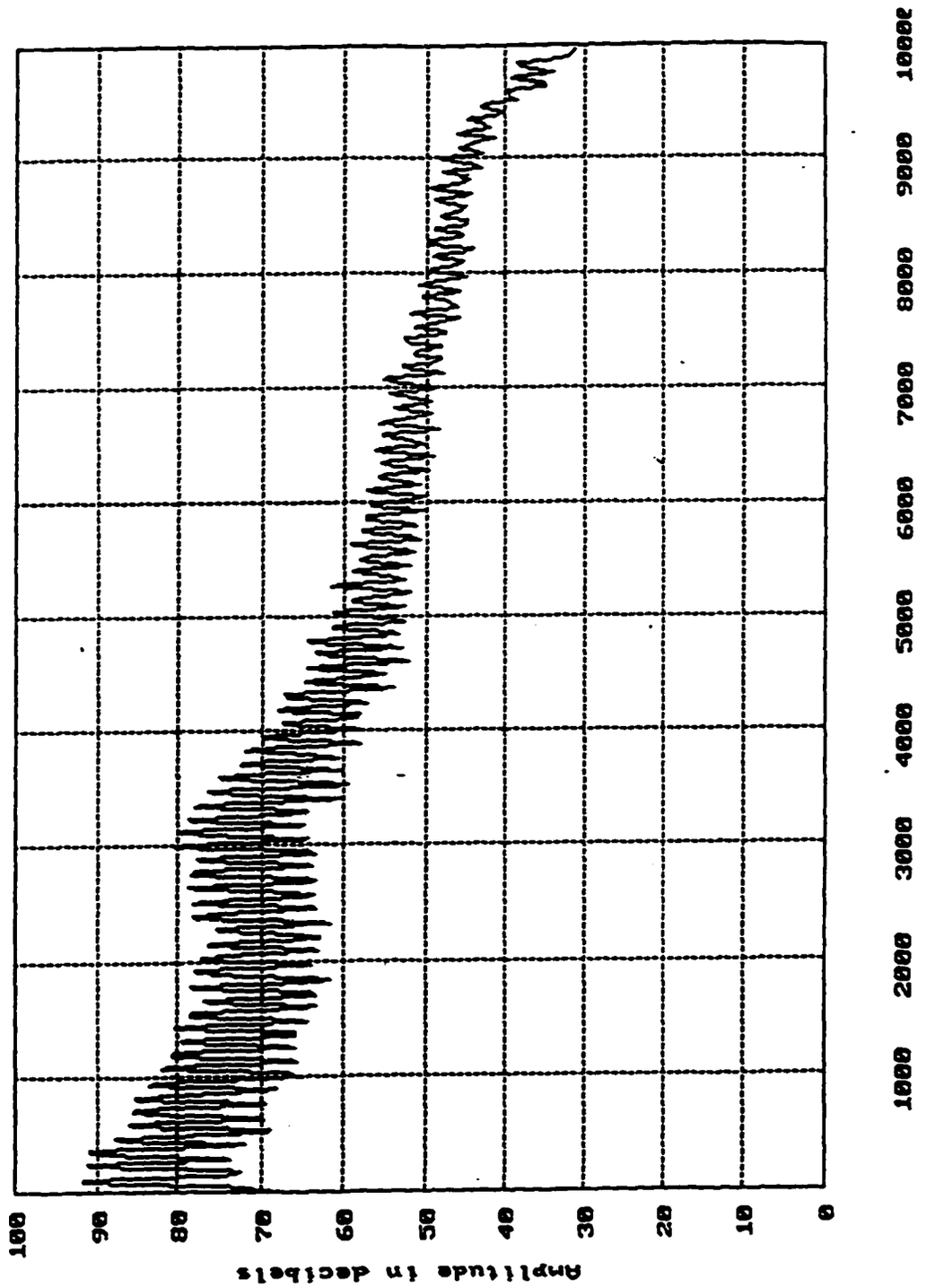
Average amplitude spectrum for male /a C/ tokens



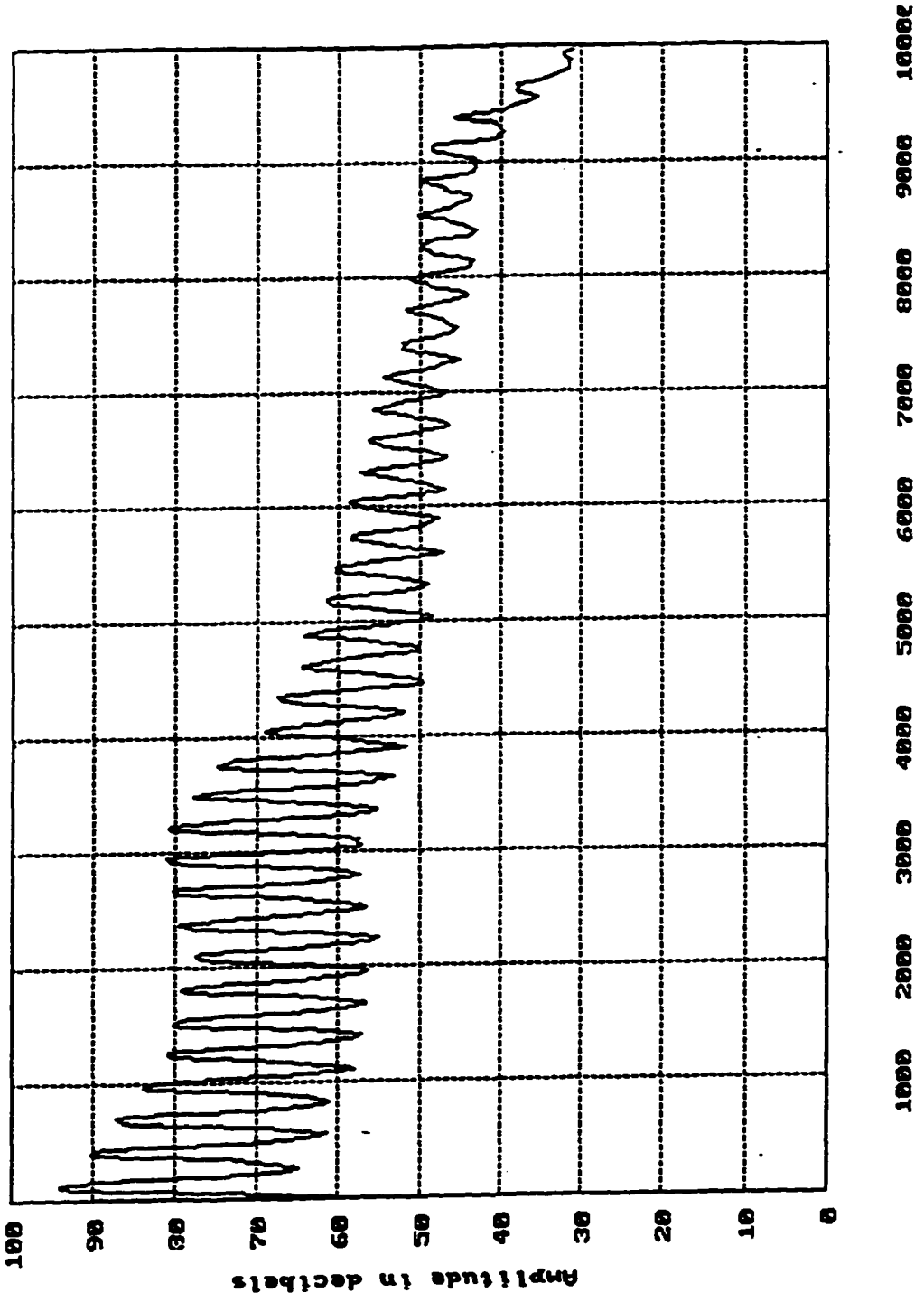
Average amplitude spectrum for speech-spectrum shaped noise



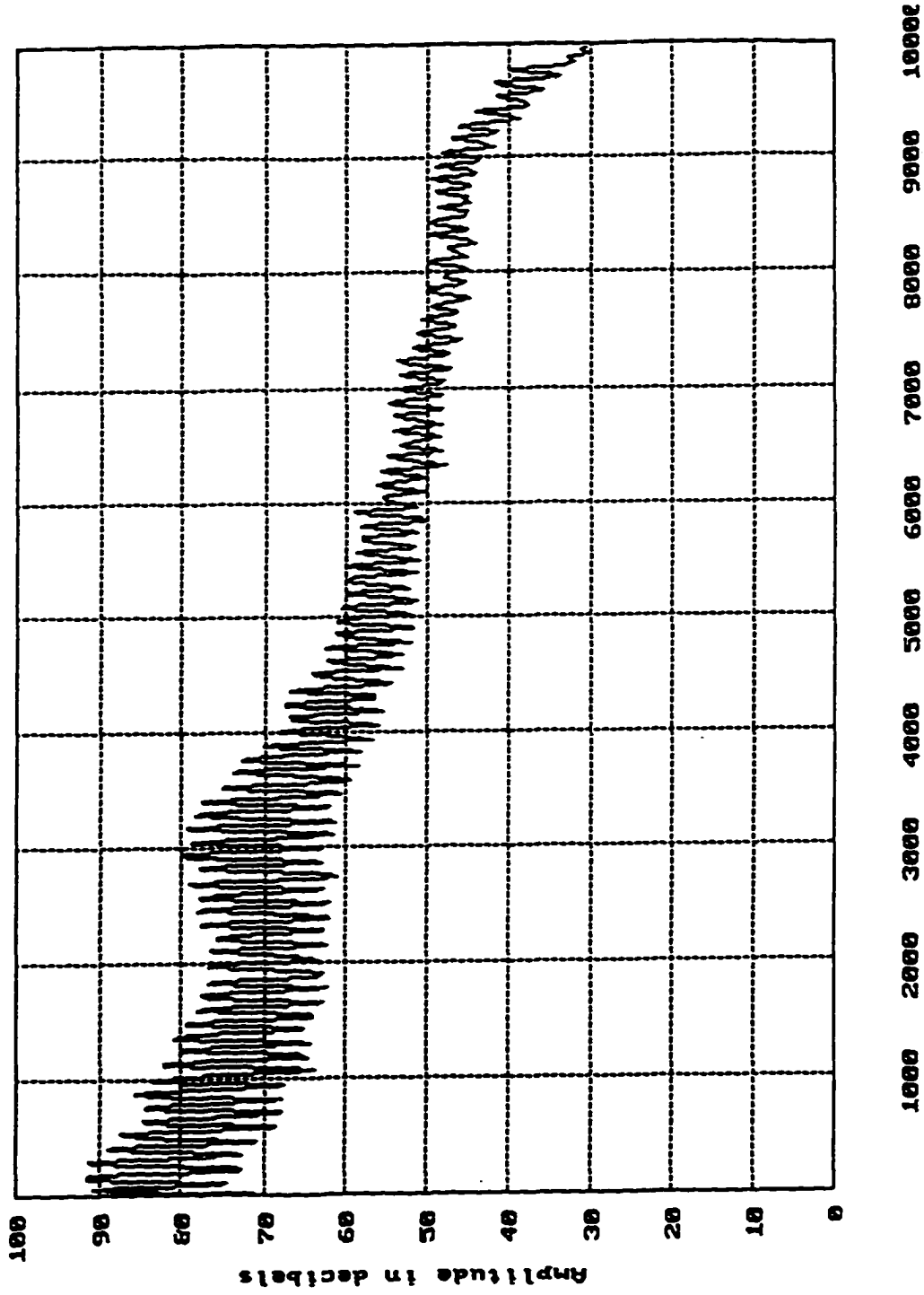
Average amplitude spectrum for female on-harmonic noise



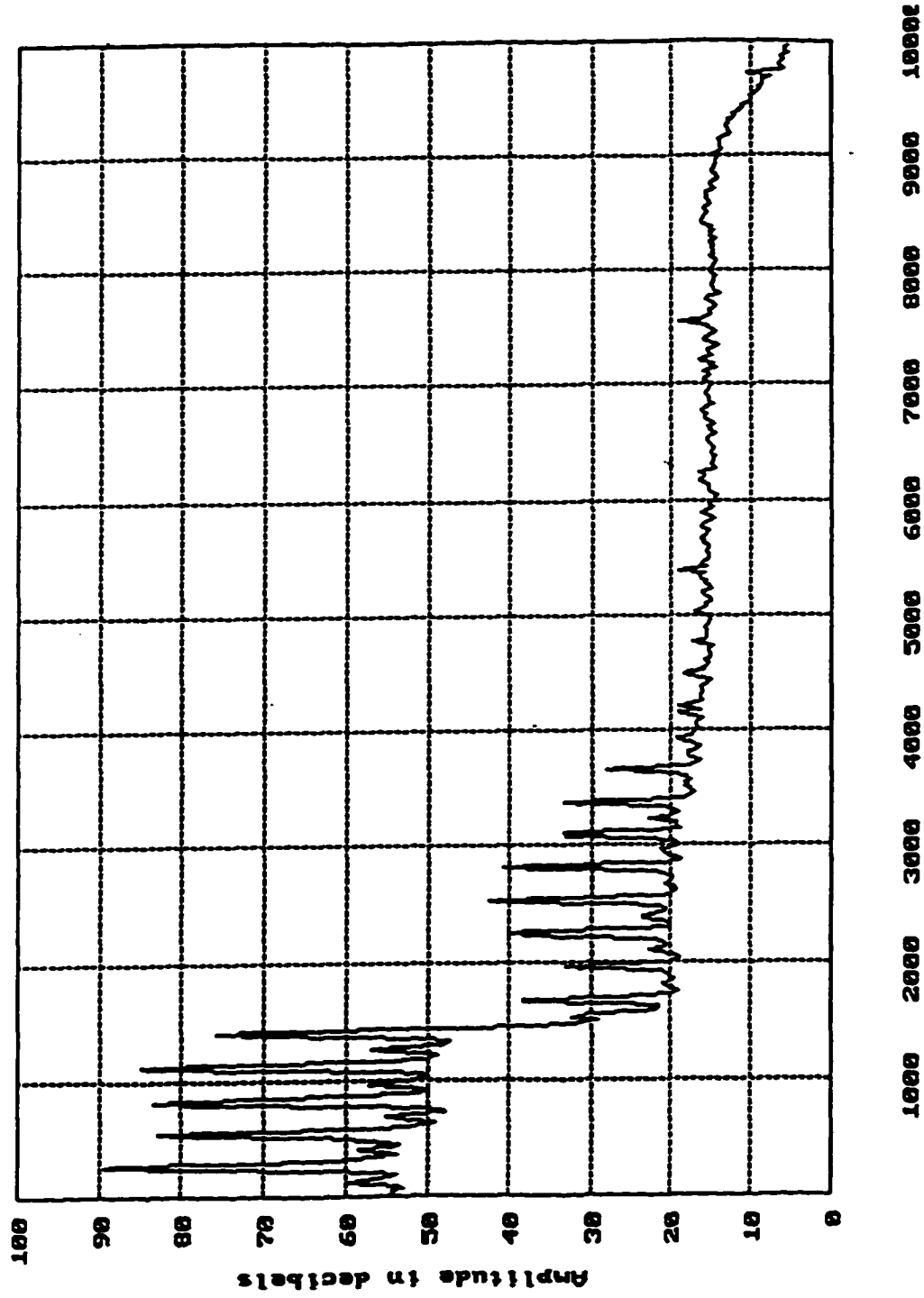
Average amplitude spectrum for male on-harmonic noise



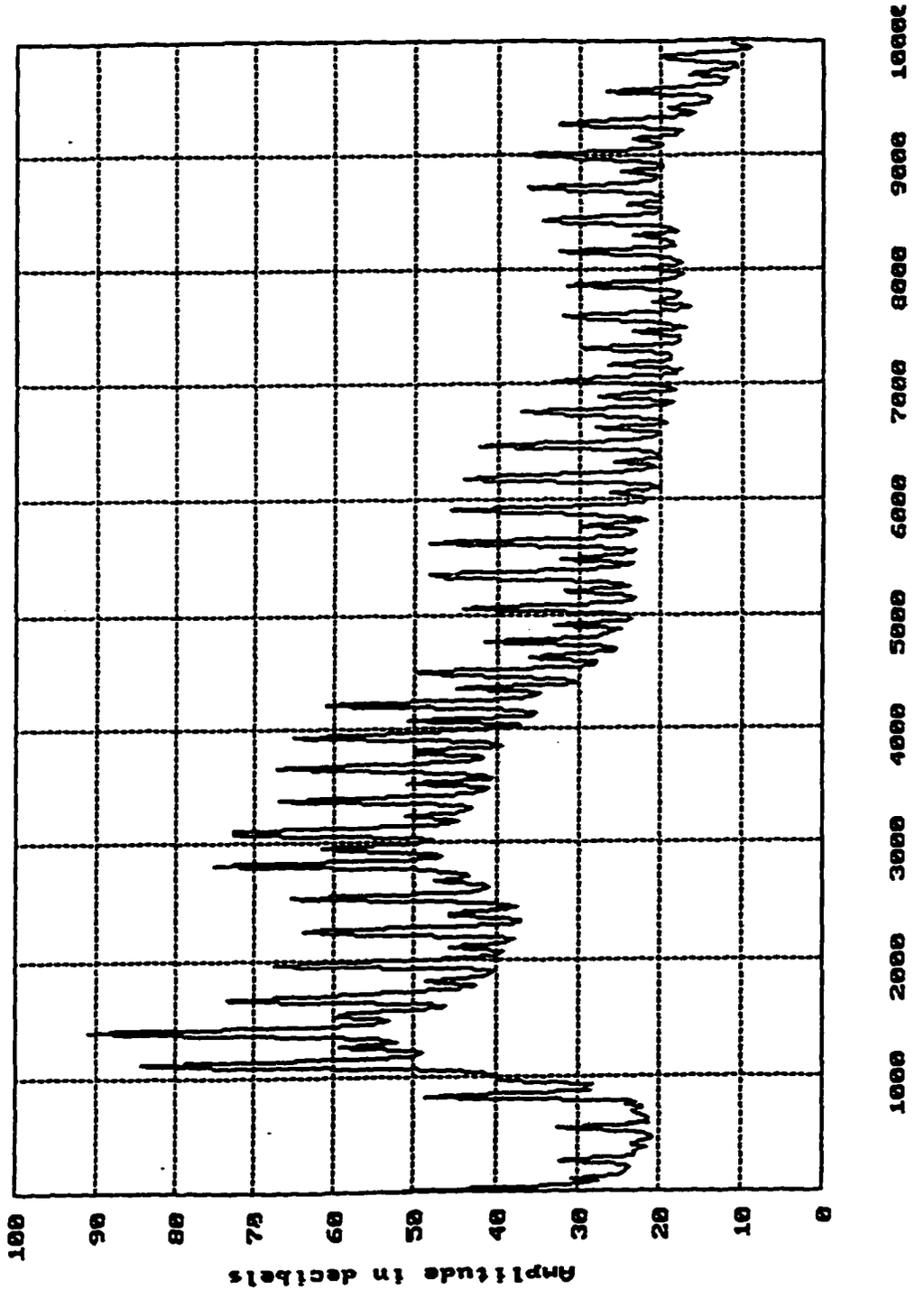
Average amplitude spectrum for female between-harmonic noise



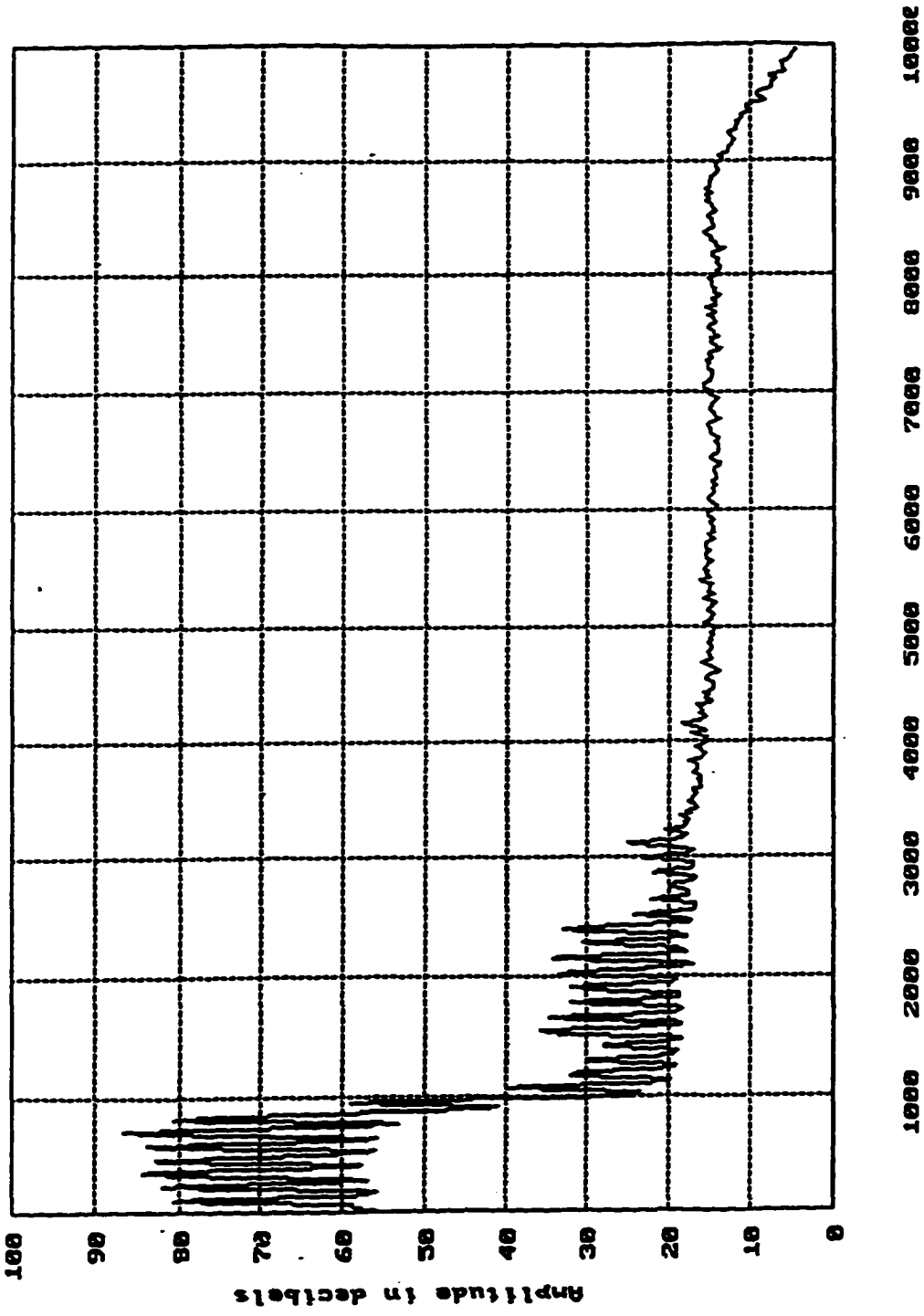
Average amplitude spectrum for male between-harmonic noise



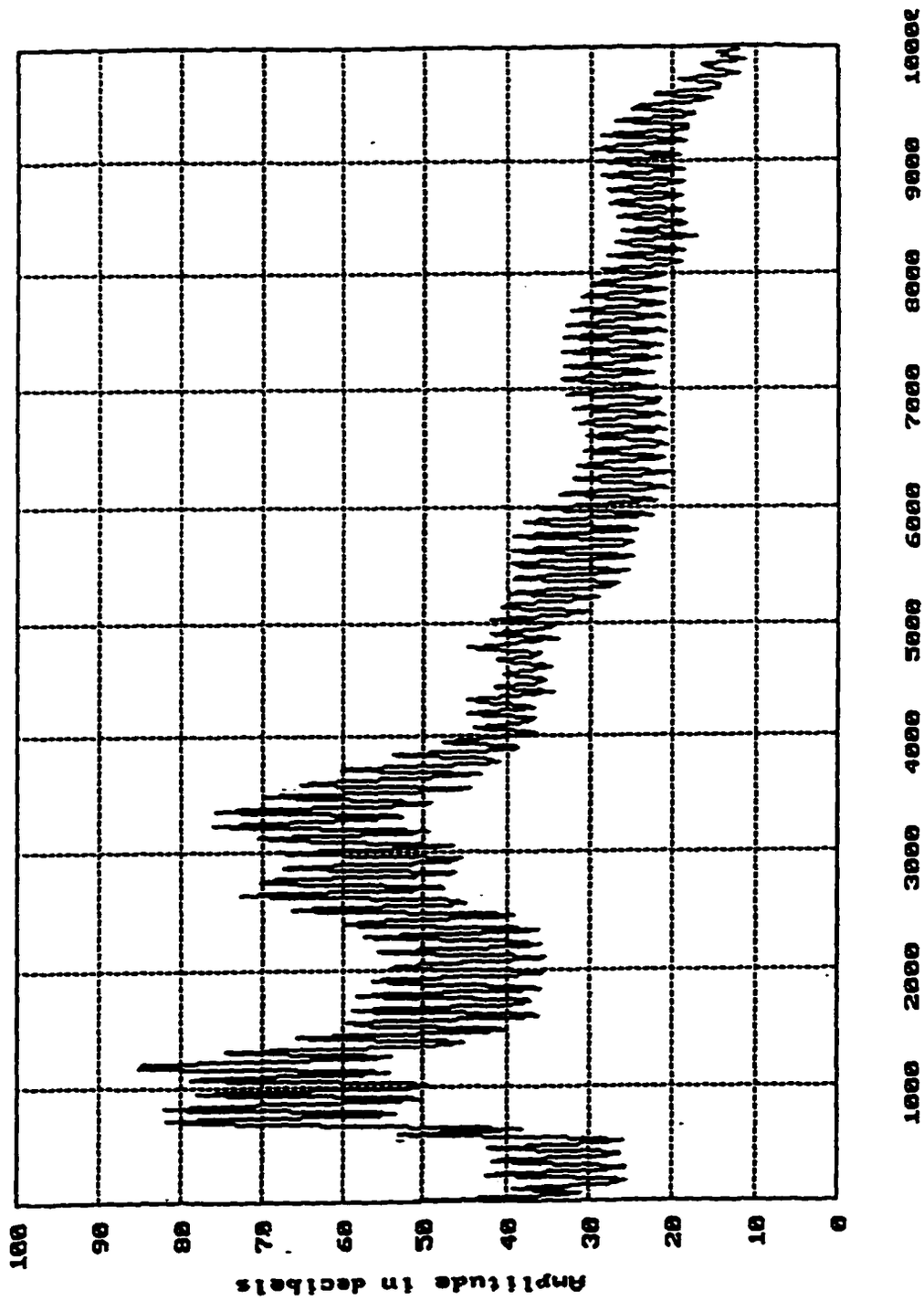
Average amplitude spectrum for low-pass filtered female /aC/ tokens



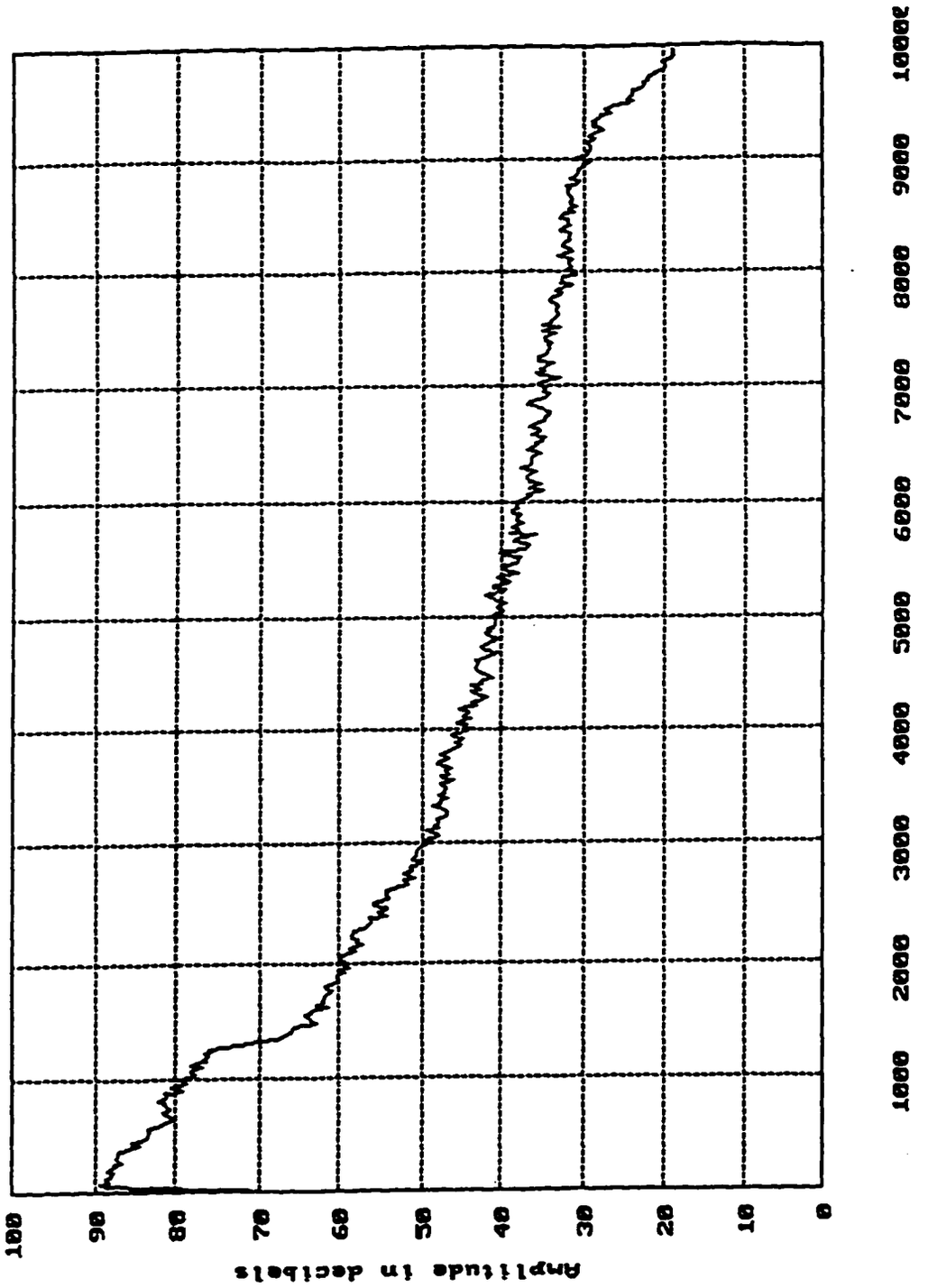
Average amplitude spectrum for high-pass filtered female /aC/ tokens



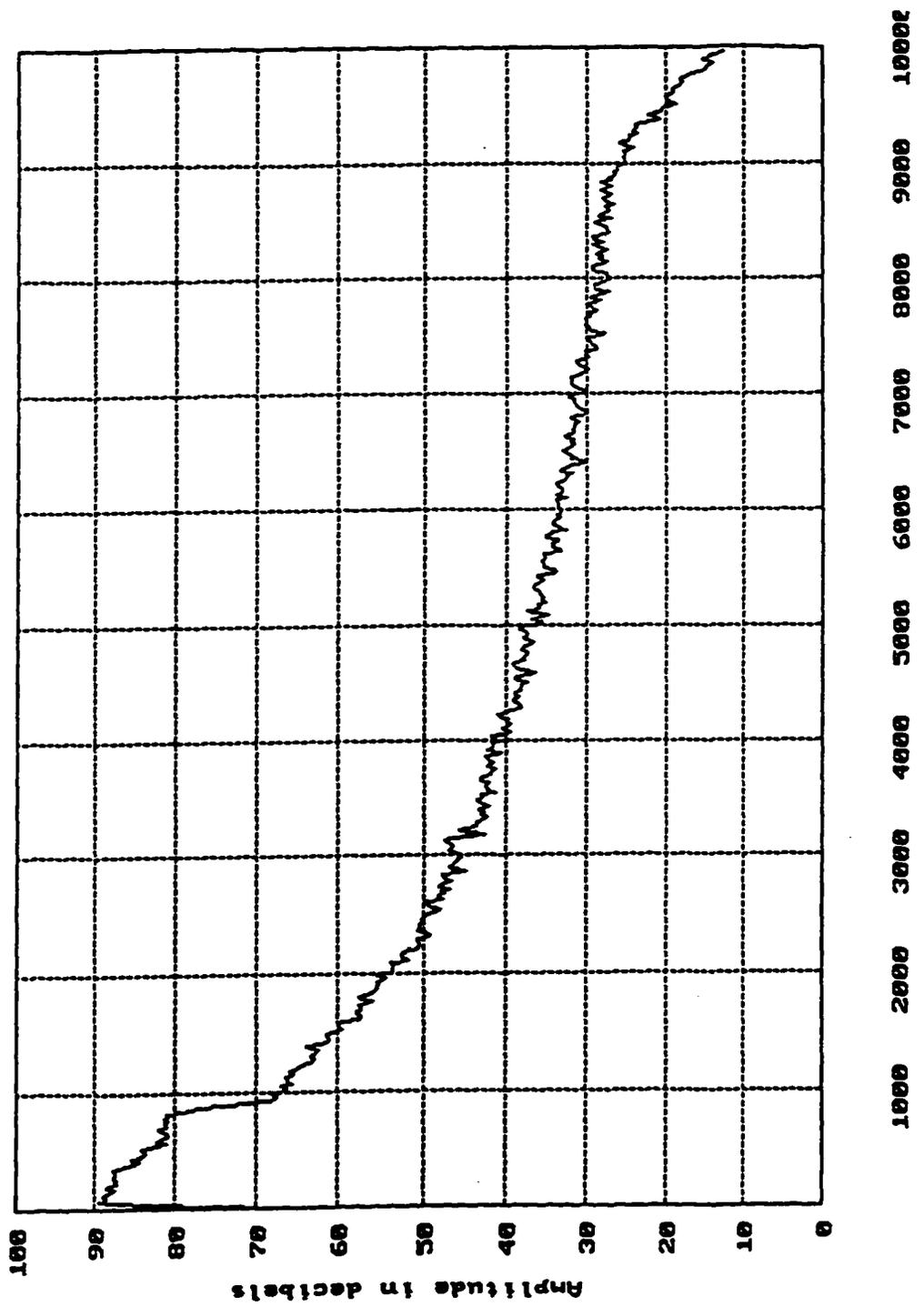
Average amplitude spectrum for low-pass filtered male /aC/ tokens



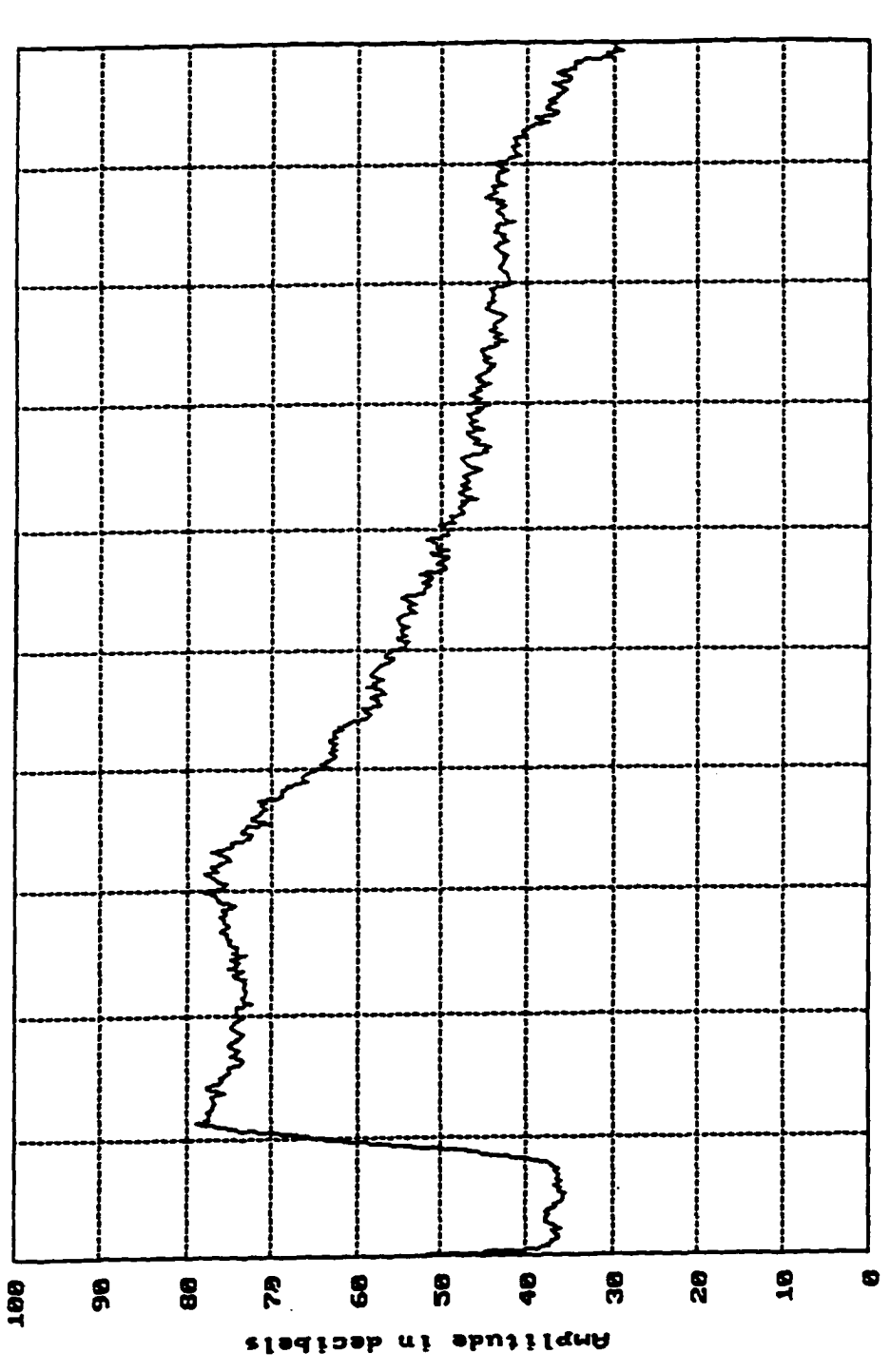
Average amplitude spectrum for high-pass filtered male /aC/ tokens



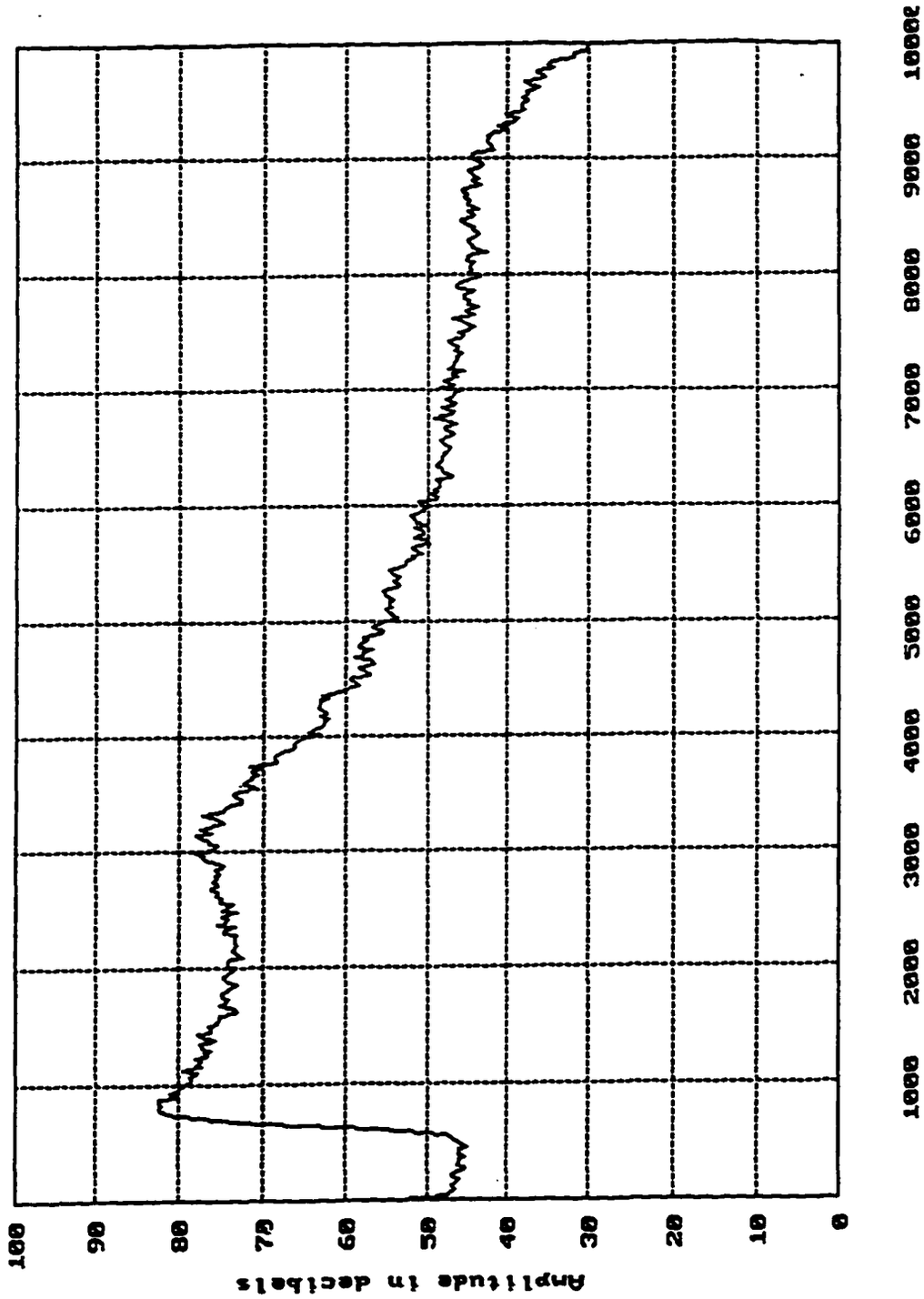
Average amplitude spectrum for low-pass filtered (1200 Hz) speech-spectrum shaped noise



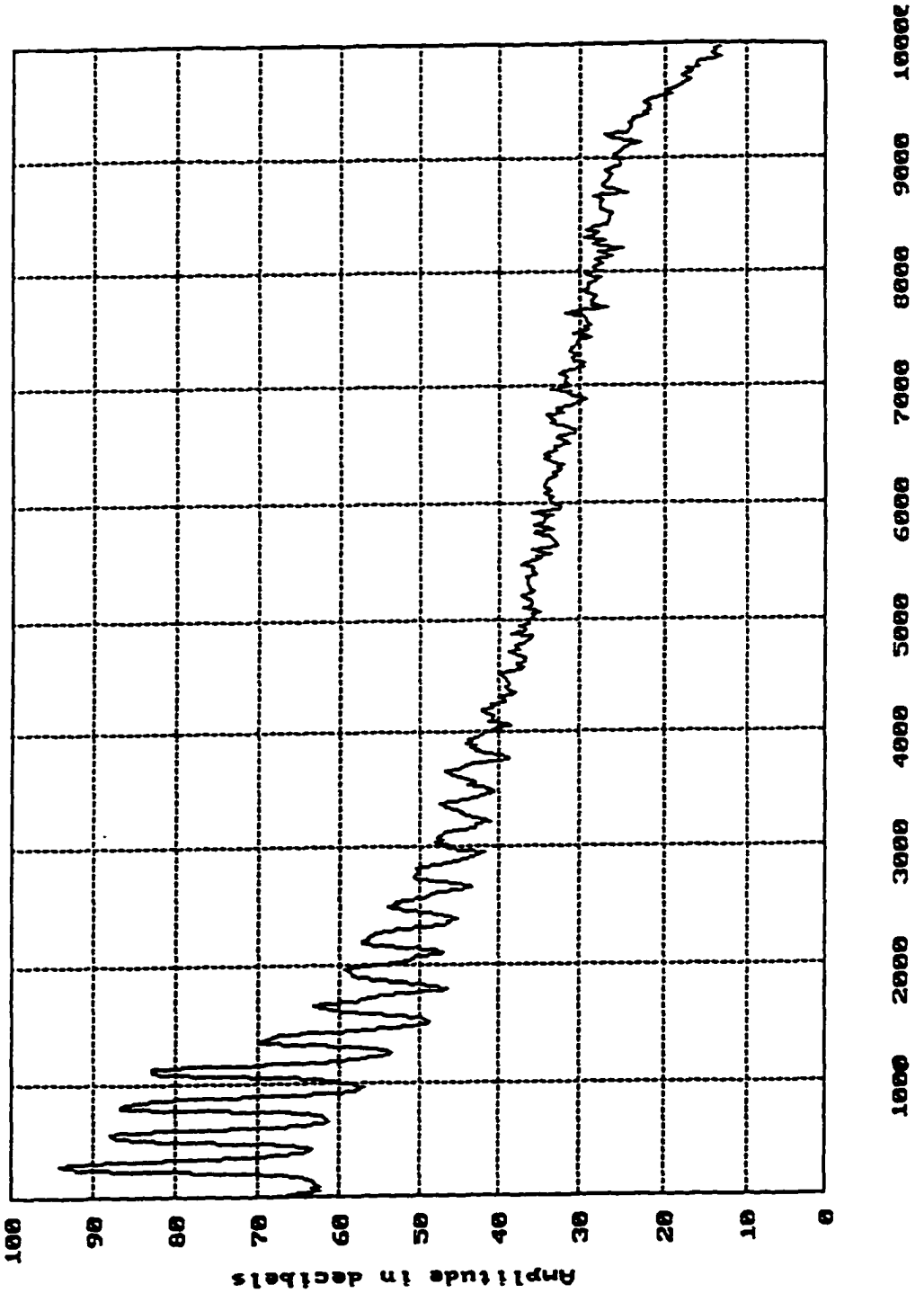
Average amplitude spectrum for low-pass filtered (800 Hz) speech-spectrum shaped noise



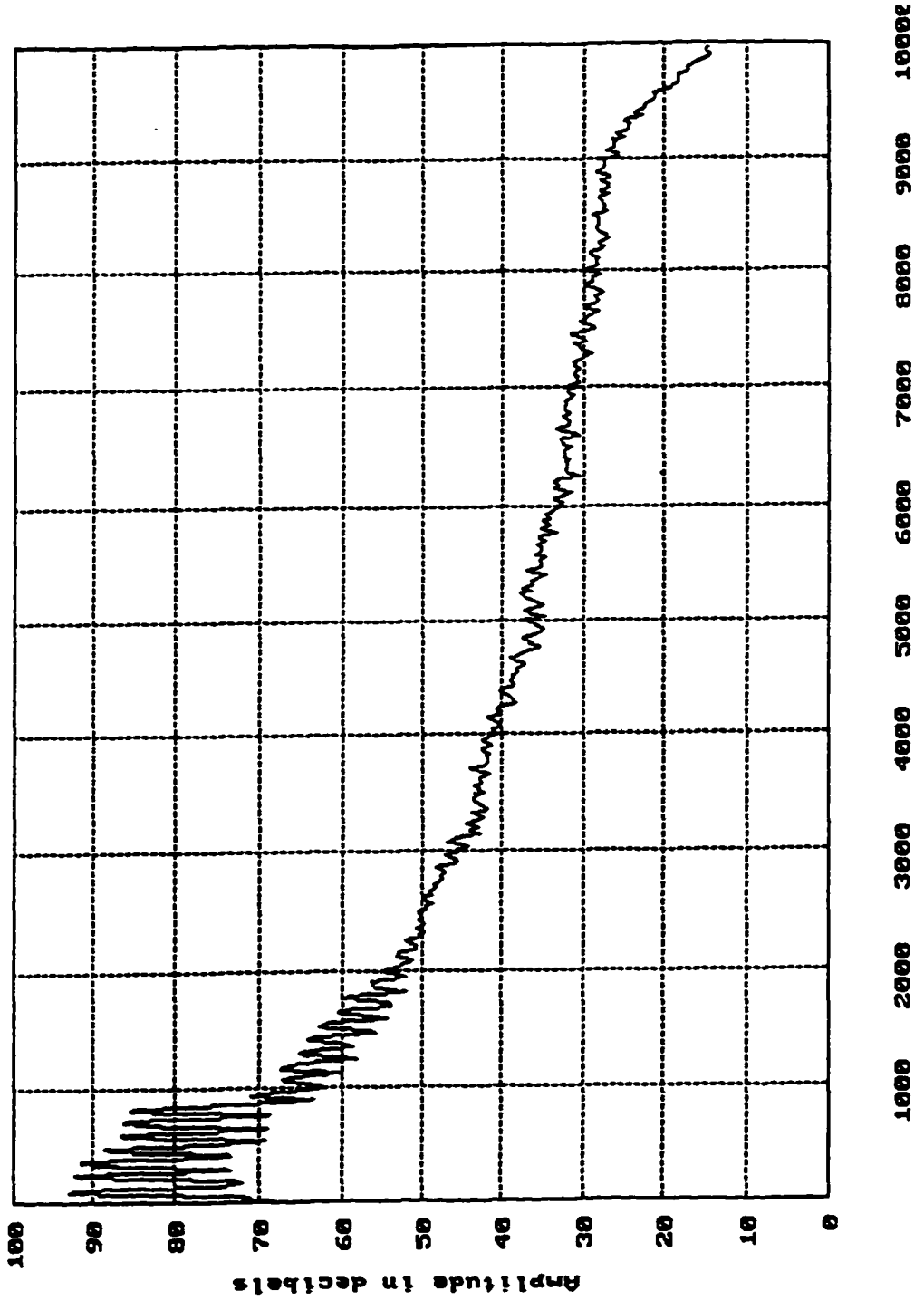
Average amplitude spectrum for high-pass filtered (1200 Hz) speech-spectrum shaped noise



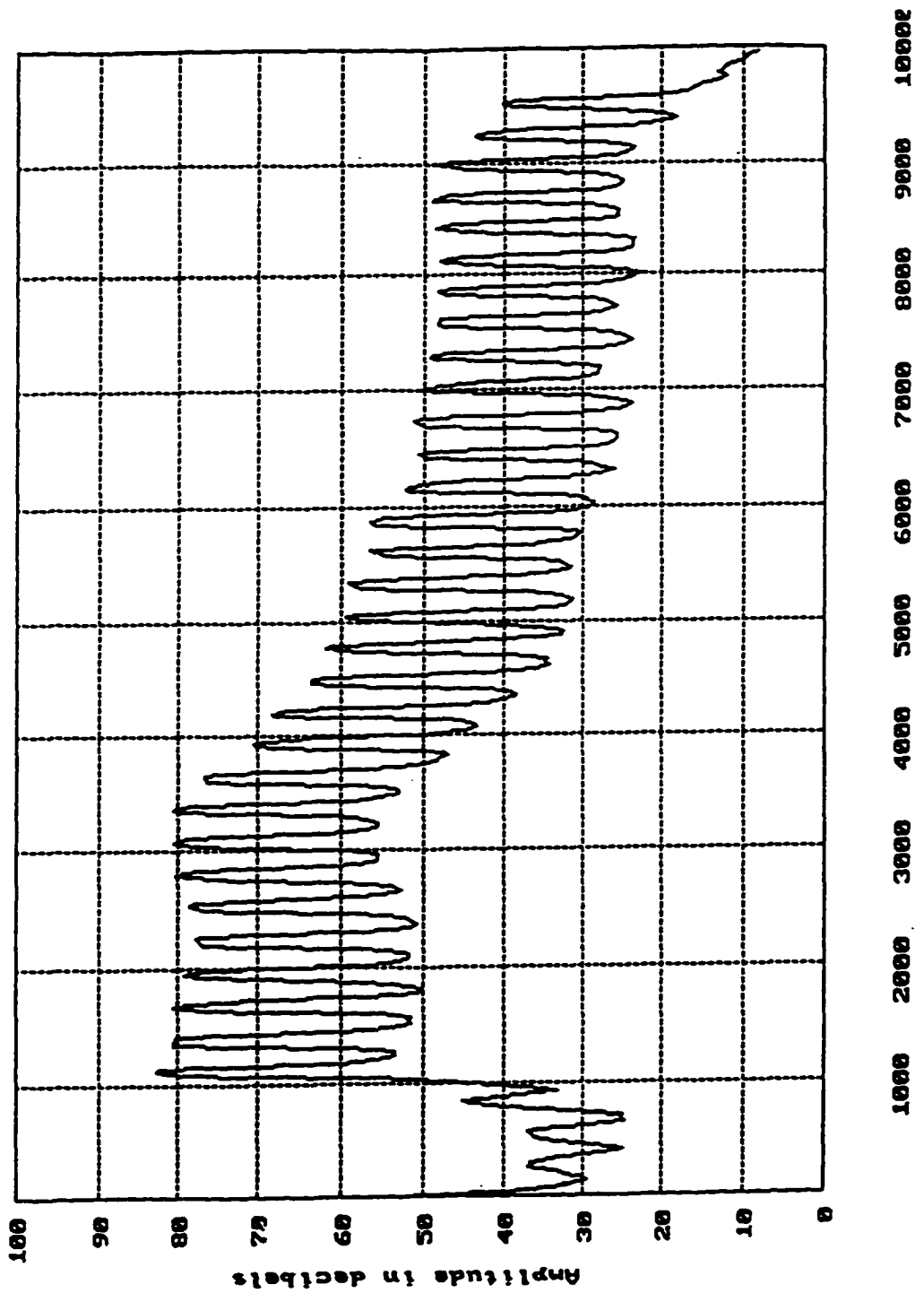
Average amplitude spectrum for high-pass filtered (800 Hz) speech-spectrum shaped noise



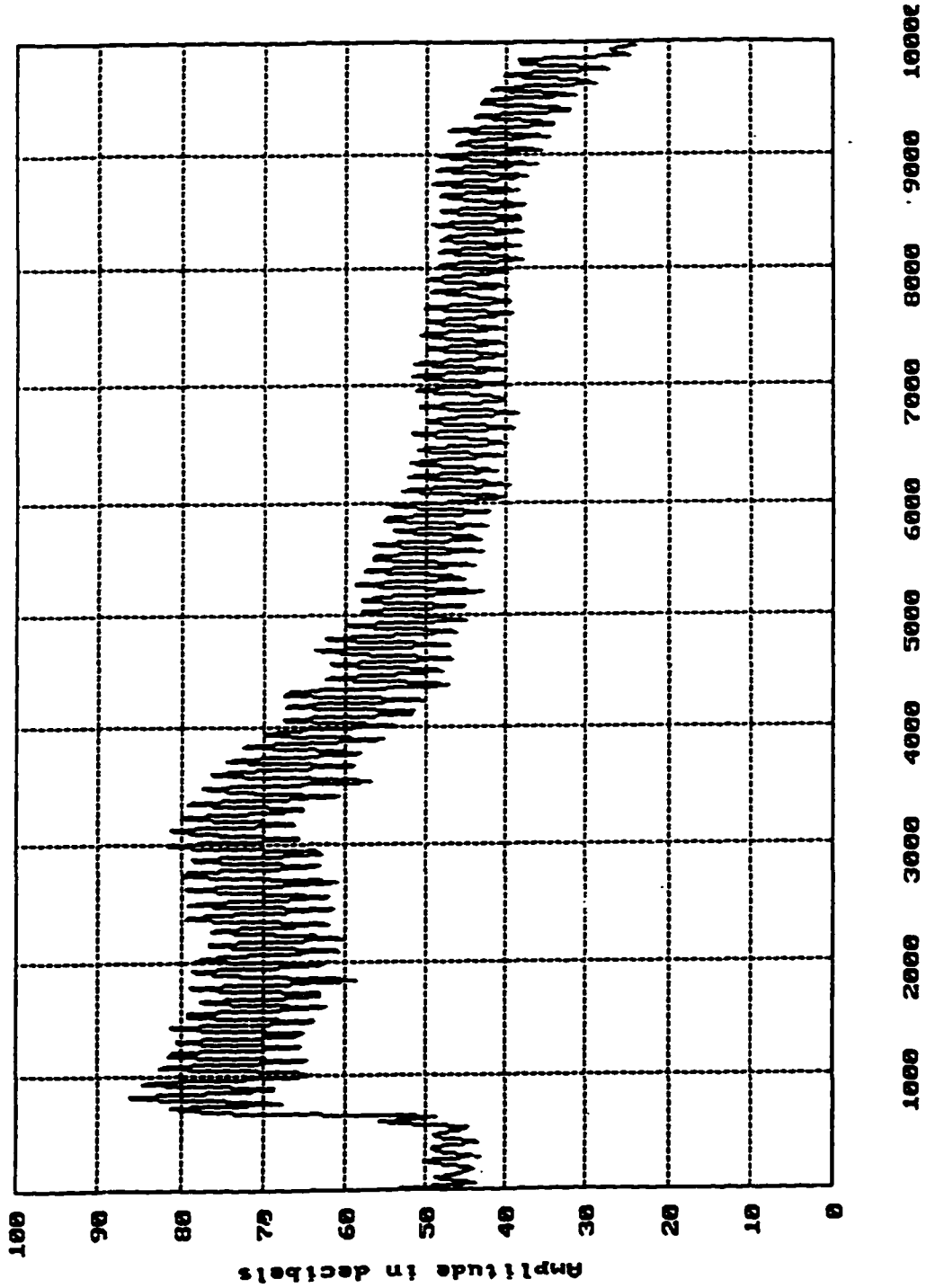
Average amplitude spectrum for low-pass filtered (1200 Hz) female on-harmonic noise



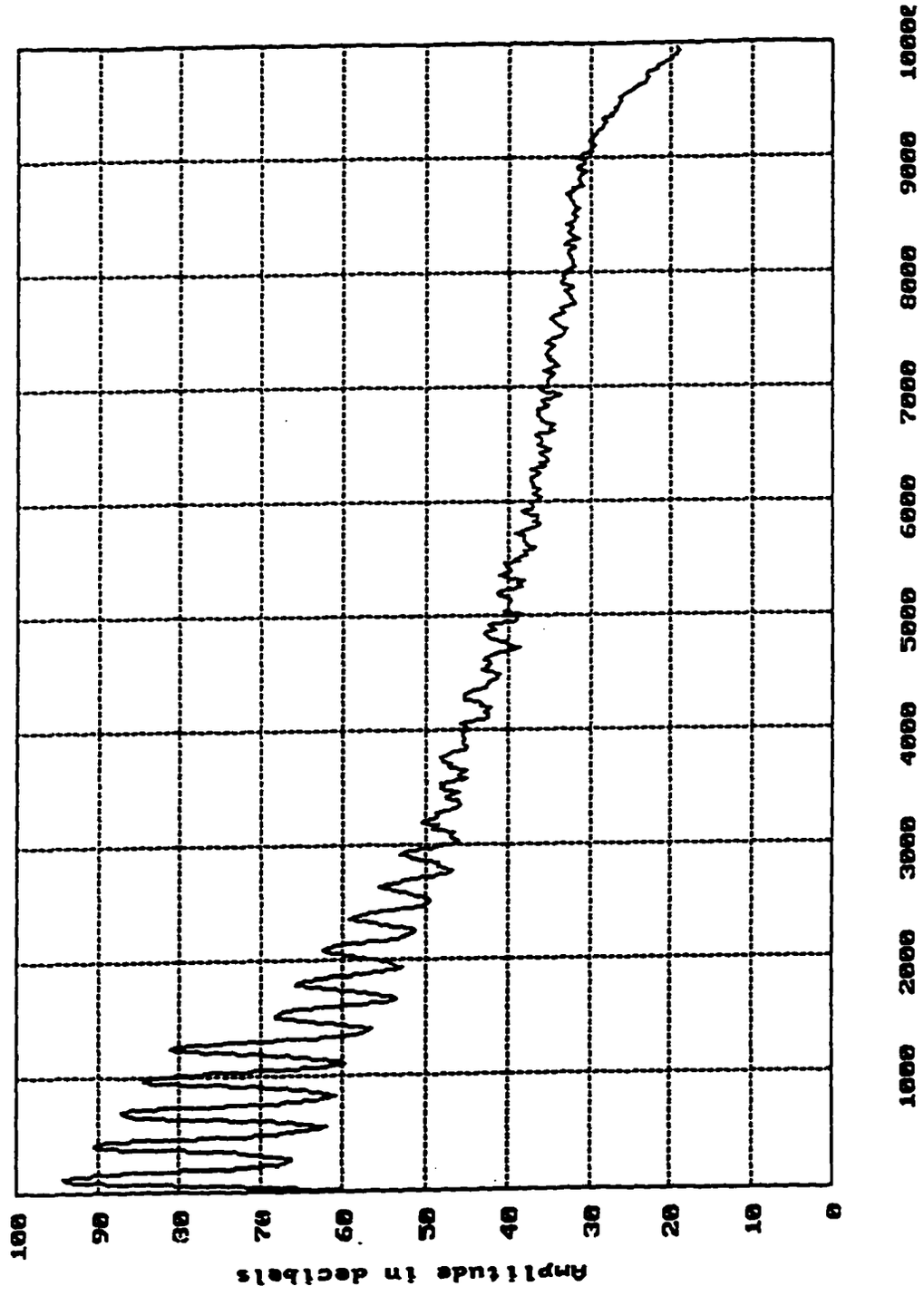
Average amplitude spectrum for low-pass filtered (800 Hz) male on-harmonic noise



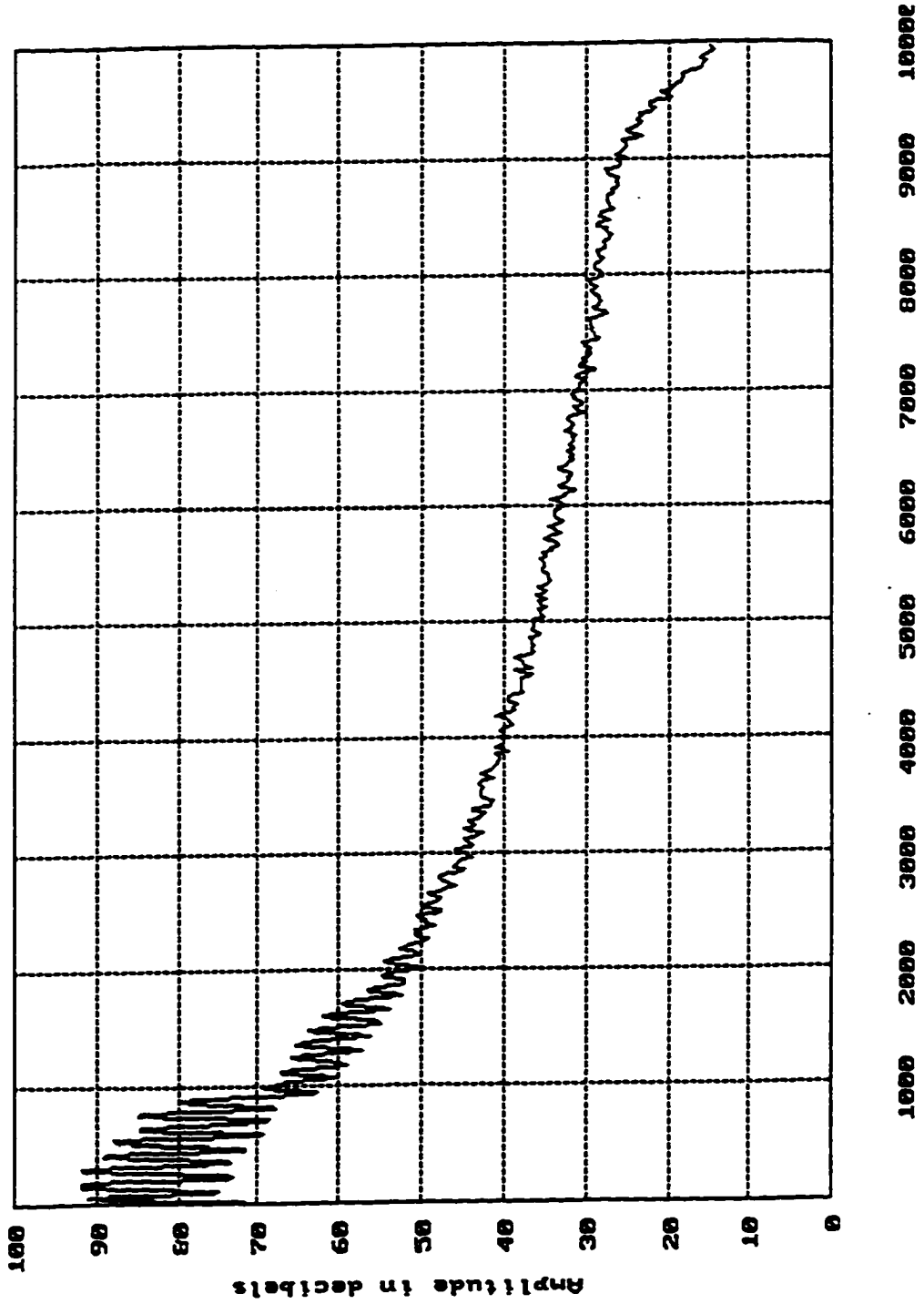
Average amplitude spectrum for high-pass filtered (1200 Hz) female on-harmonic noise



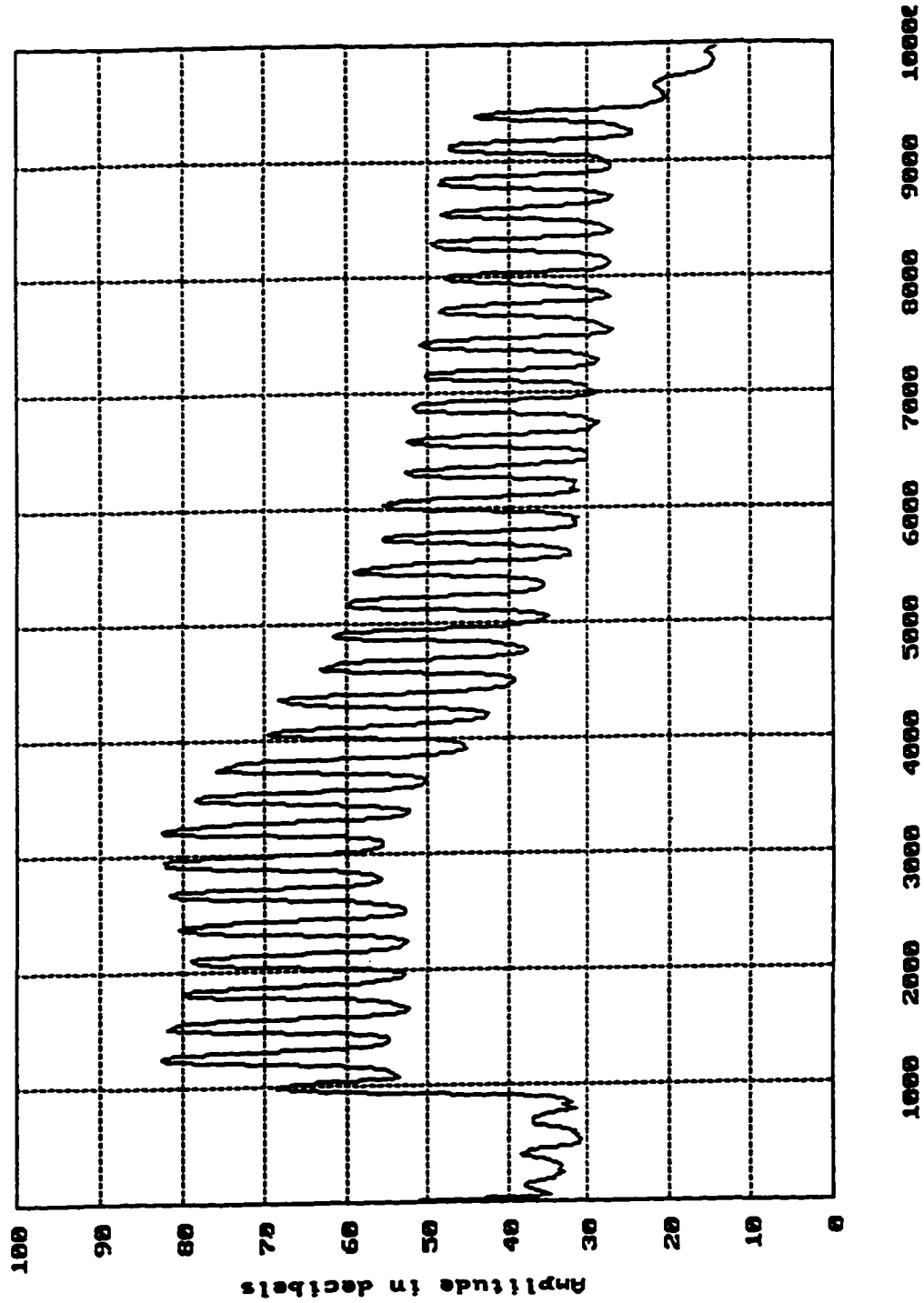
Average amplitude spectrum for high-pass filtered (800 Hz) male on-harmonic noise



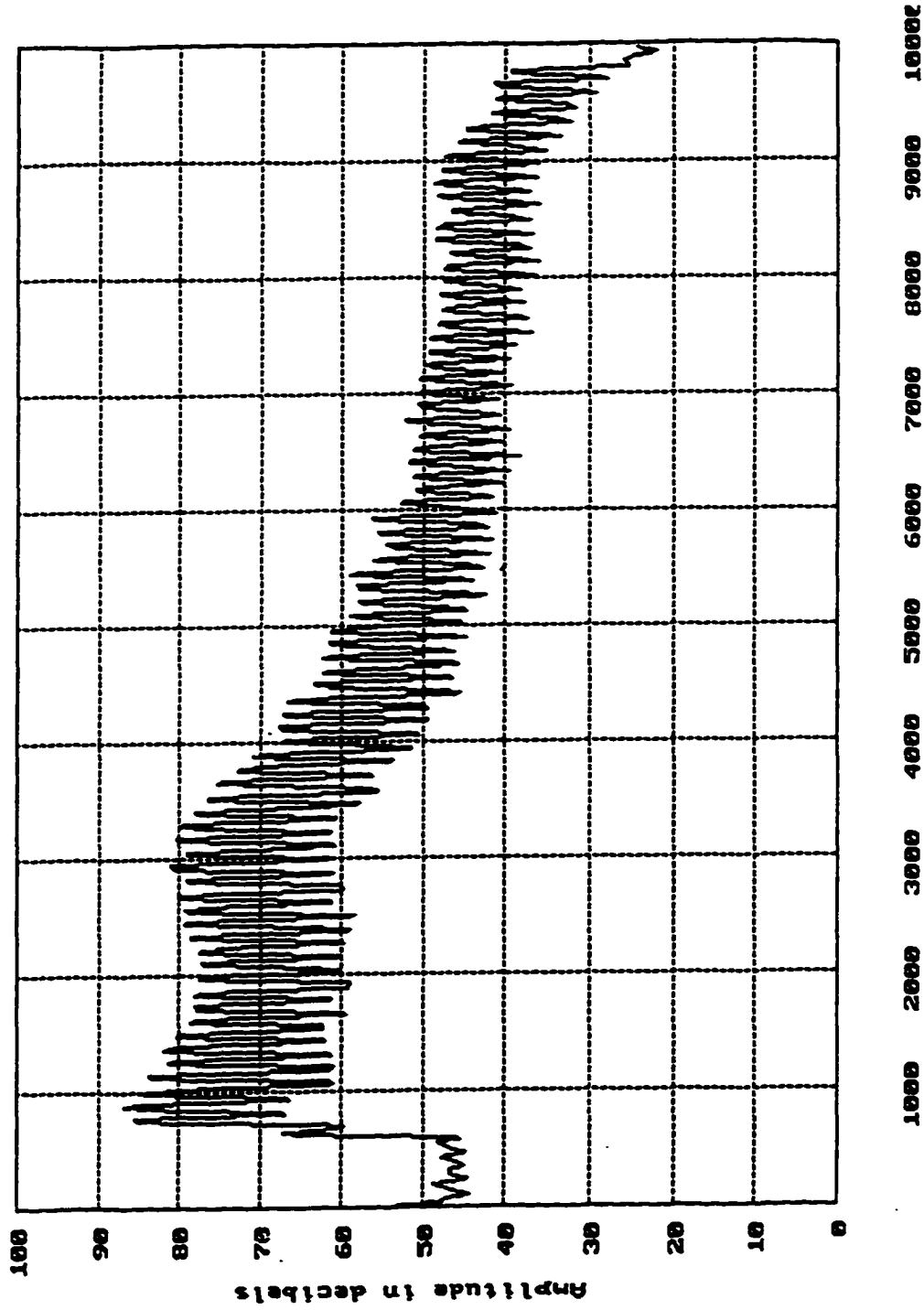
Average amplitude spectrum for low-pass filtered (1200 Hz) female between-harmonic noise



Average amplitude spectrum for low-pass filtered (800 Hz) male between-harmonic noise



Average amplitude spectrum for high-pass filtered (1200 Hz) female between-harmonic noise



Average amplitude spectrum for high-pass filtered (800 Hz) male between-harmonic noise

APPENDIX 2

SUBJECT CONSENT FORM

CONSONANT RECOGNITION IN ON-HARMONIC VERSUS BETWEEN-HARMONIC NOISE

Dissertation Investigator: Janet Reath Schoepflin, M.A.
Committee Chair: Harry Levitt, Ph.D.
Graduate School and University Center, City University of New York

The **PURPOSE** of this research is to examine consonant recognition performance in a background of noise. During the experiment you will be seated in a sound-treated room, listening through headphones to ten vowel-consonant syllables presented in different noise conditions. The syllables are /ap/, /at/, /ak/, /af/, /as/, /ab/, /ad/, /ag/, /av/, and /az/.

After a brief practice period during which you will listen to the syllables presented in quiet (no background noise), you will begin the experiment. The **PROCEDURES** to be followed will include random presentation of the syllables under several different noise conditions. After each syllable is presented, you will select the one you heard by pressing a correspondingly labeled key on a computer keyboard. Immediately following your selection, the next syllable will be presented and you will again make your selection. You may listen to any syllable again by pressing the space bar and the "return" key on the keyboard. Each experimental session involves 1800 presentations. You are free to complete a session in one or more sittings, taking as many breaks as you wish. The complete experiment will involve 5 experimental sessions, requiring a total of approximately 8 to 10 hours. The level (loudness) of the stimuli can be adjusted by you to one of comfort.

There is no known **RISK** associated with the study, and the conditions under which the research is being conducted are similar to those used in routine clinical audiological evaluations (hearing tests). The **BENEFITS** which may be expected from the research include a better understanding of how noise interferes with speech recognition, and possible improvements in the design of hearing aids and other auditory devices.

The results of the study will be analyzed by the researchers and may be published in professional journals and/or reported at professional meetings. Individual responses will be known only to the research team and all records will be kept confidential. In reporting the results of the study, subject anonymity will be maintained.

Consonant Recognition in On-Harmonic versus Between-Harmonic Noise
Consent Form - Page 2

Participants in the study may **WITHDRAW** or refuse to continue in the study at any time. Participants will be compensated at the rate of $8.00 for each hour of test time. In addition, subway or bus fare for travelling to and from test sessions will be reimbursed. Payment will be received through the mail after the final test session. Should you withdraw from the study, you will be paid for all hours that you have completed.

To date, any and all questions raised by me have been answered satisfactorily. If any further questions arise regarding the study or my participation as an experimental subject, I understand that I may contact Janet Schoepflin at (212) 481-4464 or Professor Harry Levitt at (212) 624-2359. I further understand that if I have questions concerning my rights as a participant in this study, I may call Sponsored Research, Graduate School and University Center/CUNY at (212) 642-2059.

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

NAME: _____

SIGNATURE: _____ DATE: _____

WITNESS: _____ DATE: _____

INSTRUCTIONS TO SUBJECTS

CONSONANT RECOGNITION IN ON-HARMONIC VERSUS BETWEEN-HARMONIC NOISE

Dissertation Investigator: Janet Reath Schoepflin, M.A.
Committee Chair: Harry Levitt, Ph.D.
Graduate School and University Center, City University of New York

You will be listening through headphones to 10 different vowel-consonant syllables in a background of noise. The syllables are /ap/, /at/, /ak/, /af/, /as/, /ab/, /ad/, /ag/, /av/, and /az/. After each syllable is presented, you will select the one you heard by pressing a correspondingly labeled key on a computer keyboard which will be shown to you by the examiner. Immediately following your selection, the next syllable will be presented and you will again make your selection. You may listen to any syllable again by pressing the space bar and the "return" key on the keyboard. Each experimental session involves 1800 presentations. You are free to complete a session in one or more sittings, taking as many breaks or rest periods as you wish. The complete experiment will involve 5 experimental sessions. The level of the stimuli can be adjusted by you to whatever level you feel is comfortable.

As you have previously been informed, you may ask the examiner any questions concerning the study or your rights as a participant at any time during the experiment. You may also withdraw from the study at any time without penalty.

APPENDIX 3

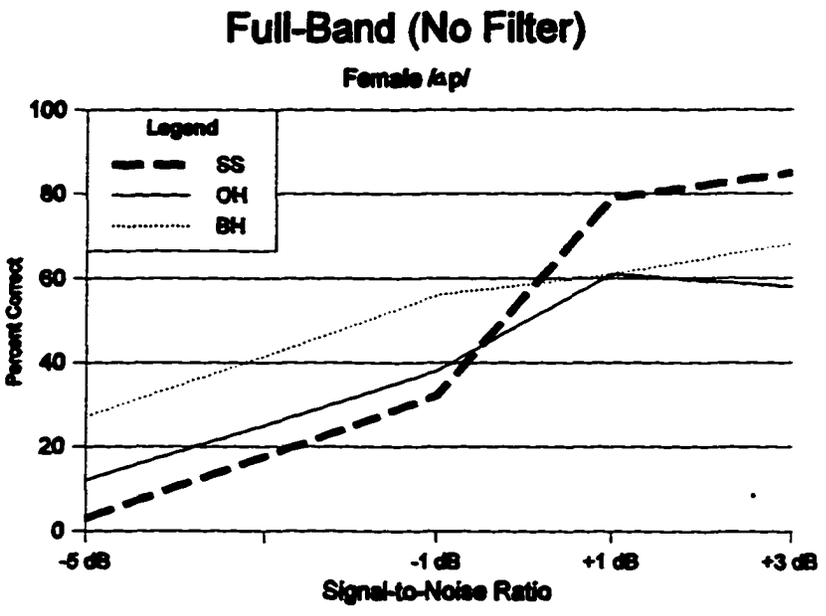
ANOVA Results for 3 Filter Conditions for the Factors Noise-type (N), Gender (G), Consonant (C), and Signal-to-Noise Ratio (S/N)

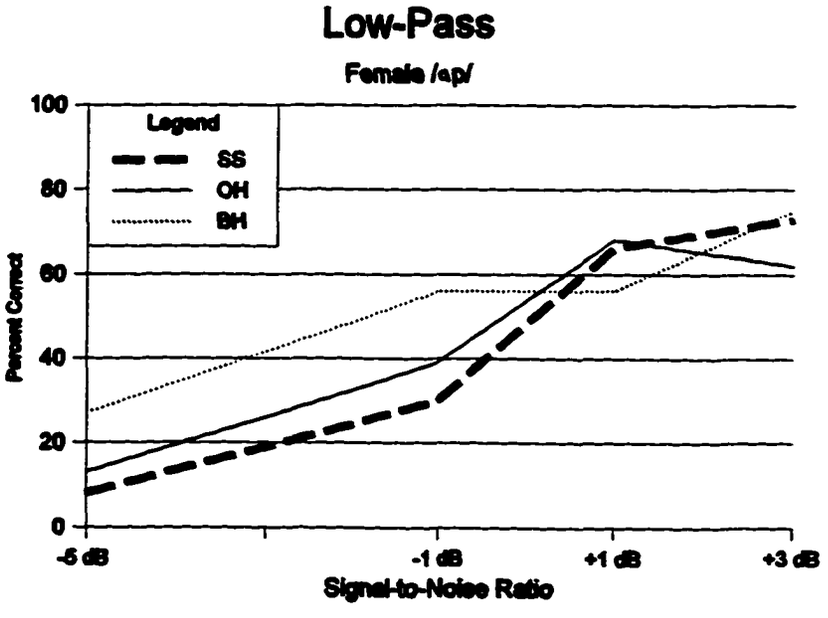
		Full-Band (No Filter)		Low-Pass Filter		High-Pass Filter	
SOURCE	**df**	**F**	**p**	**F**	**p**	**F**	**p**
Noise (N)	2, 18	38.475	<0.001	38.683	<0.001	11.299	<0.001
Gender (G)	1, 9	0.651	0.554	1.541	0.245	0.010	0.919
N x G	2, 18	2.553	0.104	16.282	<0.001	0.994	0.609
Consonant (C)	9, 81	9.424	<0.001	13.765	<0.001	6.242	<0.001
N x C	18, 162	9.569	<0.001	6.907	<0.001	6.117	<0.001
G x C	9, 81	13.101	<0.001	18.006	<0.001	15.257	<0.001
N x G x C	18, 162	3.881	<0.001	3.722	<0.001	3.113	<0.001
S/N	3, 27	108.715	<0.001	34.290	<0.001	69.718	<0.001
N x S/N	6, 54	11.296	<0.001	3.704	0.004	1.815	0.113
G x S/N	3, 27	16.877	<0.001	1.295	0.296	0.977	0.581
N x G x S/N	6, 54	1.957	0.088	7.918	<0.001	1.617	0.160
C x S/N	27, 243	7.531	<0.001	6.799	<0.001	5.047	<0.001
N x C x S/N	54, 486	4.532	<0.001	2.363	<0.001	2.117	<0.001
G x C x S/N	27, 243	6.166	<0.001	4.483	<0.001	6.653	<0.001
N x G x C x S/N	54, 486	3.805	<0.001	2.643	<0.001	2.379	<0.001

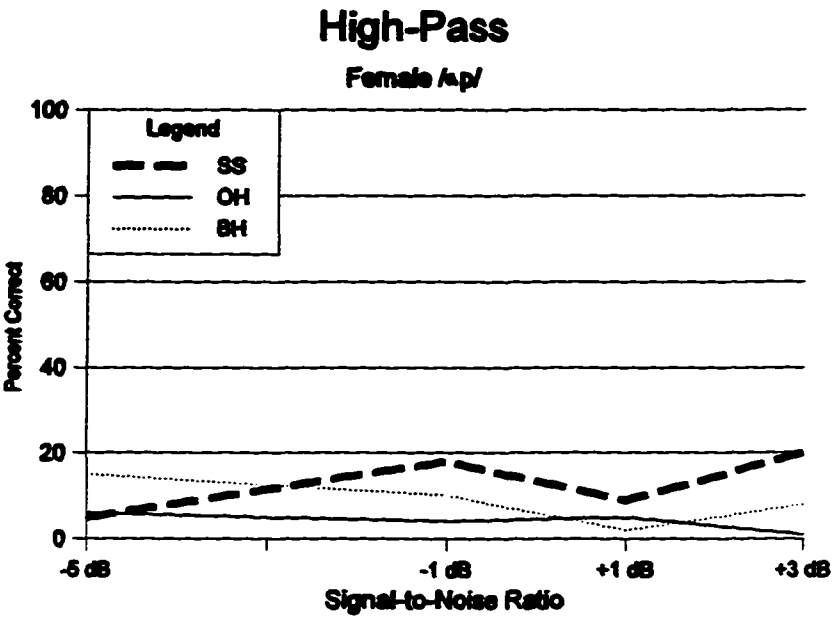
Note: **Underlining indicates statistical significance. Bold-type plus underlining indicates statistical significance below the .001 level.**

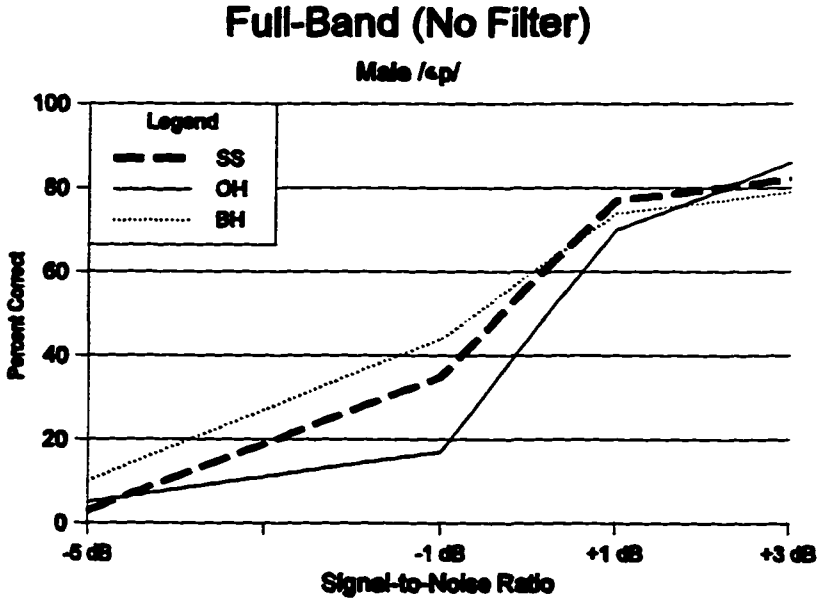
APPENDIX 4

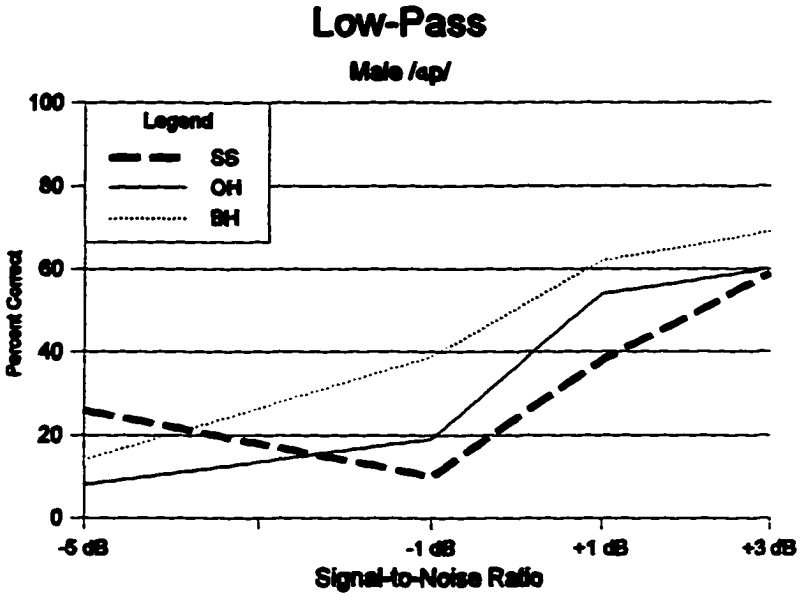
Consonant Recognition Performance for /ɑp/

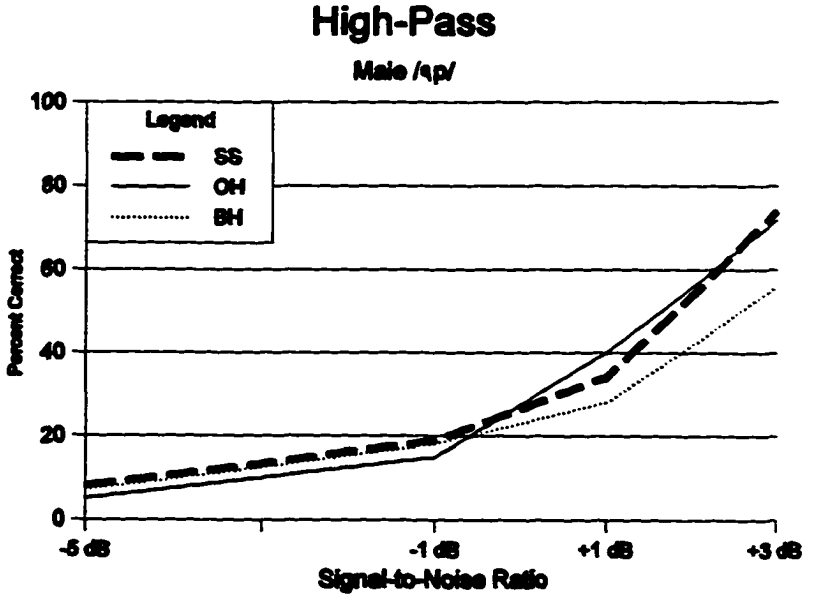




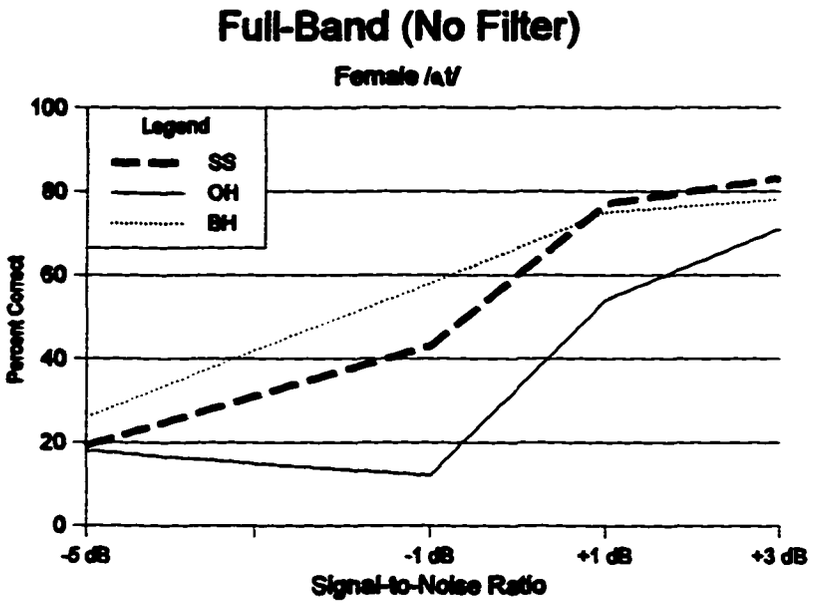


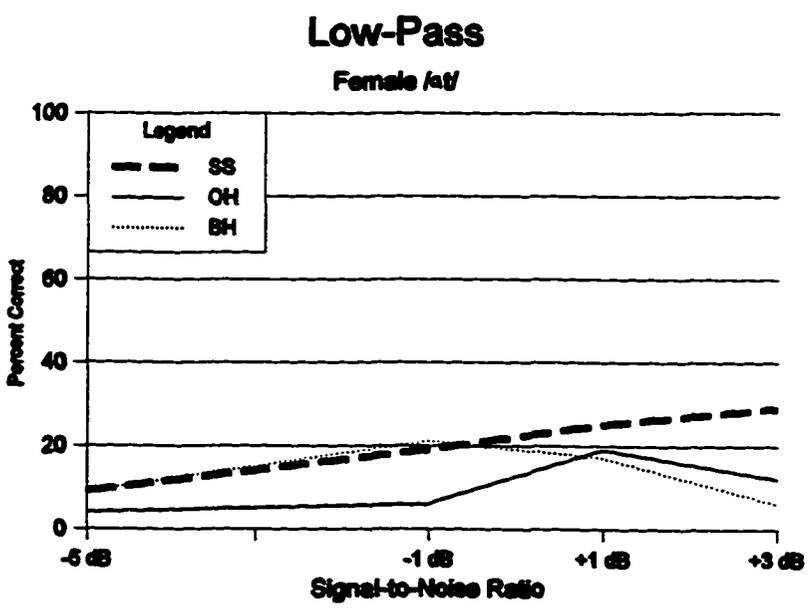


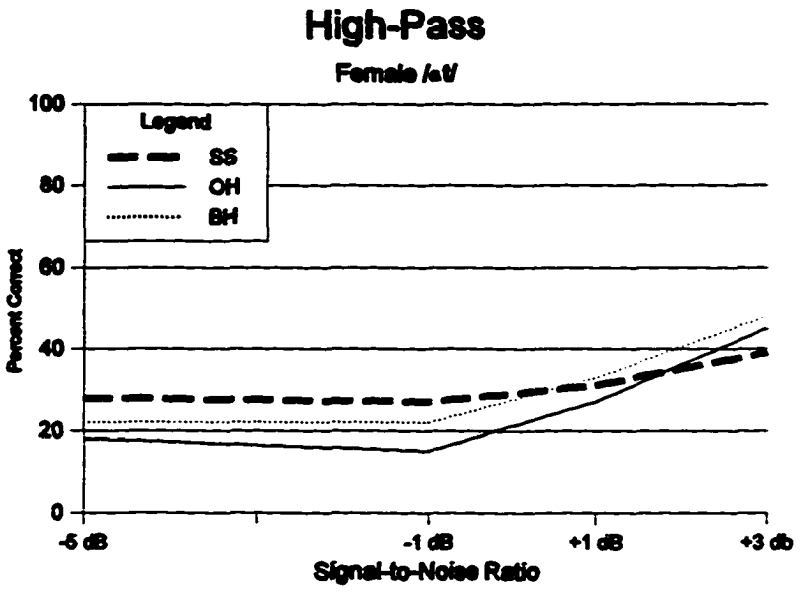


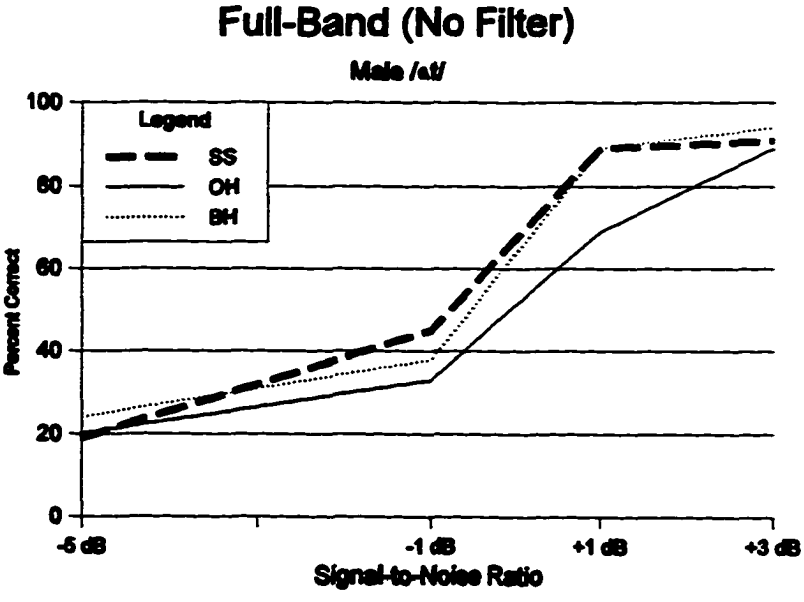


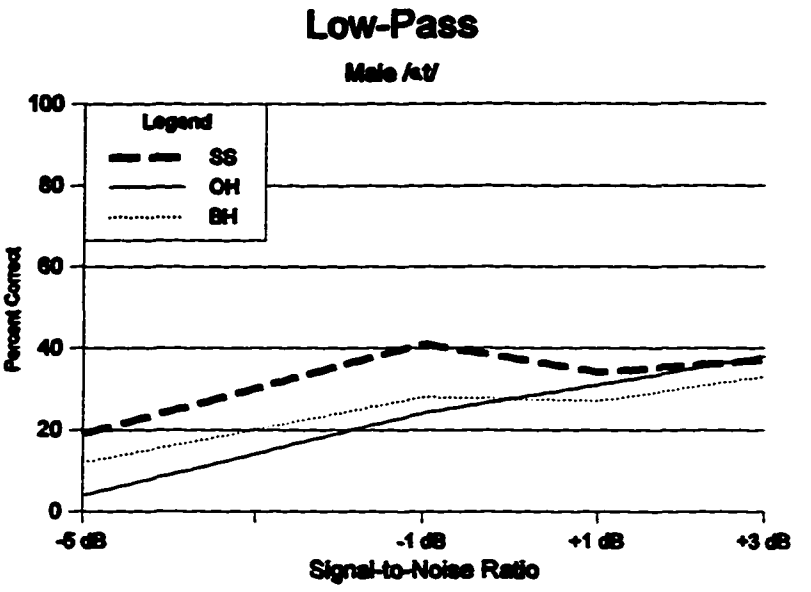
Consonant Recognition Performance for /ɑt/

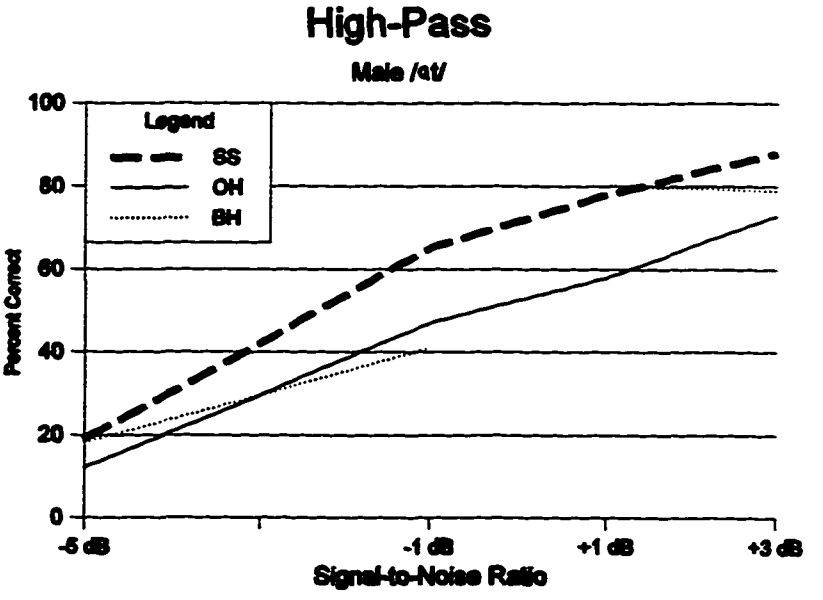




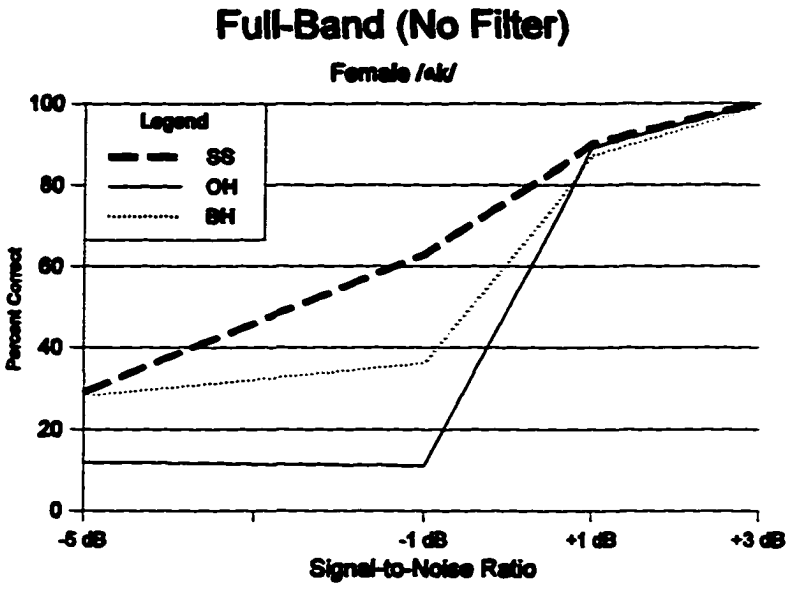


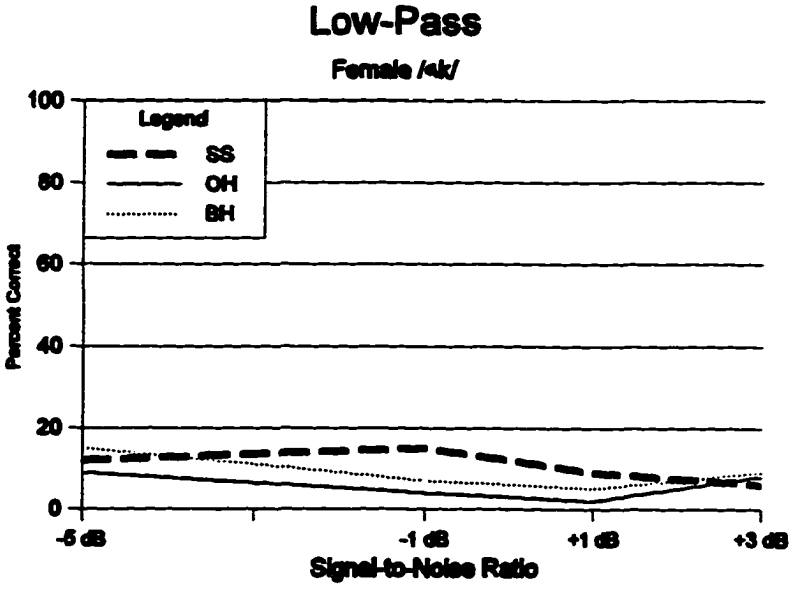


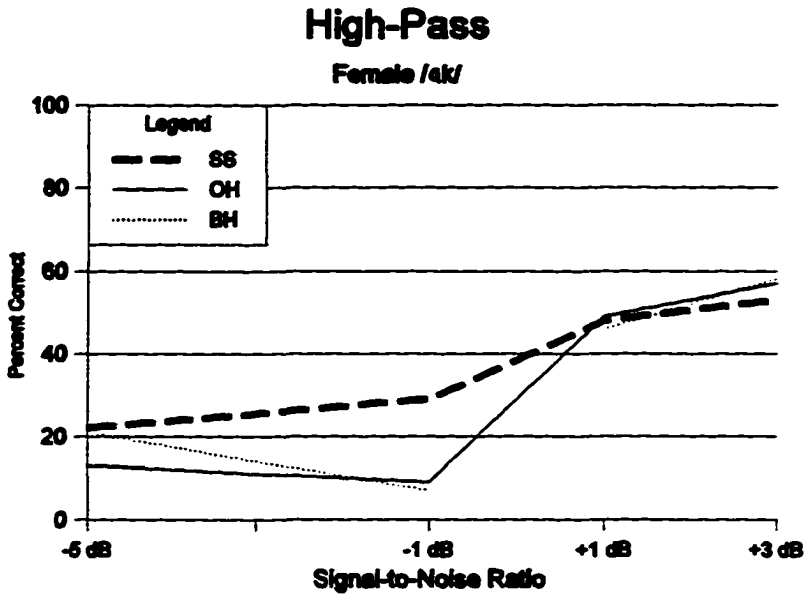


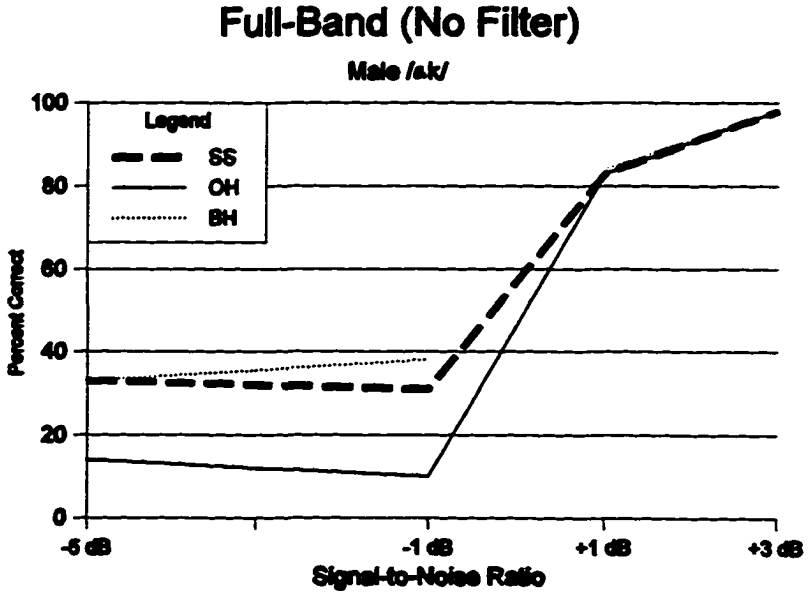


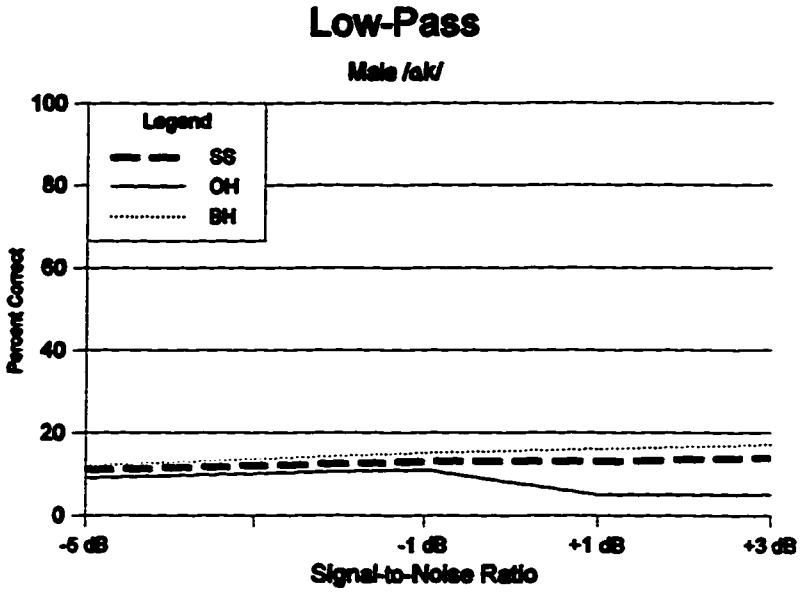
Consonant Recognition Performance for /ɑk/

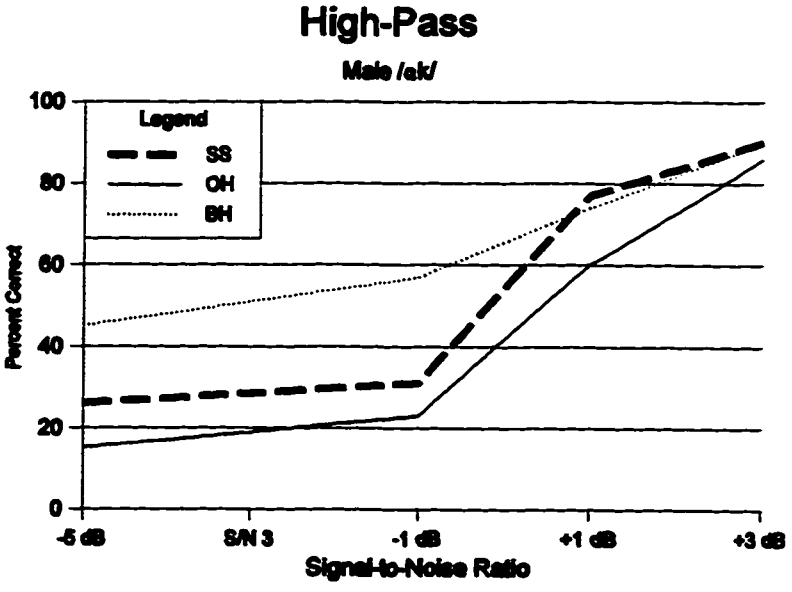




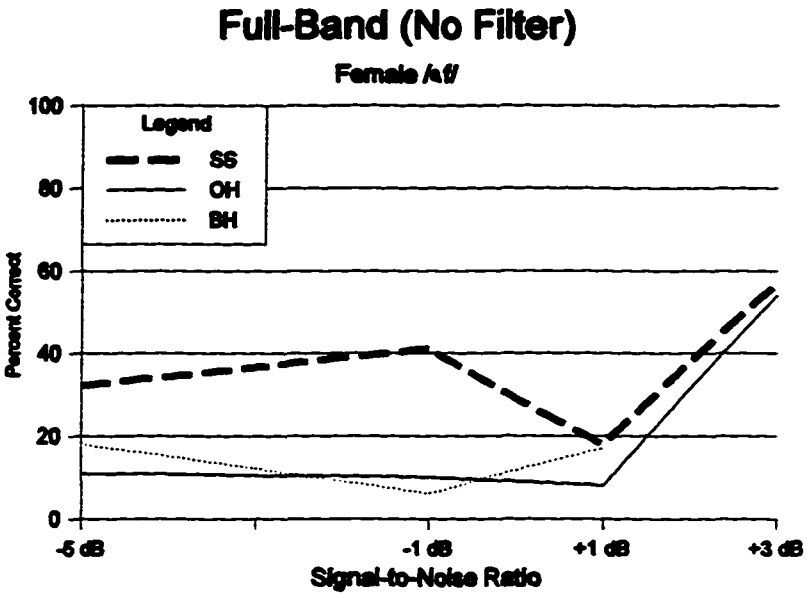


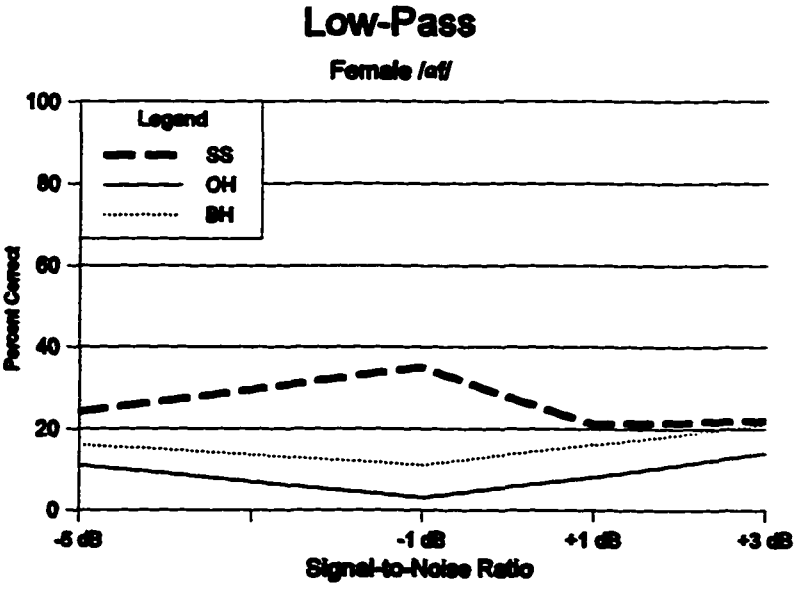


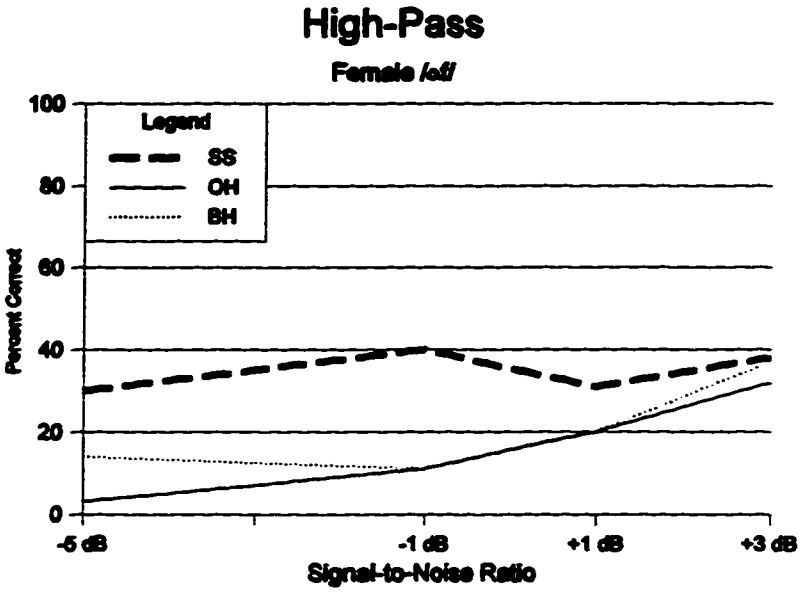


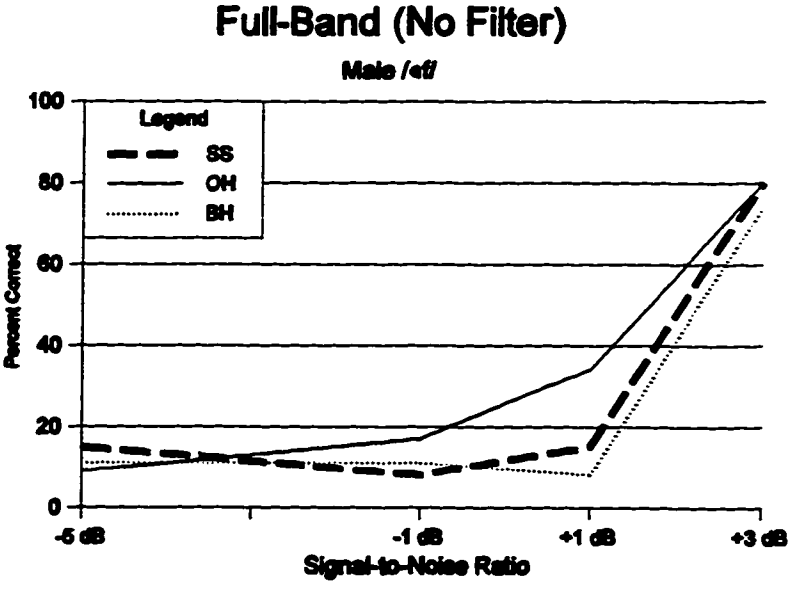


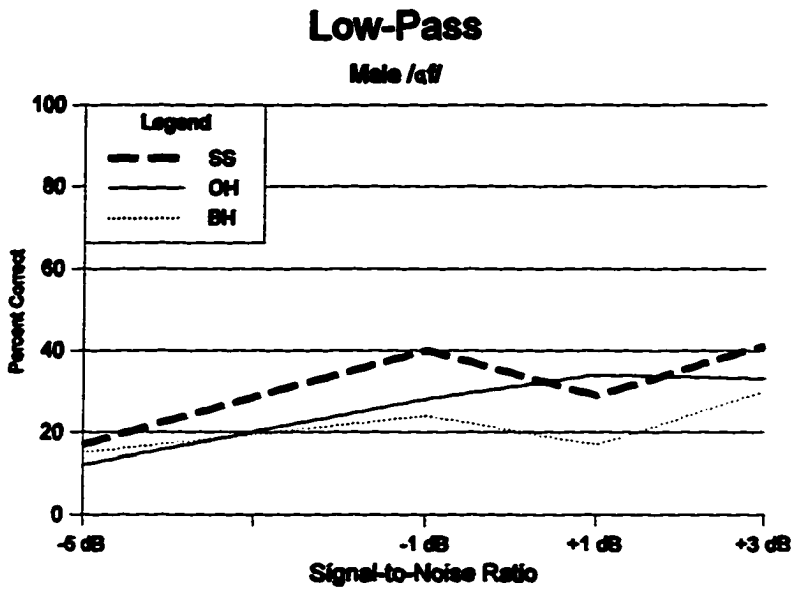
Consonant Recognition Performance for /aʃ/

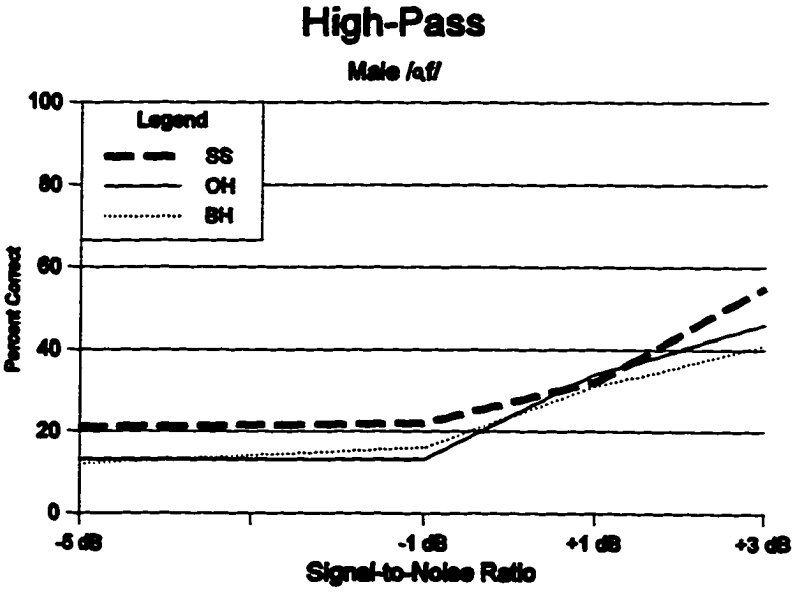




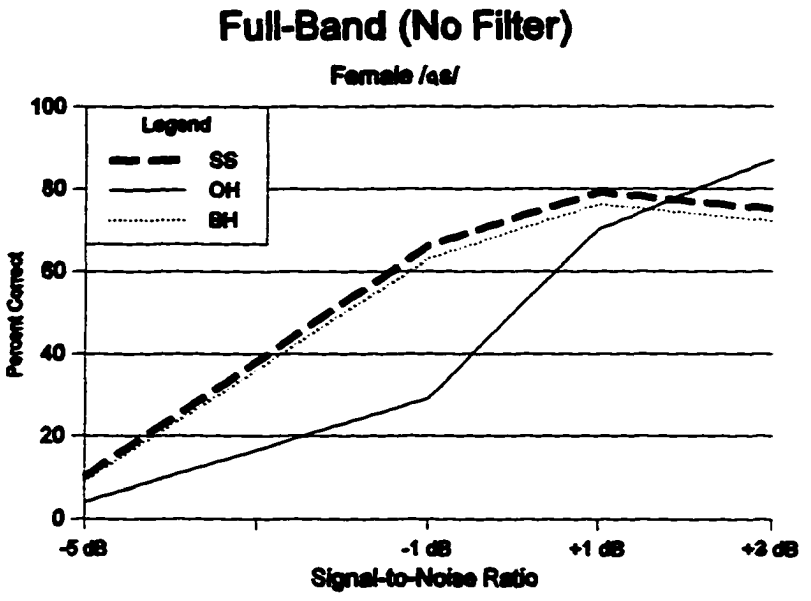


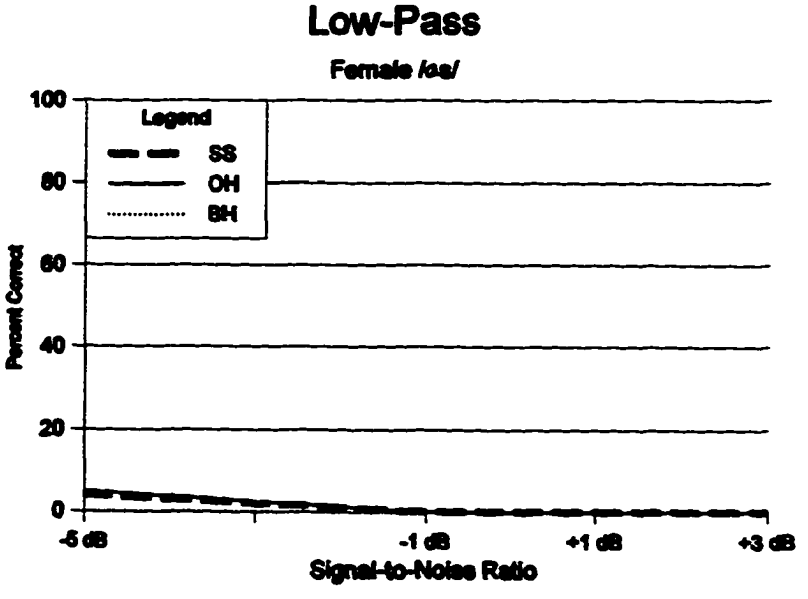


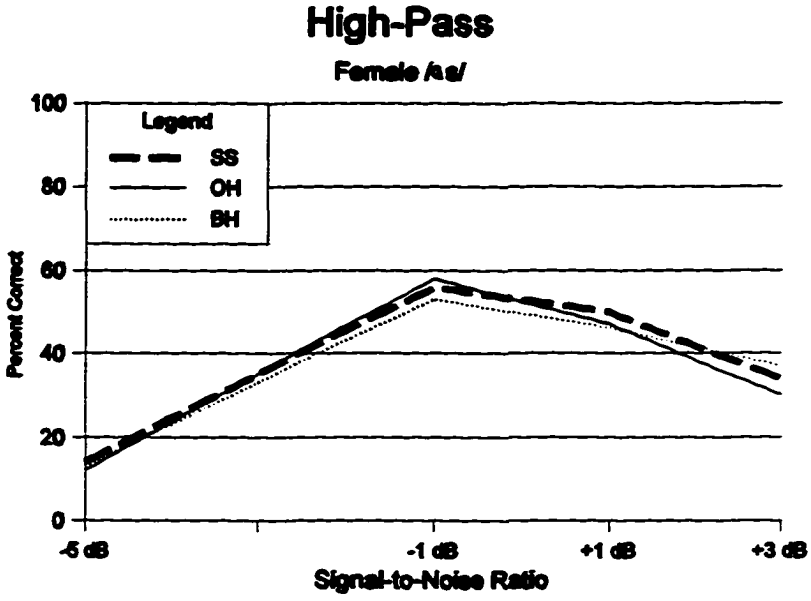


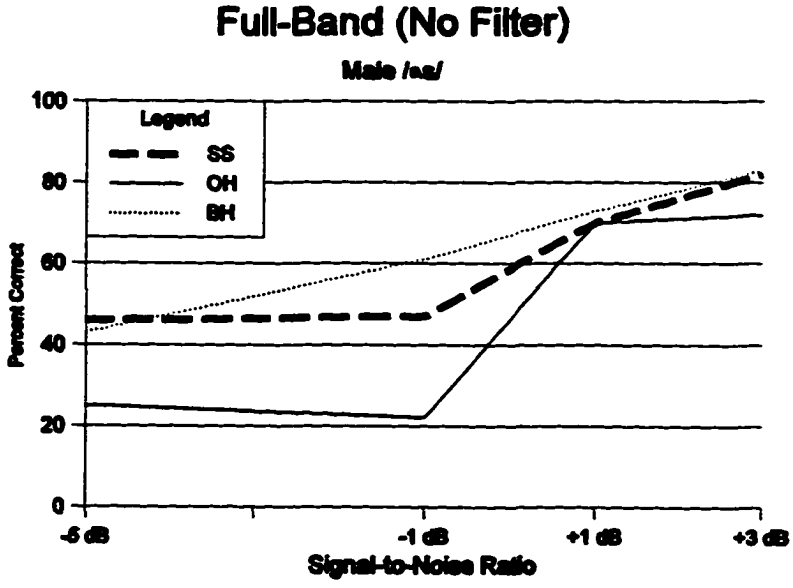


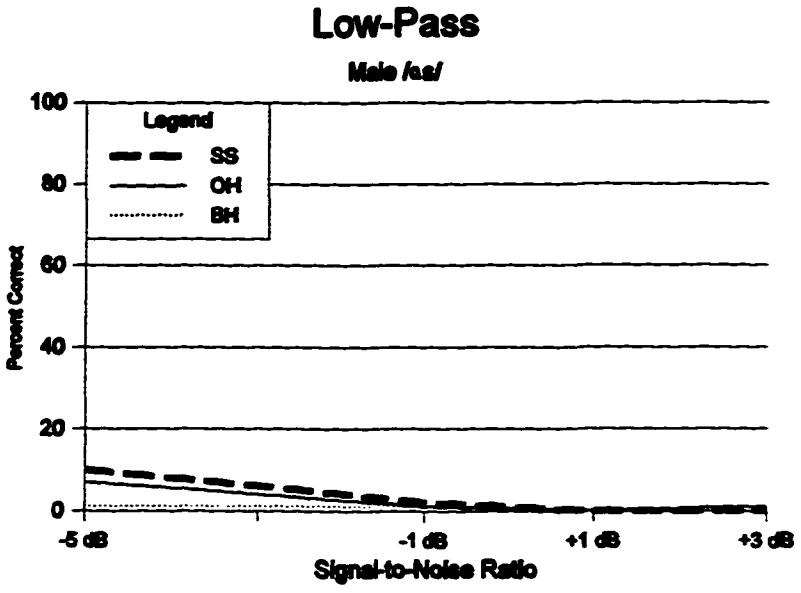
Consonant Recognition Performance for /as/

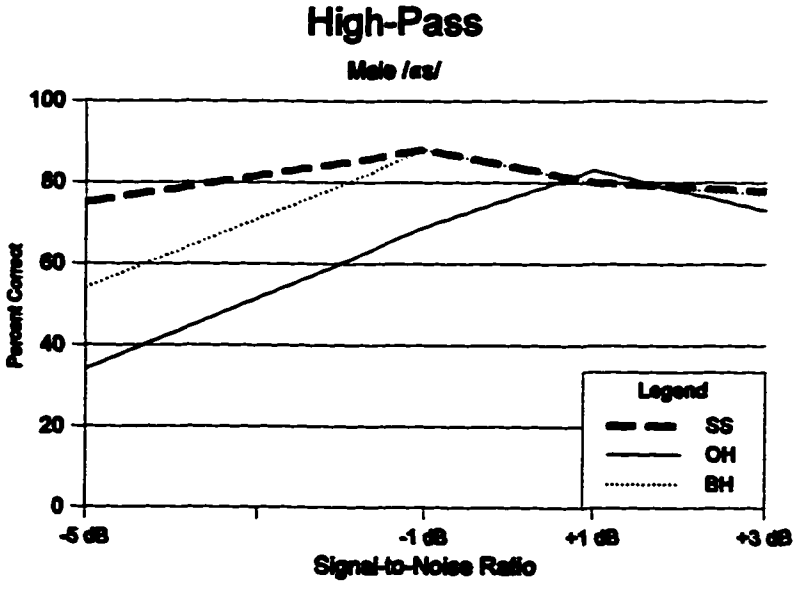




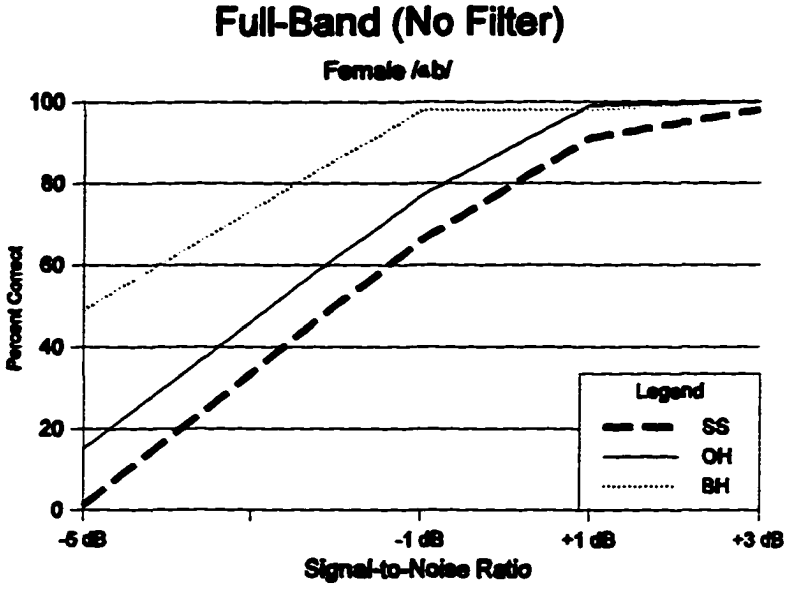


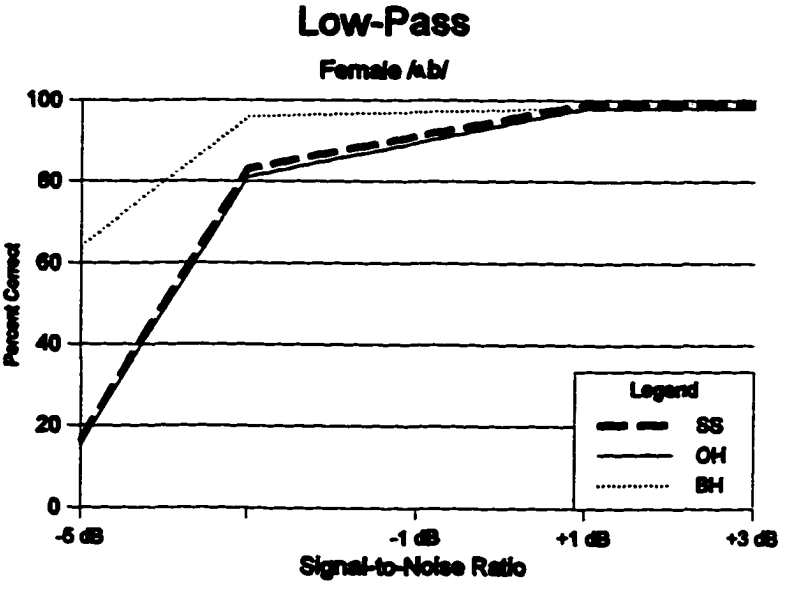


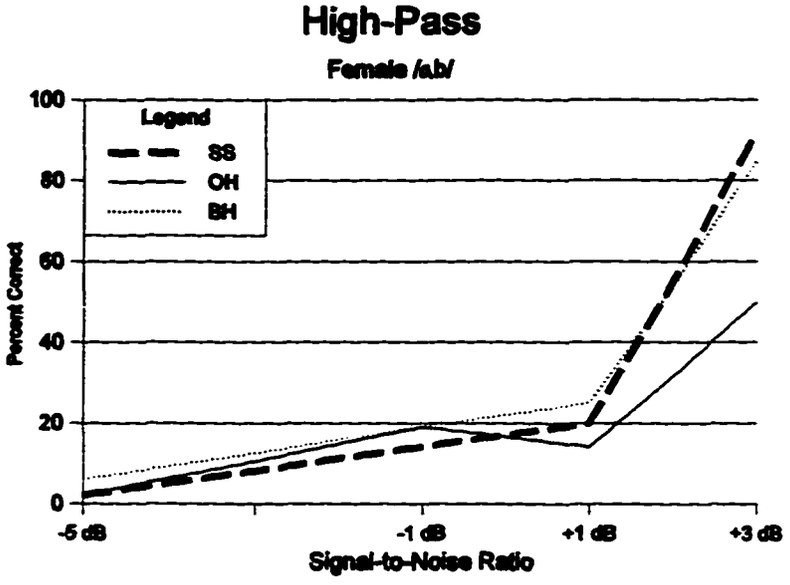


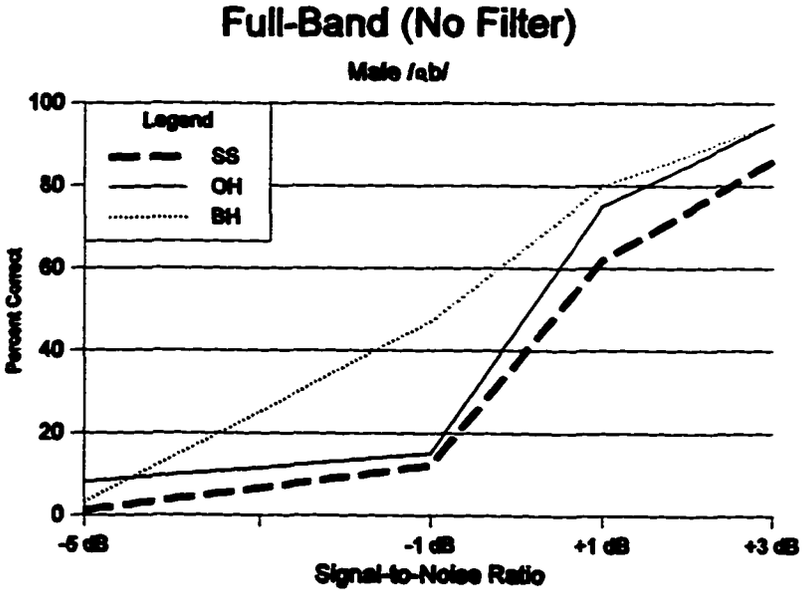


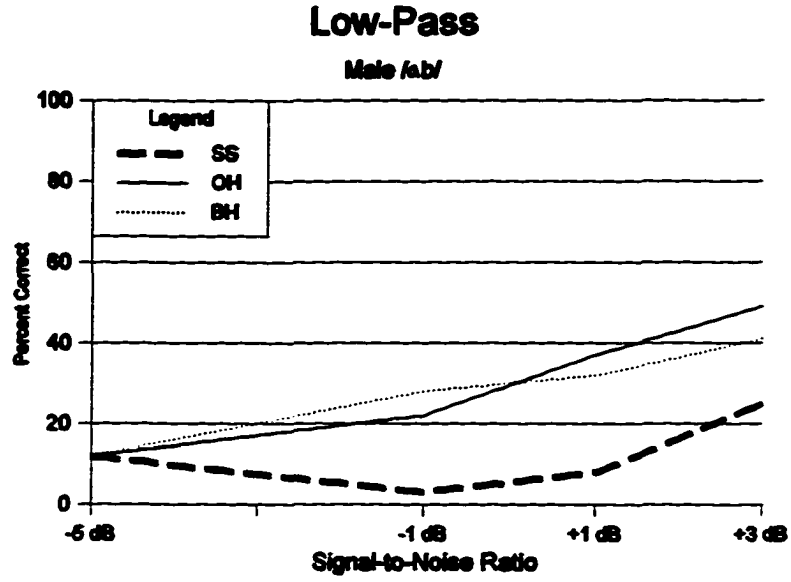
Consonant Recognition Performance for /ɑb/

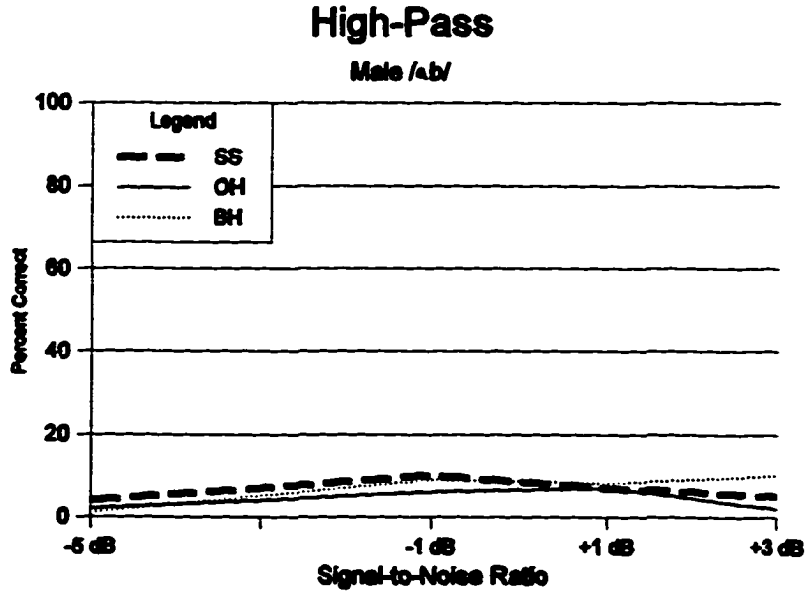




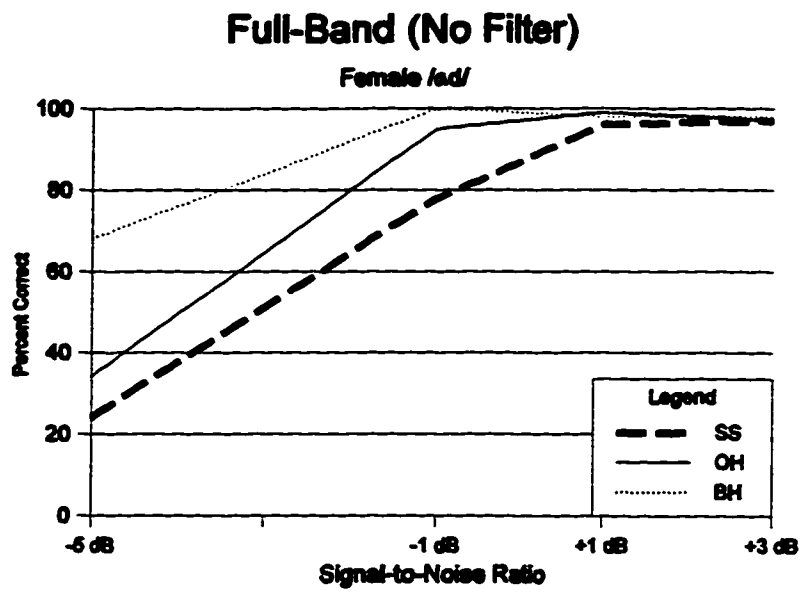


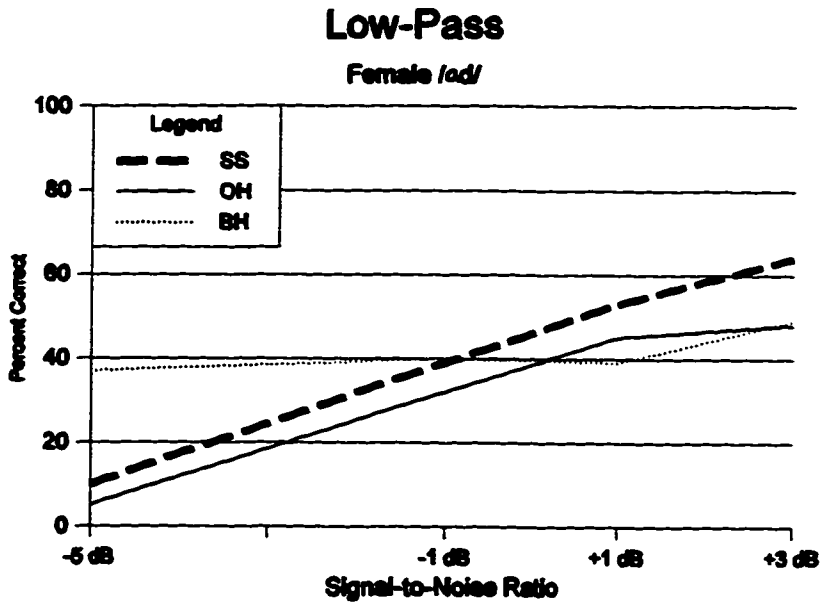


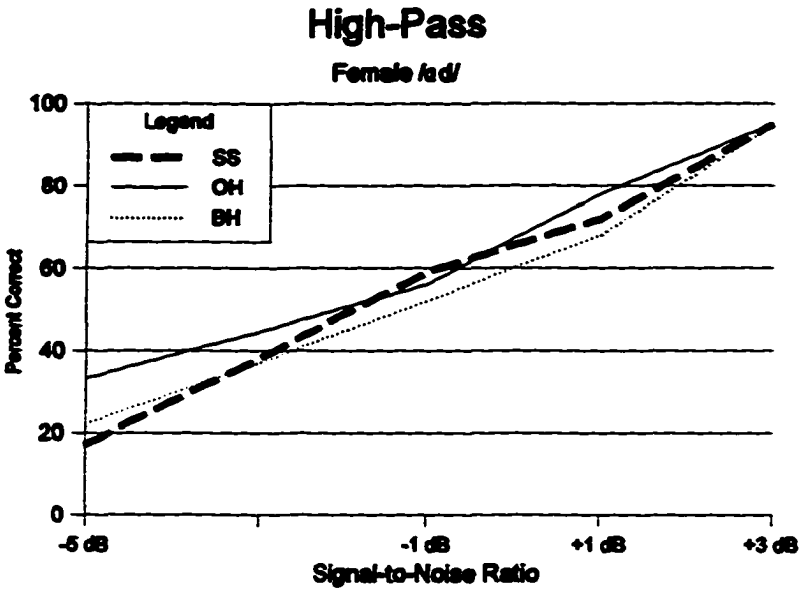


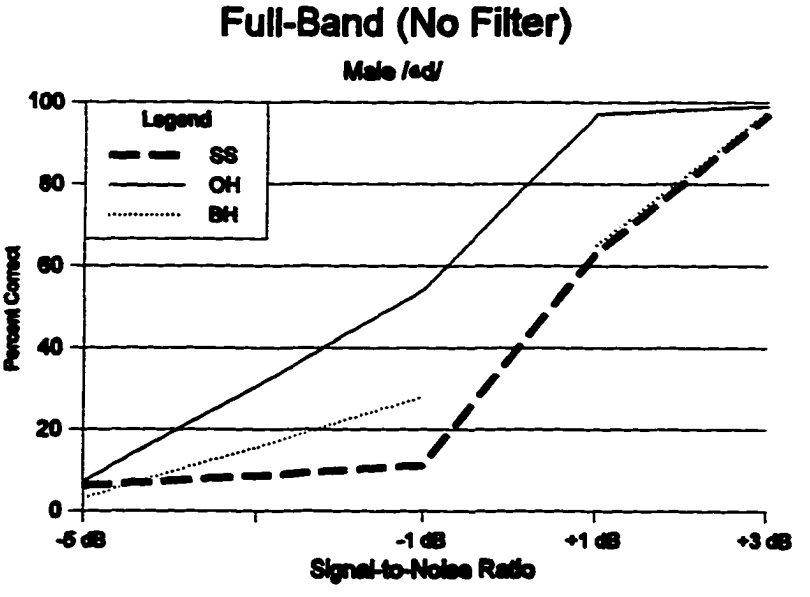


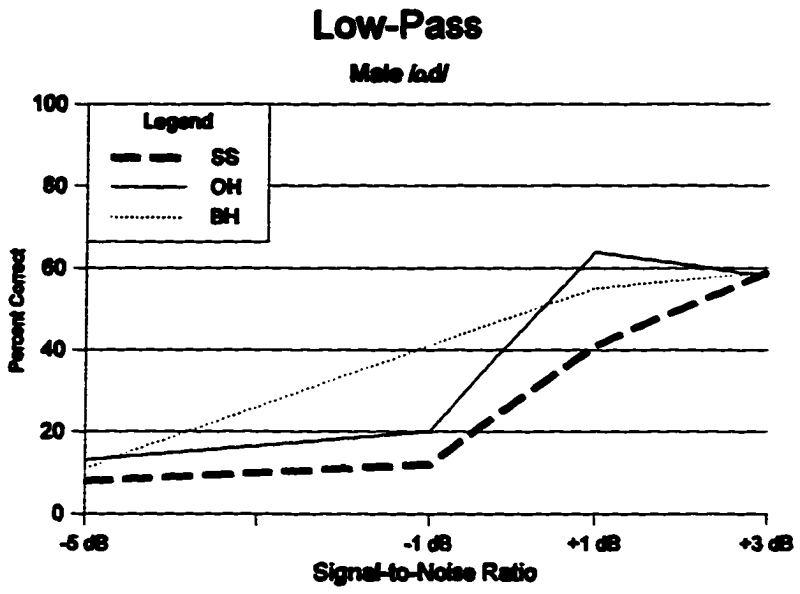
Consonant Recognition Performance for /ad/

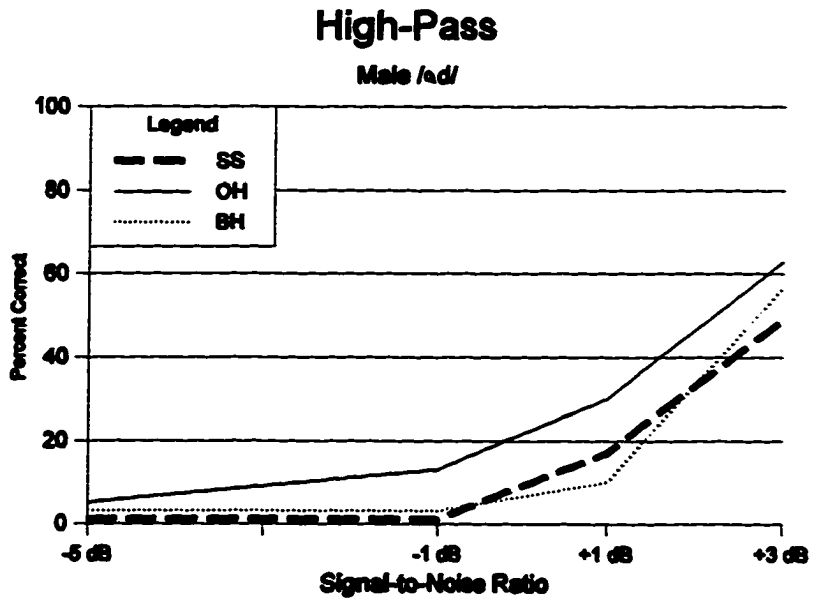




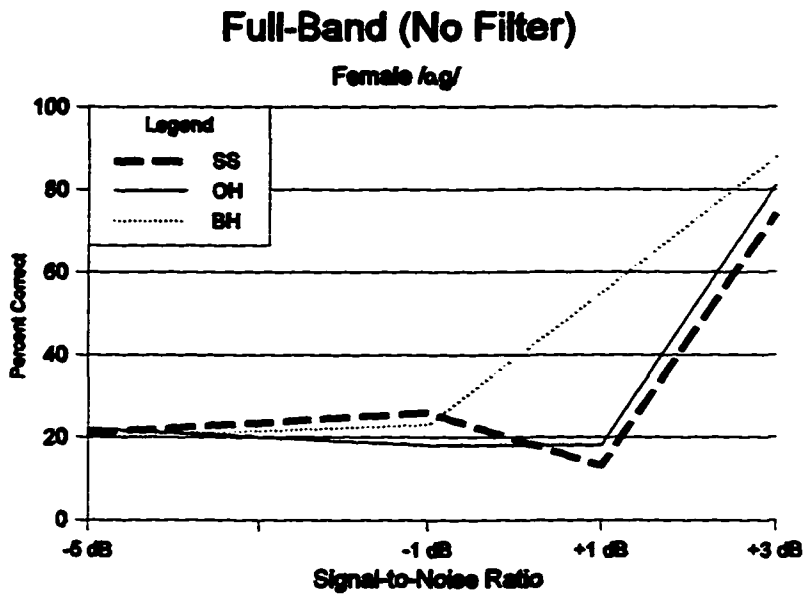


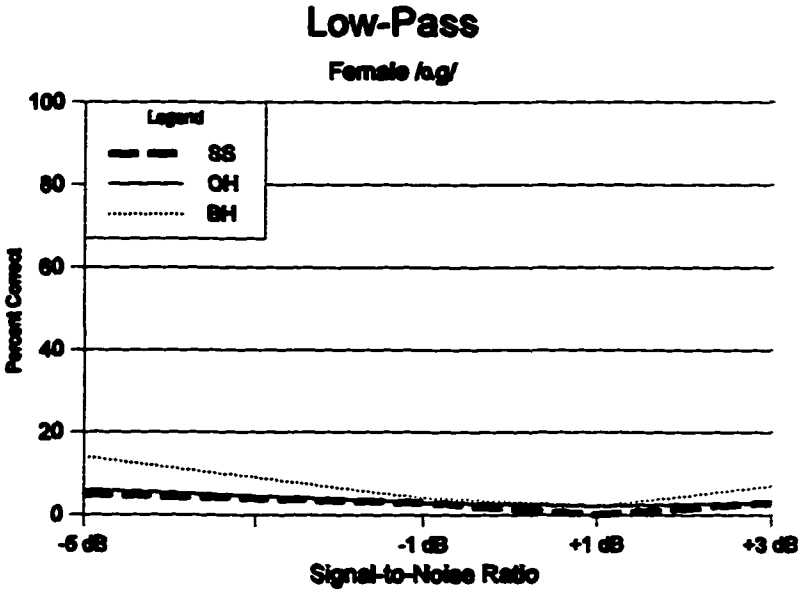


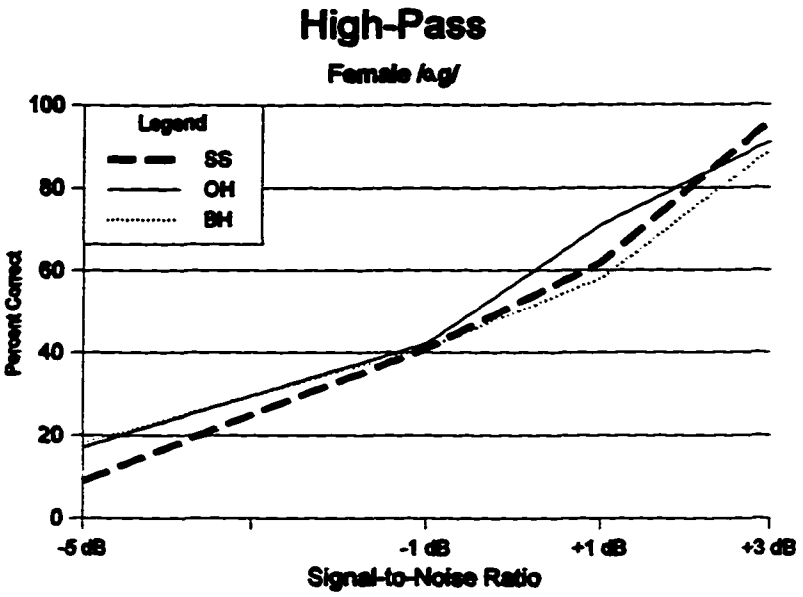


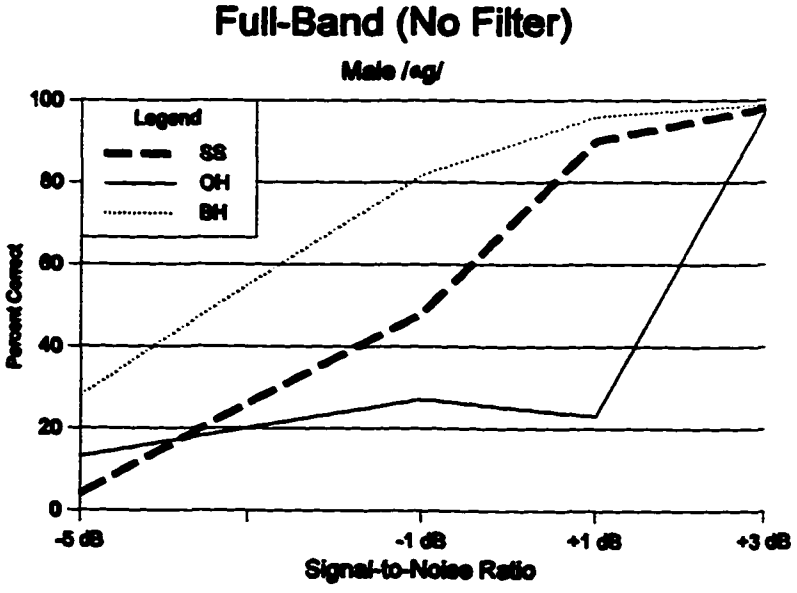


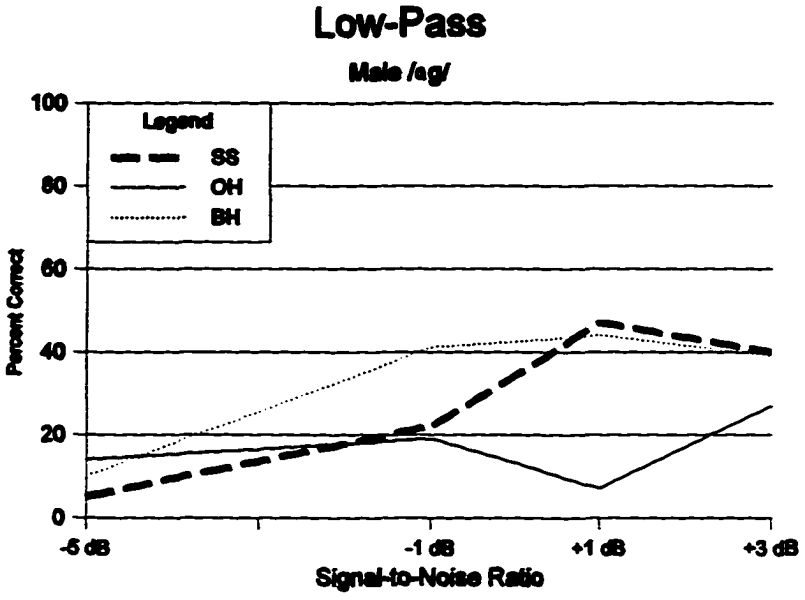
Consonant Recognition Performance for /ɑg/

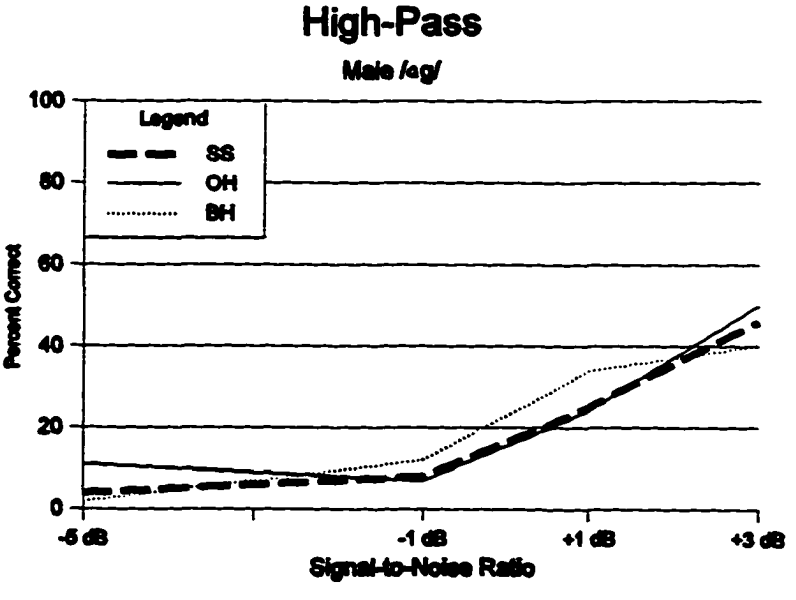




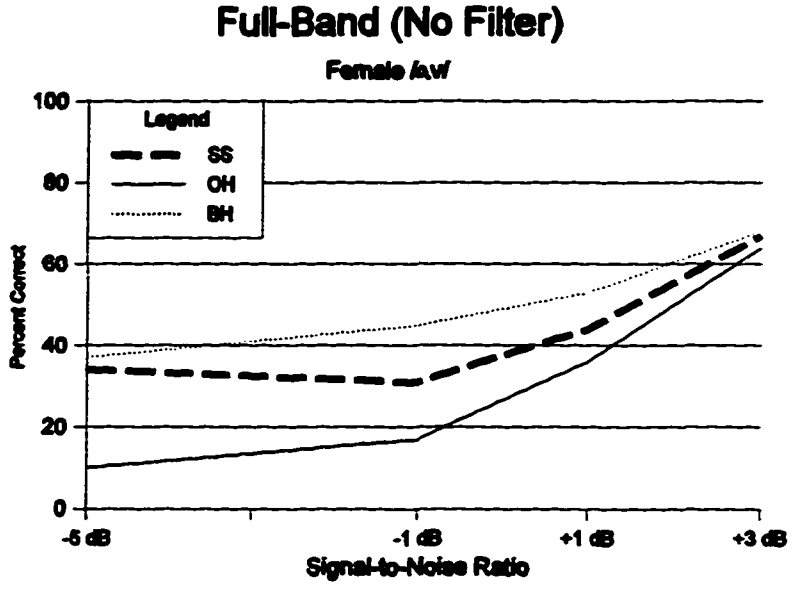


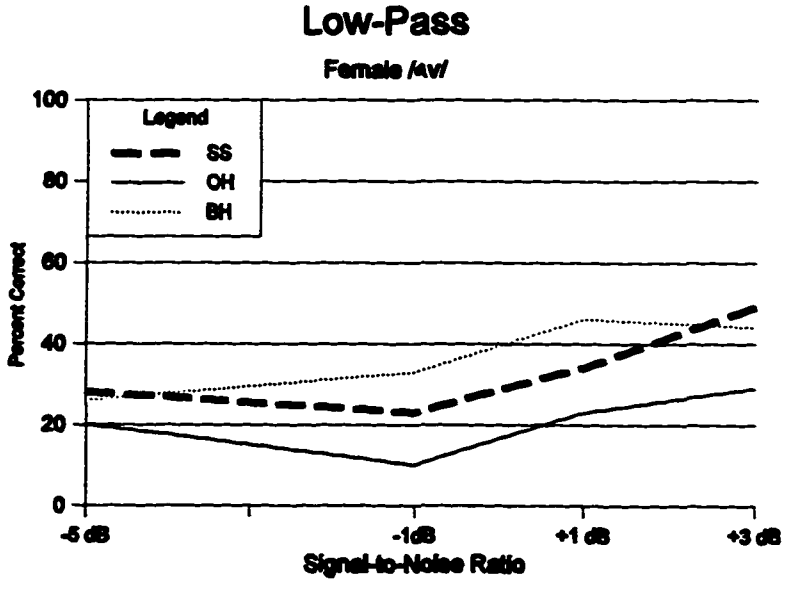


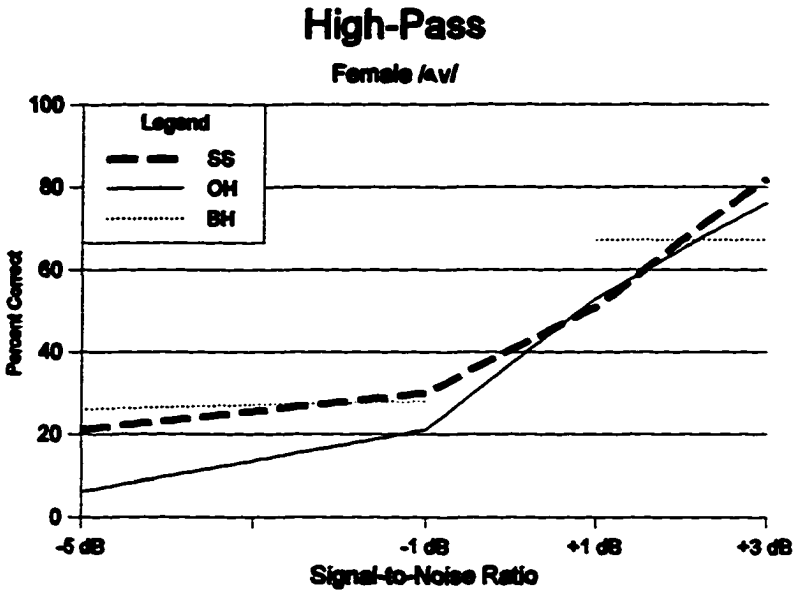


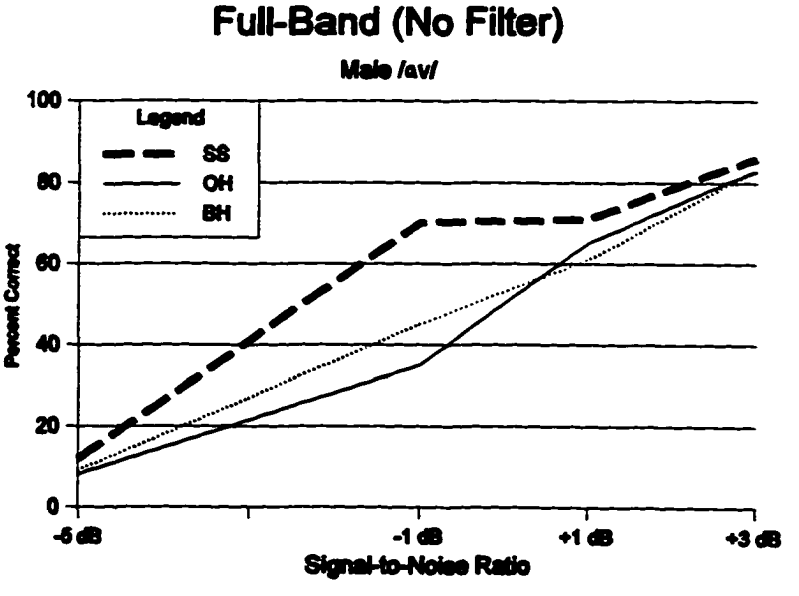


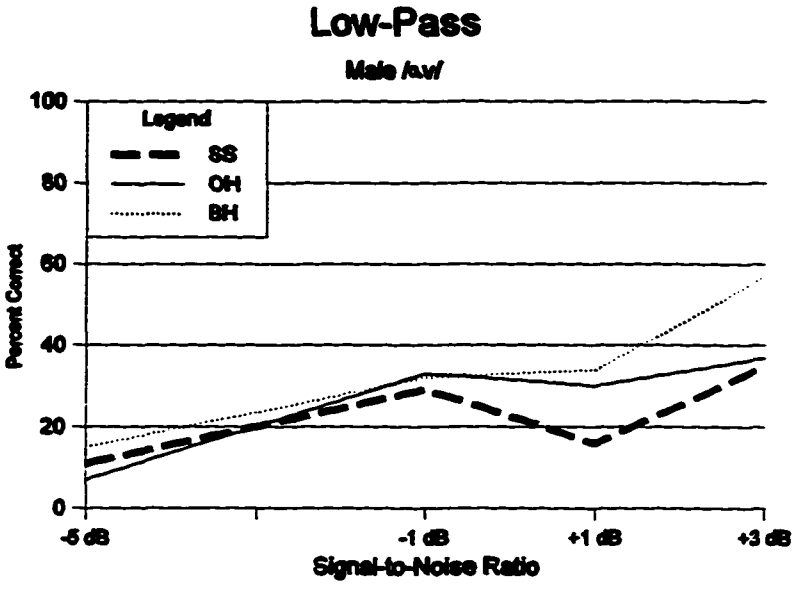
Consonant Recognition Performance for /ɑv/

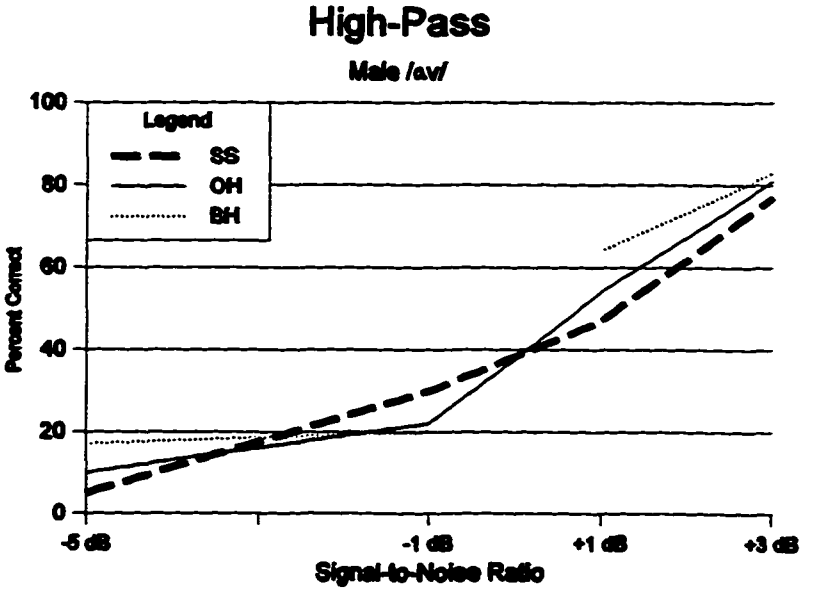




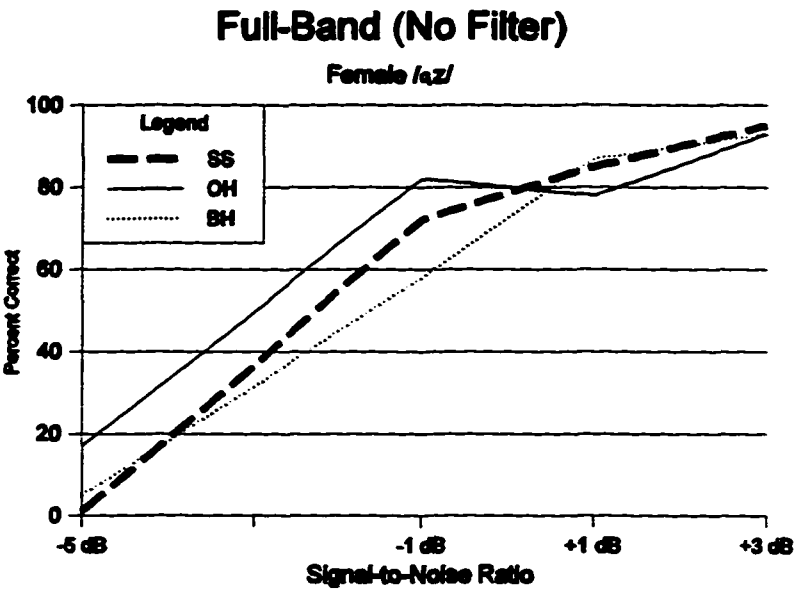


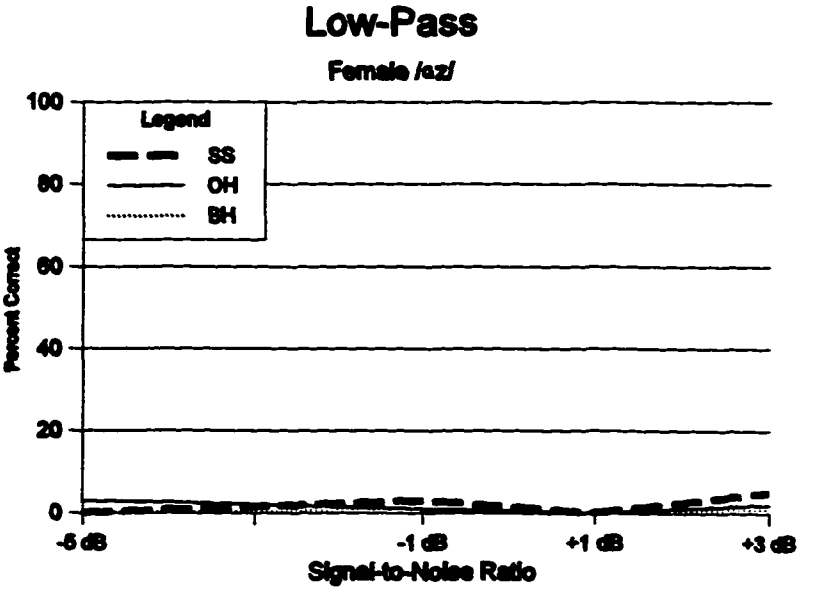


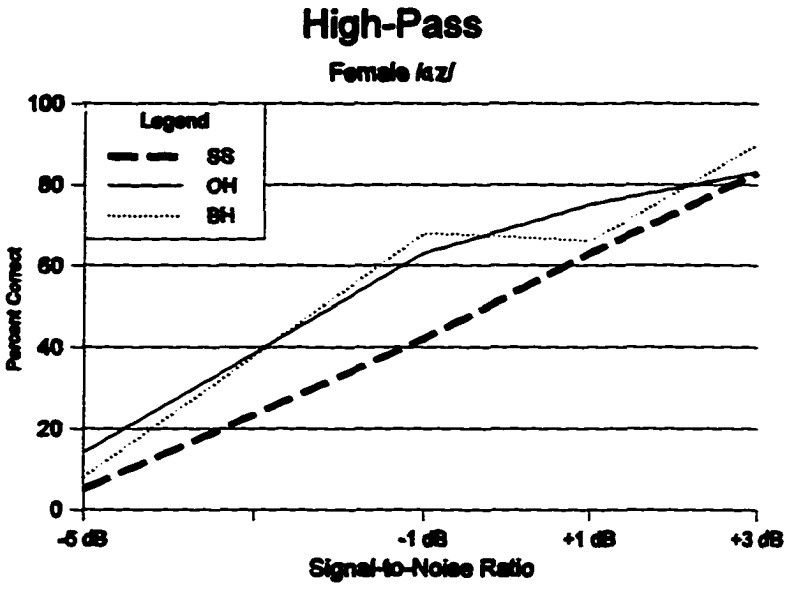


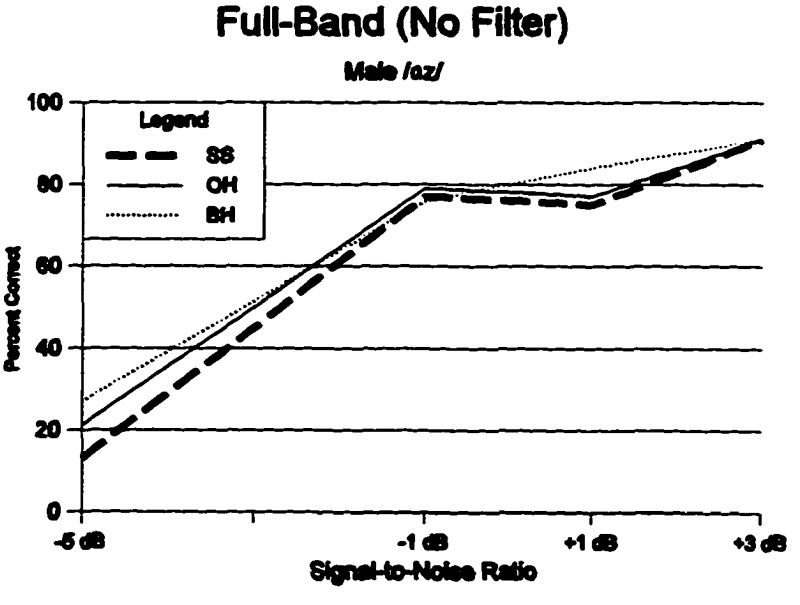


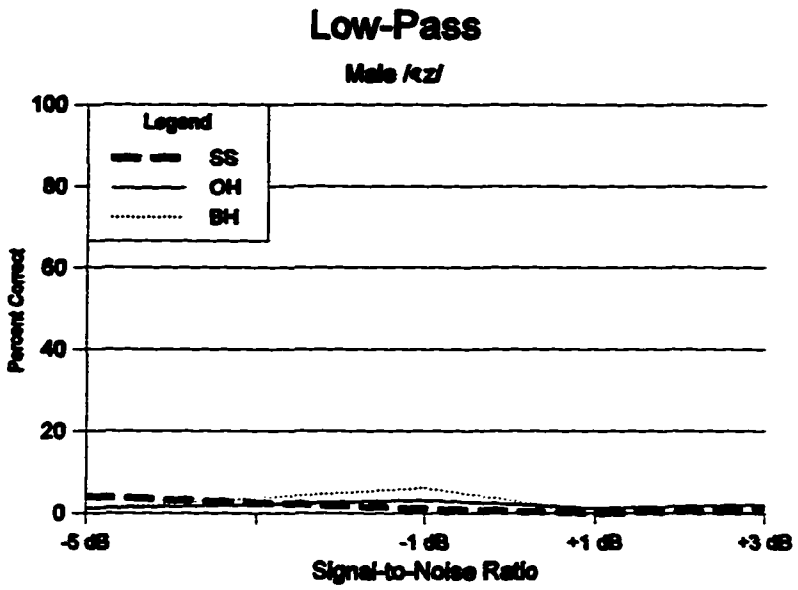
Consonant Recognition Performance for /ɑz/

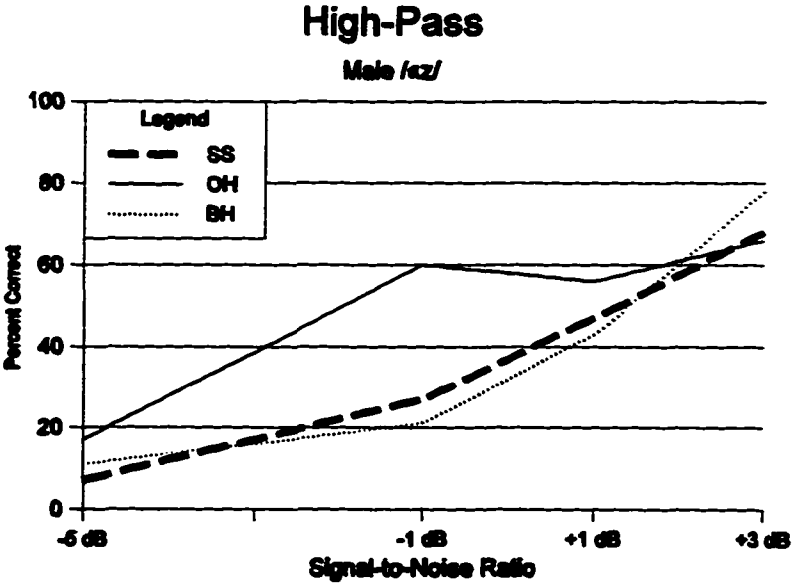












BIBLIOGRAPHY

Abramovitz, R. (1980). Frequency shaping and multiband compression in hearing aids. *J. Comm. Dis.* 13: 483-488.

American National Standards Institute (1989). Specifications for Audiometers. ANSI S3.6-1989, New York.

Brey, R.H., Chabries, D.M., Christiansen, R.W., Robinette, M.S., and Jolley, R. (1984). Improved intelligibility in noise using adaptive filtering I: Normal subjects. *Proc. of ASHA Conv.*, San Francisco, CA.

Brey, R.H., Robinette, M.S., Chabries, D.M., and Christiansen, R.W. (1987). Improvement in speech intelligibility in noise employing an adaptive filter with normal and hearing-impaired subjects. *J. Rehab. Res. Dev.* 24 (4): 75-86.

Chabries, D.M., Christiansen, R.W., Brey, R.H., and Robinette, M.S. (1982). Application of the LMS adaptive filter to improve speech communication in the presence of noise. *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing* 1: 148-151.

Christiansen, R.W., Chabries, D.M., and Lynn, D. (1982). Noise reduction in speech using adaptive filtering: I. Signal processing algorithms. *J. Acoust. Soc. Am.* 71 (1): S7.

Chung, D.Y. and Mack, B. (1979). The effect of masking by noise on word discrimination scores in listeners with normal hearing and with noise-induced hearing loss. *Scand. Aud.* 8: 139-143.

Cohen, R.L. and Keith, R.W. (1976). Use of low-pass noise in word recognition. *J. Sp. Hear. Res. 19*: 48-54.

Cooper, J.C., Jr. and Cutts, B.P. (1971). Speech discrimination in noise. *J. Sp. Hear. Res.* 14: 332-337.

Delattre, P.C., Liberman, A.M., and Cooper, F.S. (1955). Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am.* 27: 769-773.

Dentino, M., McCool, J., and Widrow, B. (1978). Adaptive filtering in the frequency domain. *Proc. IEEE* 66: 1658-1659.

Dirks, D.D., Morgan, D.E., Dubno, J.R. (1982). A procedure for quantifying the effects of noise on speech recognition. *J. Sp. Hear. Dis. 47*: 114-123.

Dubno, J.R., Dirks, D.D. and Schaefer, A.B. (1989). Stop-consonant recognition for normal hearing listeners and listeners with high frequency hearing loss. II: Articulation index predictions. *J. Acoust. Soc. Am.* 85: 355-364.

Dubno, J.R. and Levitt, H. (1981). Predicting consonant confusions from acoustic analysis. *J. Acoust. Soc. Am.* 69: 249-261.

Duggirala, V., Studebaker, G.A., Pavlovic, C.V., and Sherbecoe, R.L. (1988). Frequency importance functions for a feature recognition test material. *J. Acoust. Soc. Am.* 83: 2372 - 2382.

Elliott, L.L. (1962). Backward and forward masking of probe tones of different frequencies. *J. Acoust. Soc. Am.* 34: 1116-1117.

Ferrara, E.R., Jr. (1980). Fast implementation of LMS adaptive filters. *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. ASSP-28 (4): 474-475.

Ferrara, E.R., Jr. (1985). Frequency-domain implementations of periodically time-varying filters. *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. ASSP-33 (4): 883-892..

Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Phys.* 12: 47-65.

Frazier, R.H., Samsam, S., Braida, L.D. and Oppenheim, A.V. (1976). Enhancement of speech by adaptive filtering. *Procs. Intl. Conf. on Acoustics, Speech and Signal Processing*: 251-253.

French, N.R. and Steinberg, J.C. (1947). Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.* 19 (1): 90-119.

Fry, D.B. (1958). Experiments in the perception of stress. *Lang. Speech* 1: 126-152.

Fry, D.B. (1964). The function of the syllable. *Z. Phon. Sprachwiss. Kommunikations-forsch* 17: 215-221.

Gässler, G. (1954), Uber die Horschwelle fur Schallereignisse mit verschieden breitem Frequenzspektrum. *Acustica* 4: 408-414.

Graupe, D. and Causey, G.D. (1977). Method and means for adaptively filtering near-stationary noise from speech. US Patent #4,025,721.

Graupe, D., Grosspietsch, J.K. and Basseas, S.P. (1987). A single-microphone-based self-adaptive filter of noise from speech and its performance evaluation. *J. Rehab. Res. Dev.* 24 (4): 119-126.

Haggard, M.P., Trinder, J.R., Foster, J.R. and Lindblad, A.C. (1987). Two-state compression of spectral tilt: individual differences and psychoacoustical limitations to the benefit from compression. *J. Rehab. Res. Dev.* 24 (4): 193-206.

Harris, K.S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Lang. Speech* 1: 1-7.

Harris, R., Brey, R., Robinette, M., Chabries, D., Christiansen, R., and Jolley, R. (1988). Use of adaptive digital signal processing to improve speech communication for normally hearing and hearing-impaired subjects. *J. Sp. Hear. Res.* 31: 265-271.

Hawkins, D.B. and Yacullo, W.S. (1984). Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation. *J. Sp. Hear. Dis.* 49: 278-286.

Hirsh, I. J., Reynolds, E.G., and Joseph, M. (1954). Intelligibility of different speech materials. *J. Acoust. Soc. Am.* 26 (4): 530-538.

Humes, L.E., Dirks D.D., Bell, T.S., Ahlstron, C., and Kincaid, G.E. (1986). Application of the articulation index and the speech transmission index to the recognition of speech by normal-hearing and hearing-impaired listeners. *J. Sp. Hear. Res.* 29: 447-462.

Kamm, C.A., Dirks, D.D., and Bell, T.S. (1985). Speech recognition and the articulation index for normal and hearing-impaired listeners. *J. Acoust. Soc. Am.* 77: 281-288.

Kates, J.M. (1994). Speech enhancement based on a sinusoidal model. *J. Sp. Hear. Res.* 37: 449-464.

Keith, R. and Talis, H. (1972). The effects of white noise on PB scores of normal and hearing impaired listeners. *Audiology 11*: 177-183.

Killion, M. (1984). New insert earphones for audiometry. *Hear. Instrum.* 35: 45-46.

Kozhevnikov, V.A., and Chistovich, L.A. (1965). Speech: Articulation and Perception. Translated by the Joint Publications Research Service. Clearinghouse for Federal Scientific and Technical Information, U.S. Dept. of Commerce, Washington DC (Pub # JPRS: 30, 543; TT: 65-31233).

Ladefoged, P. (1963). Some physiological parameters in speech. *Lang. Speech* 6: 109-119.

Leshowitz, B. (1977). Speech intelligibility in noise for listeners with sensorineural hearing damage. *IPO Ann. Prog. Rep.* 12: 11-23.

Levitt, H. (1986). Hearing impairment and sensory aids: A tutorial review. *J. Rehab. Res. Dev.* 23: xiii-xviii.

Liberman, A. M., Delattre, P. C., Cooper, F.S., and Gerstman, L. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Monogr.* 68: 1-13.

Lieberman, P. (1967). *Intonation, Perception and Language*. Cambridge, MA: MIT Press.

Lieberman, P. and Michaels, S.B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *J. Acoust. Soc. Am.* 34: 922-927.

Lim, J.S. (1982). Signal processing for speech enhancement. In *The Vanderbilt Hearing Aid Report*, Studebaker, G.A. and Bess, F.H. (eds). Upper Darby, PA: Monographs in Contemporary Audiology, 124-129.

Lim, J.S. , ed. (1983). *Speech Enhancement*. Englewood Cliffs: Prentice-Hall, Inc.

Lim, J.S. and Oppenheim, A.V. (1979). Enhancement and bandwidth compression of noisy speech. *Proc. IEEE* 67: 1586-1604.

Lim, J.S., Oppenheim, A.V., and Braida, L.D. (1978). Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition. *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. ASSP-26: 354-358.

Lindblom, B. and Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw and larynx movement. *J. Acoust. Soc. Am.* 50: 1166.

Ling, D. (1976). *Speech and the Hearing-Impaired Child: Theory and Practice*. Washington, DC: The Alexander Graham Bell Association for the Deaf, Inc.

Lippman, R., Braida, L., and Durlach, N. (1981). Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. *J. Acoust. Soc. Am.* 69: 524-531.

Martin, E. and Pickett, J. (1970). Sensorineural hearing loss and upward spread of masking. *J. Sp. Hear. Res.* 13: 426-437.

McAulay, R.J., and Quatieri, T.F. (1986). Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34: 744-754.

Miller, G.A. (1947). The masking of speech. *Psych. Bull.* 44: 105-129.

Miller. G.A., Heise, G.A., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *J. Exp. Psych.* 41: 329-335.

Moore, B.C.J. (1987). Design and evaluation of a 2-channel compression hearing aid. *J. Rehab. Res. Dev.* 24: 181-192.

Moore, B.C.J. and Glasberg, B.R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hear. Res.* 28: 209-225, 1987.

Morton, J. and Carpenter, A. (1963). Experiments relating to the perceptio of formants. *J. Acoust. Soc. Am.* 35: 475-480.

Nabelek, I. (1983). Performance of hearing-impaired listeners under various types of amplitude compression. *J. Acoust. Soc. Am.* 74 (3): 776-791.

Niemeyer, W. (1965). Speech audiometry with phonetically balanced sentences. *Intl. Audiol. 4*: 97-101.

Pavlovic, C.V. (1987). Derivation of primary parameters and procedures for use in speech intelligibility predictions. *J. Acoust. Soc. Am.* 82: 413-422.

Pavlovic, C.V. (1989). Speech spectrum considerations and speech intelligibiltiy predictions in hearing aid evaluations. *J. Sp. Hear. Dis.* 54: 3-8.

Pavlovic, C.V., Studebaker, G.A., and Sherbecoe, R.L. (1986). An articulation index-based procedure for predicting the speech recognition performance of hearing impaired individuals. *J. Acoust. Soc. Am.* 80: 50-57.

Pearsons, K., Bennett, R.S. and Fidell, S. (1976). *Speech levels in various environments*, Bolt, Beranek and Newman, Inc., Report No. 3281. Prepared for the Office of Resources and Development, EPA.

Peterson, G. and Barney, H. (1952). Control methods used in a study of vowels. *J. Acoust. Soc. Am.* 24: 175-184.

Peterson, P.M., Durlach, N.I., Rabinowitz, W.M., and Zurek, P.M. (1987). Multimicrophone adaptive beamforming for interference reduction in hearing aids. *J. Rehab. Res. Dev.* 24 (4): 103-110.

Plomp, R. (1964). The ear as a frequency analyser. *J. Acoust. Soc. Am.* 36: 1628-1636.

Plomp, R. (1977). Acoustical aspects of cocktail parties. *Acustica* 38: 186-191.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *J. Acoust. Soc. Am.* 83 (6): 2322- 2327.

Plomp, R. and Mimpen A.M. (1979). Speech reception threshold for sentences as a function of age and noise level. *J. Acoust. Soc. Am.* 66: 1333-1342.

Ross, M., Huntington, D.A., Newby, H.A., and Dixon, R.F. (1965). Speech discrimination of hearing-impaired individuals in noise. *J. Aud. Res.* 5: 47-72.

Sambur, M.R. (1978). Adaptive noise cancelling for speech signals. *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. ASSP-26 (5): 419-423.

Schafer, R. (1982). Speech processing for the hearing impaired. In *The Vanderbilt Hearing Aid Report*, Studebaker, G.A. and Bess, F.H. (eds). Upper Darby, PA: Monographs in Contemporary Audiology, 130-132.

Scharf, B. (1961). Complex sounds and critical bands. *Psychol. Bull.* 58: 205-217.

Scharf, B. (1970). Critical Bands. In *Foundations of Modern Auditory Theory*, J.V. Tobias (ed.), New York: Academic Press, pp. 157-202.

Schum, D.J. (1990). Noise reduction strategies for elderly, hearing-impaired listeners. *J. Am. Acad. Audiol.* 1 (1): 31-36.

Schwander, T.J. and Levitt, H. (1987). Effect of two-microphone noise reduction on speech recognition by normal-hearing listeners. *J. Rehab. Res. Dev.* 24 (4): 87-92.

Sigelman, J. and Preves, D.A. (1987). Field trials of a new adaptive signal processor hearing aid circuit. *Hear. J.* 40 (4): 24-29.

Skinner, M. W. and Miller, J.D. (1983). Amplification bandwidth and intelligibility of speech in quiet and noise for listeners with sensorineural hearing loss. *Audiol.* 22: 253- 279.

Stach, B.A., Speerschneider, J.M., and Jerger, J.F. (1987). Evaluating the efficacy of automatic signal processing hearing aids. *Hear. J.* 40 (3): 15-19.

Stein, L. K. and Dempsey-Hart, D. (1984). Listener-assessed intelligibility of a hearing aid self-adaptive noise filter. *Ear Hear.* 5 (4): 199-204.

Studebaker, G.A., Pavlovic, C. V., and Sherbecoe, R.L. (1987). A frequency importance function for continuous discourse. *J. Acoust. Soc. Am.* 81: 1130-1138.

Studebaker, G.A., and Sherbecoe, R.L. (1989). Estimated versus observed performance on a monosyllabic word test. Paper presented at the annual meeting of the American Speech-Language-Hearing Association, St. Louis, MO.

Studebaker, G.A., and Sherbecoe, R.L. (1991). Frequency importance and transfer functions for recorded CID W-22 word lists. *J. Sp. Hear. Res.* 34: 427-438.

Studebaker, G.A. and Sherbecoe, R.L. (1993). Frequency-importance functions for speech recognition. In Studebaker, G.A. and Hochberg, I., eds. *Acoustical factors affecting hearing aid performance*, 2nd ed. Needham Heights: Allyn and Bacon.

Suter, A. H. (1985). Speech recognition in noise by individuals with mild hearing impairments. *J. Acoust. Soc. Am.* 78: 887-900.

Tukey, J.W. (1953). *The problem of multiple comparisons.* Princeton: Princeton University.

Tyler, R.S. and Kuk, F.K. (1989). The effects of "noise suppression" hearing aids on consonant recognition in speech-babble and low-frequency noise. *Ear Hear.* 10: 243- 249.

Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. *Lang. Speech* 3: 223-234.

Van Tassell, D. J., Larsen, S.Y., and Fabry, D.A. (1988). Effects of an adaptive filter hearing aid on speech recognition in noise by hearing-impaired subjects. *Ear Hear.* 9 (1): 15-21.

Walker, G., Byrne, D., and Dillon, H. (1984). The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise. *J. Acoust. Soc. Am.* 76 (3): 746-757.

Wegel, R.L. and Lane, C.E. (1924). The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.* 23: 266-285.

Widrow, B. (1966). *Adaptive filters I: Fundamentals.* Stanford, CA: Stanford Electronics Lab Report SU-SEL-66-126.

Winer, B.J. (1971). *Statistical principles in experimental design,* 2nd ed. New York: McGraw-Hill, Inc.

Wolinsky, S. (1986). Clinical assessment of a self-adaptive noise filtering system. *Hear. J.* 39 (10): 29-32.

Zwicker, E. (1954). Die Verdeckung von Schmalbandgerauschen durch Sinustone. *Acustica* 4: 415-420.

Zwicker, E., Flottorp, G., and Stevens, S.S. (1957). Critical bandwidth in loudness summation. *J. Acoust. Soc. Am.* 29: 548-557.