EVALUATION OF

INFINITE PEAK CLIPPING

AS A MEANS OF AMPLITUDE COMPRESSION

by

Karen Ann Silletto

B.S.E.E., University of California at Davis
(1982)

Submitted to the Department of
Electrical Engineering and Computer Science
in Partial Fulfillment of the
Requirements for the
Degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1984

Signature of Author
Department of Electrical Engineering and Computer Science
July 13, 1984

Certified by: Dr. P.M. Zurek, Thesis Supervisor

Accepted by: A.C. Smith, Chairman
Departmental Committee on Graduate Students

EVALUATION OF
INFINITE PEAK CLIPPING
AS A MEANS OF AMPLITUDE COMPRESSION

by

KAREN ANN SILLETTO

Submitted to the Department of Electrical Engineering and Computer Science on July 13, 1984 in partial fulfillment of the requirements for the Degree of Master of Science in Electrical Engineering

ABSTRACT

Infinite peak clipping is a simple and effective technique for processing speech prior to transmission over an analog channel with limited dynamic range. However, the technique has yet to be applied seriously to the problem of limited auditory dynamic range accompanying sensorineural hearing loss. In this thesis, an attempt was made to evaluate the potential for infinite peak clipping as a means of amplitude compression. Measurements were made of the compression of the range of speech levels when clipping is followed by post-filtering (as is presumed to occur when clipped speech is analyzed by the ear). Spectral distortions caused by clipping were also assessed.

The widths of level distributions of clipped and filtered speech were found to be relatively independent of the characteristics of both the pre-filter and the post-filter. Measured ranges between the 10% and 90% cumulative levels were about 10-15 dB as compared to the input ranges of 30 to 40 dB.

The clipped spectra of unvoiced speech sounds can be predicted analytically. The clipped spectra of voiced sounds cannot be easily predicted and so several cases were examined empirically. In general, despite the radical distortion of the input time wave, spectral distortions are not severe. Pre-filtering schemes designed to preserve relevant cues in the clipped spectra are discussed. It is concluded that infinite peak clipping as a means of extreme amplitude compression deserves detailed study with hearing-impaired listeners.

Thesis Supervisor: Dr. Patrick M. Zurek
Title: Research Scientist, Research Laboratory of Electronics

ACKNOWLEDGMENT

First, I wish to thank Pat Zurek for his continued guidance and support on this project. I would also like to thank all the members of the Communications Biophysics Group for helping to make the laboratory such an enjoyable place to work. Although I cannot mention all the people who have been wonderfully supportive and encouraging, I would like to thank my office-mates, Dan Leotta, and Jean Reid (not forgetting Mineola Minos), for making the office such a home away from home. Also, special thanks to Dan Leotta, Mike McConnell, and Diane Bustamante for their extra help when I really needed it. Finally, I would like to thank my parents and family for their love, and continued support.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Problem Definition

Sensorineural hearing impairments are characterized by elevated hearing thresholds without a corresponding elevation in the discomfort threshold. This reduction in dynamic range, if severe enough, can necessitate a compromise between speech intelligibility and comfort. Consequently, amplitude compression of speech has been extensively studied as a means of matching the dynamic range of speech and other acoustic signals to the reduced dynamic range of the impaired listener.

The most common means of amplitude compression has been automatic gain control (AGC) in which the gain of an output amplifier is controlled by an estimate of the input amplitude. Results of studies of AGC in hearing aids are reviewed by Braida et al. (1979). While there are clear benefits and applications for such systems, there are also limitations in the hearing-aid application. For instance, due to the "attack time" of an AGC system, the effectiveness in protecting the user from high-level transient sounds is limited. Another limitation is that the actual degree of

compression is less than that specified (DeGennaro et al., 1981).

Infinite peak-clipping is a simple method of amplitude compression that drastically reduces the dynamic range of the speech signal by clipping the input time-wave so that a rectangular wave of constant amplitude results. Only the zero-crossings of the original waveform are preserved. Since the output (envelope) amplitude is constant, infinite peak clipping produces maximal range reduction of the wideband signal.

The obvious cost of this range compression is severe waveform distortion. However, it has been found that, despite this extreme distortion, infinitely-clipped speech can be highly intelligible (Licklider, 1946; Licklider, Bindra, Pollack, 1948). Thus, it would seem that peak clipping should be considered as a means of amplitude compression in hearing aids and also perhaps cochlear implants and tactile aids, since it achieves large compression of the amplitude range of speech with relatively little loss in intelligibility.

Though clipping produces a constant-amplitude wideband signal, the perceived loudness may not be constant because the ear effectively filters incoming sounds into frequency bands (critical bands) whose outputs are believed to be

relevant to loudness (Green and Swets, 1966, Chapter 10). The amplitude range of speech after post-filtering has not yet been studied and clearly needs to be understood in order to evaluate clipping as a means of amplitude compression.

The effects of clipping on important cues for speech intelligibility are also not well understood. In particular, while it is generally accepted that the short-time spectral amplitude pattern of speech is important, changes in this pattern resulting from infinite peak clipping are not understood. It is possible that, even though wideband clipping produces fair intelligibility, multiple narrowband clippers could better preserve spectral amplitude patterns and consequently increase intelligibility.

The general problem addressed in this thesis is how infinite peak clipping might be used to compress the amplitude range of speech for the hearing-impaired. More specifically, the thesis will investigate the influence of pre- and post-filtering on the amplitude distributions and spectral patterns of clipped speech. The range of filtered-clipped-filtered speech will be measured and the spectra of clipped waveforms will be analyzed in order to evaluate the potential for clipping as a means of amplitude compression, as well as to understand past studies which

examined the intelligibility of clipped speech.

1.2 Previous Research

Peak clipping was first introduced in radio transmitters during WWII as a means for maximizing the intelligibility over a transmission channel with limited dynamic range. Licklider (1946) studied different types of amplitude distortion and concluded that peak-clipping was the least detrimental to intelligibility. Licklider's experiments showed that since infinite peak clipping preserves only the zero crossings of the input signal, the pattern of instantaneous amplitudes in the speech waveform is not essential for intelligibility.

Pollack (1952) confirmed Licklider's results and showed that the intelligibility of infinitely-clipped speech, relative to unmodified speech, is a function of SNR (with signals equated in terms of peak amplitude and with noise added after the clipper) and roughly independent of the frequency range of the speech signal. Clipped speech is more intelligible than normal speech at low SNR's because the square speech has more power per unit peak amplitude than normal speech. At high SNR's, distortion introduced by the clipper degrades intelligibility relative to unmodified speech.

Intelligibility of clipped speech is improved by highpass filtering the speech before clipping (Licklider, Bindra, Pollack, 1948; Pollack, 1952; Thomas and Niederjohn, 1970; Thomas and Ravindran, 1971). One interpretation of this finding (Thomas and Niederjohn, 1970) is that a highpass pre-filter with cutoff at 1100 Hz reduces high-amplitude, low-frequency signals and results in better preservation of the second formant, which is very important for intelligibility.

There has been one investigation of clipped speech with hearing-impaired listeners. Thomas and Sparks (1971) compared the intelligibility of filtered-clipped speech with linearly-amplified speech by testing 17 ears of 16 hearing-impaired subjects. For 13 of the 17 cases, the filtered-clipped speech was more intelligible than linearly-amplified speech with the two types of speech equated in terms of their overall SPLs. However, these results are open to the criticism that the linear reference condition may have been less than optimal.

Hildebrant (1982) investigated a multiband clipping system that filtered the incoming speech into frequency regions corresponding to the ranges of the first three formants. The frequency bands were individually clipped, filtered again and then summed together. A reduced dynamic

range was created for normal listeners by adding white noise after processing and restricting the wideband signal to be below an artificial "discomfort threshold". Intelligibility with clipped speech was much better than with linearly-amplified speech. A similar system was studied by Guidarelli (1981), who found that intelligibility was greatest with 4 octave-wide or 6 two-third-octave-wide channels.

The general conclusion that can be drawn from past studies is that infinite peak clipping reduces the dynamic range of wideband speech while maintaining a fair degree of intelligibility. This fundamental result, that the "amplitude information" in a speech waveform can be discarded without a drastic degradation of intelligibility, initially appears quite surprising. However, the key to understanding this result lies in the realization that "amplitude information" is also coded in the zero-crossings of a wave. Although it is not at all obvious from inspection of the wave, the spectra of clipped signals are often very similar to the spectra of unclipped signals.

Licklider, Bindra, and Pollack (1948) clearly emphasized the importance of examining the spectral changes caused by clipping in order to understand the relatively small effect of clipping on intelligibility. Figure 1.1,

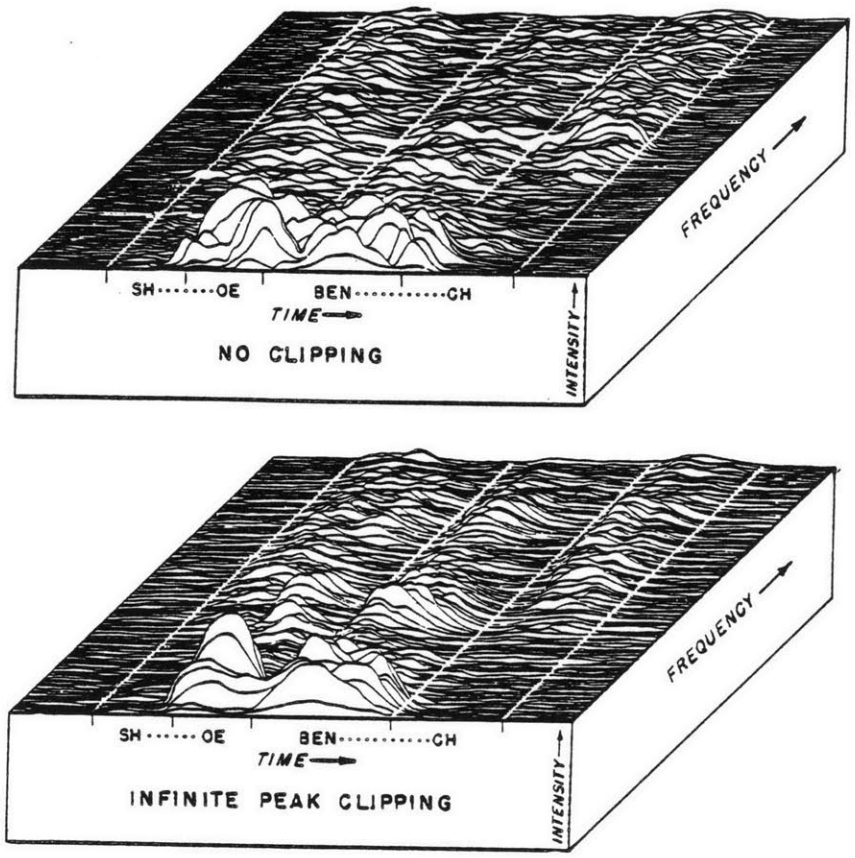


Figure 1.1 A running spectrum of a two-word utterance, before and after infinite peak clipping (taken from Licklider et al., 1948).

taken from their paper, shows a running spectrum (a 3-D plot of amplitude vs. time and frequency) of a two-word utterance before and after infinite peak clipping. From this frequency analysis it is clear why clipping has little effect on intelligibility - the spectral patterns of clipped and unclipped speech are not drastically different.

1.3 Rationale for Present Study

The rationale for studying the effect of pre- and post-filters on the range of clipped speech and the spectral distortions caused by clipping follows from an underlying model and set of assumptions about the operation of the ear and the perception of speech. First, it is generally believed that the ear analyzes incoming speech into frequency bands that are approximately one-third of an octave wide. In the present study, it is assumed that the quantity relevant to dynamic range is the envelope of these bandpass signals. A major portion of this thesis addresses the question: what are the effects of post-filtering on the envelope distributions of infinitely-clipped speech waves?

A second important assumption is that the primary cues to intelligibility lie in the spectral (magnitude) patterns of speech. In particular, it is assumed that local maxima in the spectra (formants) are vital to intelligibility. The effects of clipping on these patterns will be studied as well as ways, such as pre-filtering, in which the spectral distortions can be minimized.

Chapter 2

Methods

A block diagram of the system used in the present study is shown in Figure 2.1. Input signals (speech and other test signals) were either first filtered with a pre-filter and clipped, or they were unmodified. The spectra of the two signals at point 'A' were compared. The dynamic ranges of these signals were assessed by measuring the distributions of the signal envelopes after bandpass filtering.

2.1 Inputs

Harvard sentences (IEEE, 1969) spoken by a male talker were chosen as input signals. It was desired to use a speech sample that had an amplitude distribution that was characteristic of spoken English, but yet was of manageable size for digital processing. The number of sentences was determined by preliminary measurements described below (Section 2.5).

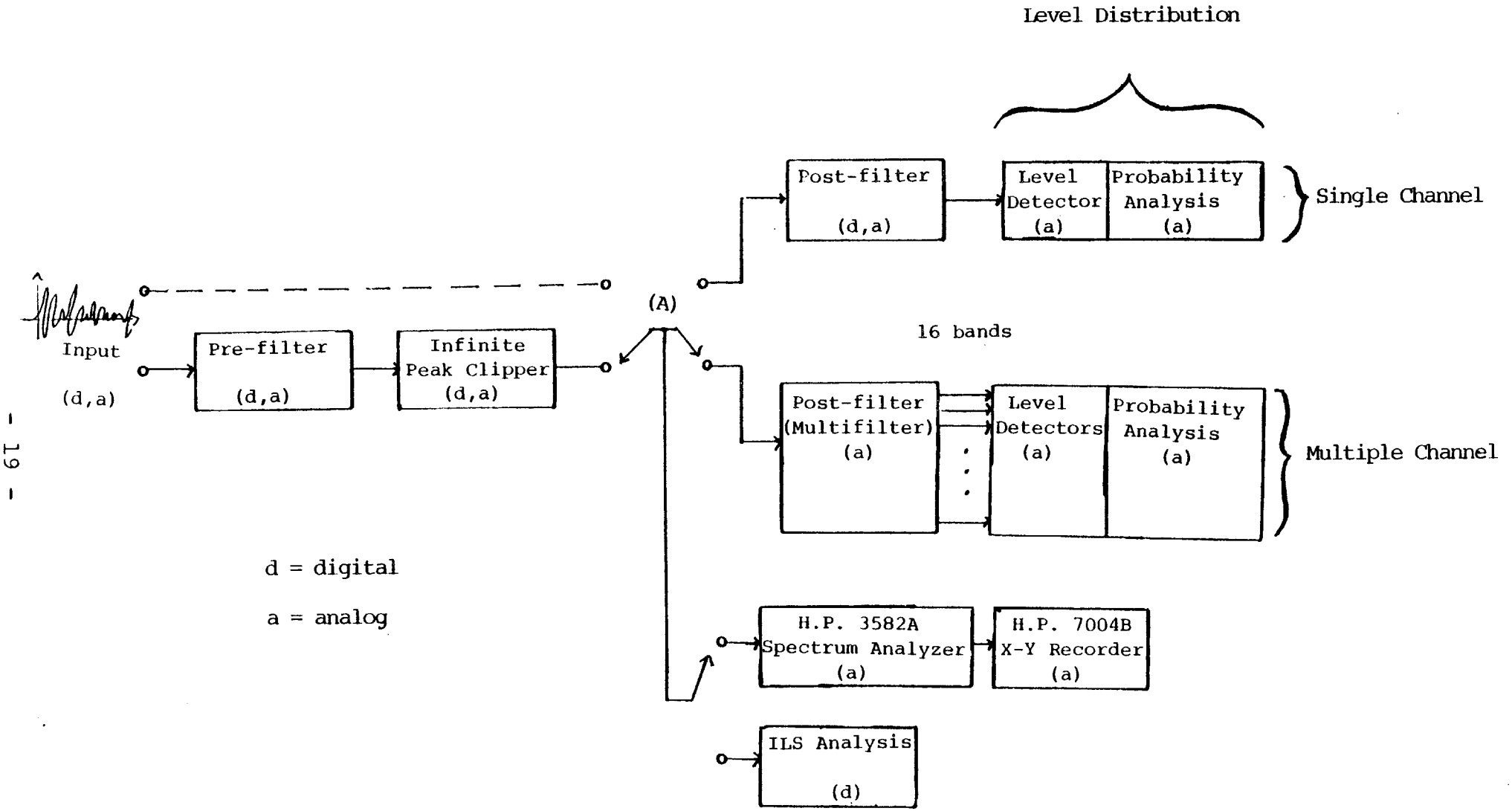


Figure 2.1 System block diagram

2.2 Filtering and Clipping

Many of the operations shown in Figure 2.1 could be implemented either digitally (not in real time), or with analog equipment. The letters 'd' and 'a' in Figure 2.1 indicate digital and analog capability.

2.2.1 Pre-filters

The analog pre-filter was a 2-pole, 1100-Hz highpass filter suggested by Thomas and Niederjohn (1970). For ease of reference, this filter is termed "optimal" because Thomas and Niederjohn found it to give the best intelligibility of a number of pre-filters. Also for ease of reference, the condition in which the signal is clipped without pre-filtering will be termed "allpass" pre-filtering.

A software package, ILS (Interactive Laboratory System) written by Signal Technology Incorporated, was used to design the digital filters and filter waveforms. An infinite impulse response Butterworth design was used for the digital filters. Digital pre-filters were a digital "optimal" filter, an allpass filter (i.e. no pre-filtering), and a single-pole, 6 dB/octave 8000-Hz highpass filter, which will be termed "differentiator" for

obvious reasons. Speech was also digitally pre-filtered into the three formant regions that Hildebrant (1982) described. The first formant filter was a bandpass filter between 200 Hz and 900 Hz, the second formant filter was 900 Hz to 2800 Hz, and the third formant filter was 2800 Hz to 6000 Hz. Finally, narrowband pre-filters with bandwidths less than 231 Hz and center frequency at 1000 Hz were used for the purpose of exploring the effects of the pre-filter bandwidth and slope.

2.2.2 Clippers

Analog clipping was performed by a Schmitt trigger with an adjustable hysteresis "dead zone" (Hildebrant, 1982). The dead zone was adjusted to a minimum value which prevented triggering by internal noise so that no output signal resulted from a null input signal. Digital clipping was performed by keeping only the sign of the digitized waveform values (except zero, which remained zero).

2.2.3 Post-filters

The clipped waveform (at point 'A' in Figure 2.1) could be post-filtered (before detection) into adjacent "critical-band" filters using a General Radio (1925)

multifilter, or digitally post-filtered into a single frequency band. The multifilter contains 30 one-third octave, 6-pole (3-zero), Butterworth filters with center frequencies that range from 25 Hz to 8000 Hz. The filter skirt slopes near the passband are about 36 dB/octave and flatten out to about 18 dB/octave due to the three zeros in the transfer function.

Only the outputs from 19 one-third octave bands between 125 Hz and 8000 Hz are used, since the most important cues for intelligibility lie in this frequency range. Because critical bands are proportionately broader at lower frequencies, outputs from the three lowest of these one-third octave bands are summed to form channel 1, and the outputs from third-octaves centered at 250 Hz and 315 Hz are summed to form channel 2. The remaining channels are the outputs of single third-octave bands. The bandwidths and center frequencies of the 16 frequency channels are listed in Table I.

For greater flexibility in examining the effect of post-filter skirts and bandwidths, single band post-filters were implemented digitally. Post-filter skirts were varied by changing the number of poles in the filter transform from 2 to 8 poles. The bandwidth of the postfilter was varied from 231 Hz to 29 Hz. The filtered bandpass output was

Table I

Channel number	Center frequency	Bandwidth	RC-Filter Time Constant	RC-Filter -3 dB Frequency	R
1	160 Hz	113 Hz	17825.0 usecs	9 Hz	12.50
2	280	130	9390.0	17	7.65
3	400	93	5730.0	28	3.32
4	500	115	4456.0	36	3.19
5	630	146	3390.0	47	3.11
6	800	185	2657.9	60	3.08
7	1000	231	1989.4	80	2.89
8	1250	289	2021.0	79	3.66
9	1600	371	1559.7	102	3.63
10	2000	463	1368.7	117	3.95
11	2500	579	859.4	185	3.12
12	3150	729	859.4	185	3.94
13	4000	926	251.5	531	1.74
14	5000	1158	198.9	795	1.46
15	6300	1459	99.9	1592	0.92
16	8000	1852	39.9	3979	0.47

input into a single analog detector.

2.3 Dynamic Range Analysis

2.3.1 Level Detection

The hardware used to detect the envelope of the signal is part of a multi-channel Automatic Gain Control compression system designed and developed by Coln (1979). Each level detector performs a short-term rms calculation in which the bandpass signal is first squared and then filtered by an RC-lowpass filter. The output of one of these detectors is an estimate of the logarithm of the envelope squared. There are 16 level detectors, one for each of the 16 bandpass outputs of the multifilter. When a single-band digital post-filter is employed, the output of the filter (after D/A conversion) is presented directly to the level detector with an 80 Hz RC-filter (channel 7).

Each RC-lowpass filter time constant can be adjusted by varying the filter capacitor. In channels 1 through 12, the time constants were adjusted to keep the second-harmonic ripple down to 6 percent at twice the center frequency of the signal. However, in the two lowest channels the bandwidth is larger than one-third octave, therefore a

greater percentage of the signal envelope will be filtered, resulting in a "smoother" signal envelope. In channels 13 through 16, the RC-filter bandwidth is larger in proportion to the input signal center frequency than in the other channels, so more second harmonic ripple will be included in the estimate of the envelope. The filter time-constants are listed in Table I.

2.3.2 Probability Distribution Analysis

Samples from the 16 level detector outputs were digitized, and "bins" corresponding to the decibel level and channel of each sample were incremented. The signal in each band was sampled at a rate of 400 samples/second, giving a total sampling rate of 6400 samples/second. The binwidth was 1/2 dB and the dynamic range, which was determined by the level detector, was 80 dB in each band.

Normalized histograms of the decibel level of the signal envelope are plotted for each of the 16 frequency channels. The distributions are normalized to the largest histogram value in the 16 channels. The rms level of each distribution is calculated along with the 10% and 90% amplitude levels. The 10% level is the amplitude level which only 10% of the speech samples exceed; 90% of the samples have amplitudes greater than the 90% point.

Henceforth in this thesis, the term "range" applied to a level distribution will mean the difference between the 10% and 90% levels.

Samples of silence or low-level noise may corrupt the distributions. In cases where the signal and noise distributions are clearly separated, the noise samples can be simply deleted. The modified file will produce a more accurate signal amplitude distribution. Such distributions will be labeled "edited" in the following sections.

2.4 Spectral Analysis

Spectral analysis was performed using two methods. The H.P. 3528A spectrum analyzer computed and averaged the spectral magnitudes of a number of time samples. Spectral graphs were plotted with the H.P. 7004B X-Y recorder. The ILS software package was also used to compute and plot the magnitude spectrum of a single time sample. Both methods used the Fast Fourier Transform (FFT) algorithm.

2.4.1 H.P. Spectrum Analyzer

The H.P. 3582A spectrum analyzer transforms a finite segment of discrete time data into a discrete frequency spectrum using an FFT implementation. In single-channel

mode, 1024 samples of the input signal are transformed by the FFT into 256 complex values. The duration, T, of a time frame is determined by the selected frequency spans and the number of samples,

T = 1024/(4*frequency span) = 256/frequency span.

For example, to compute the spectrum over a 5000 Hz span, a 51.2 msec time segment is sampled. The discrete frequency points are spaced at equal intervals, $\nabla F=1/T$. In this example frequency resolution is 19.5 Hz.

The H.P. analyzer also allows for up to 256 spectra to be averaged to give a better estimate of the spectrum of a random signal. Time samples were weighted with a Hanning window in order to minimize both the amplitude and frequency uncertainty in the spectrum.

2.4.2 ILS Spectral Analysis

The spectrum of a finite digital wave can be computed and graphed with an ILS routine which computes a 1024-point FFT. The 1024-point time sample was weighted with a Hamming window in order to minimize the frequency and amplitude distortion. Since only 1024 points are used, static time waves which repeat periodically in less than 1024 sample points (such as static vowels) are appropriate signals for this analysis.

2.5 Control Measurements and System Evaluation

2.5.1 Evaluation of the Level Distribution Measurement System

In order to assess the performance of the level detector system, ranges of signals with known distributions were measured. Analog narrowband random noise and tones were input directly into the level detector in channel 7, which has an RC-filter cutoff of 80 Hz. Narrowband random noise was used as a test signal because its envelope is known to follow the Rayleigh probability density. The narrowband noise was centered at 1000 Hz, and the bandwidth of the noise ranged from 3 Hz to 231 Hz.

The measured range for a 10-Hz band of noise was 11 db, 2 dB smaller than the expected range of 13.4 dB for a Rayleigh-distributed variable. The measured range for a 10-Hz tone was 13 dB, 3 dB smaller than the expected range of 16 dB for a tone.

The discrepancy between the measured and expected ranges of the envelope distributions can be attributed to the influence of the RC-lowpass filter. Calculated probability densities for narrowband noise and tones after squaring and lowpass filtering are given in Appendix A. The

calculated and measured ranges for noisebands and tones are graphed as a function of R, the ratio of input signal bandwidth (or frequency) to the RC-filter bandwidth in Figures 2.2 and 2.3, respectively. It is evident from these figures that the measured ranges follow the calculated ranges to within about 1 dB. As the noise bandwidth or the tone frequency increases, the difference between the input envelope range (dashed line) and the lowpass-filtered envelope range increases. If the RC-filter bandwidth is much larger than the noise bandwidth (or tone frequency), and if the second-harmonic ripple can be neglected, then the effect of lowpass filtering will be small.

Residual deviations between the calculated and measured ranges may be attributable to sampling errors or to circuit imperfections. A possible reason for the measured range being slightly greater than the calculated range for noise with bandwidths larger than approximately 80 Hz is that the analysis did not include the effects of second harmonic ripple. As the noise bandwidth increases, the signal envelope will include more second-harmonic ripple which will increase the measured level range.

It is expected that the decibel underestimation of the input envelope range will be the same for signals, such as speech, with level ranges that are much wider than those of

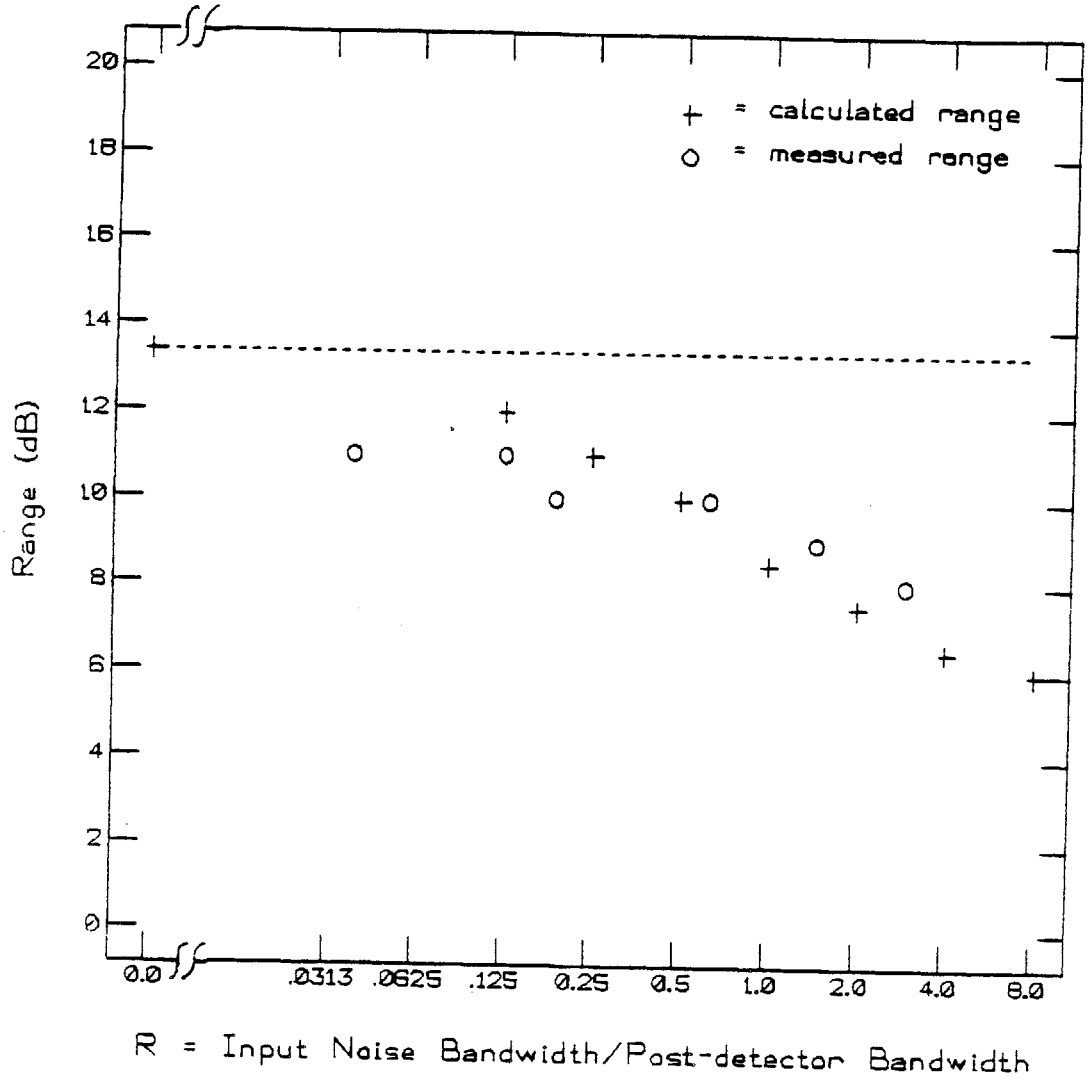


Figure 2.2 Measured and calculated ranges for narrowband noise (centered at 1000 Hz) after squaring and lowpass filtering with an RC-lowpass filter. Range is graphed as a function of the input noise bandwidth divided by the RC-filter bandwidth. Dashed line represents the expected range for a Rayleigh distributed variable.

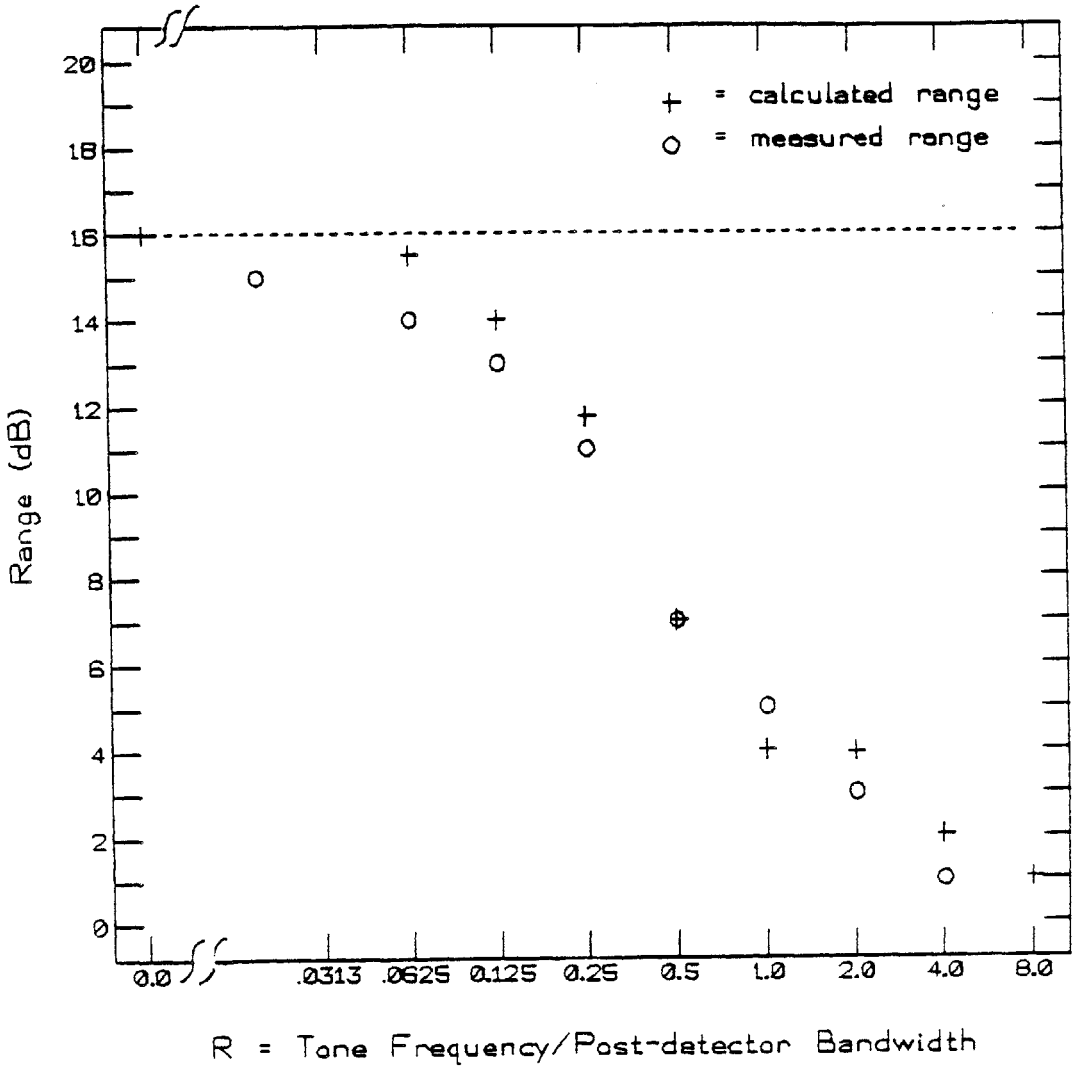


Figure 2.3 Measured and calculated ranges for tones after squaring and lowpass filtering with an RC-lowpass filter. Range is graphed as a function of the tone frequency divided by the RC-filter bandwidth.

noise or tones. If we assume that such wide distributions result from a large but slow modulation of a bandpass process, then the resulting level distribution should be the convolution of the distributions of the modulation and the bandpass envelope. If the modulation is sufficiently slow, its variation will be unaffected by the lowpass filter. Thus, for example, if there is a 5-dB underestimation of the input 13.4 dB range of narrowband noise, the underestimation will still be about 5 dB if the noise is slowly modulated over a range of, for instance, 40 dB.

2.5.2 Determination of Sample Size for Sentences

The range of a bandpass speech signal estimated with an input of n sentences should converge to a constant value as n is increased. The sample size used in the main study was chosen by preliminary measurements in which the number of sentences in the sample was increased until there was only a small change in the range for an increase in the sample size.

The level ranges of single sentences were measured after digitally post-filtering by a one-third octave (231 Hz) filter centered at 1000 Hz. The ranges were found to

vary from 25 dB to 46 dB among a sample of 10 sentences (Figure 2.4). Since this range, for 10 single sentences, varied 21 dB, a larger sample of sentences was considered.

The sample size was increased to four, ten, or twenty sentences and the range was measured (Figure 2.4). (The 10-sentence sample contained the original 4 sentences, and the 20-sentence sample contained those same 10 sentences.) It appeared that a sample size of 10 or 20 sentences would be appropriate since the measured ranges, 36 dB and 38 dB, were close to the results obtained by Dunn and White (1939) for one-half octave wide distributions above 500 Hz, and by Krieg (1980) and DeGennaro et al. (1981) for a third-octave distribution.

The measured ranges in 13 frequency channels for an input sample of the 10 and 20 sentences were compared (Figure 2.5). In order to examine all channels, the sentences were bandpass filtered with the analog one-third octave multifilter. (The ranges in channels 14, 15, and 16 were omitted because the input was filtered at 4.5 kHz). In all 13 frequency channels the ranges differed by at most 3 dB. Since the difference was small, 10 Harvard sentences were chosen for the sample data. The 10 sentences chosen are listed in Appendix B. The sentences are aproximately 3 seconds long, so 30 seconds of data at 400 samples/second,

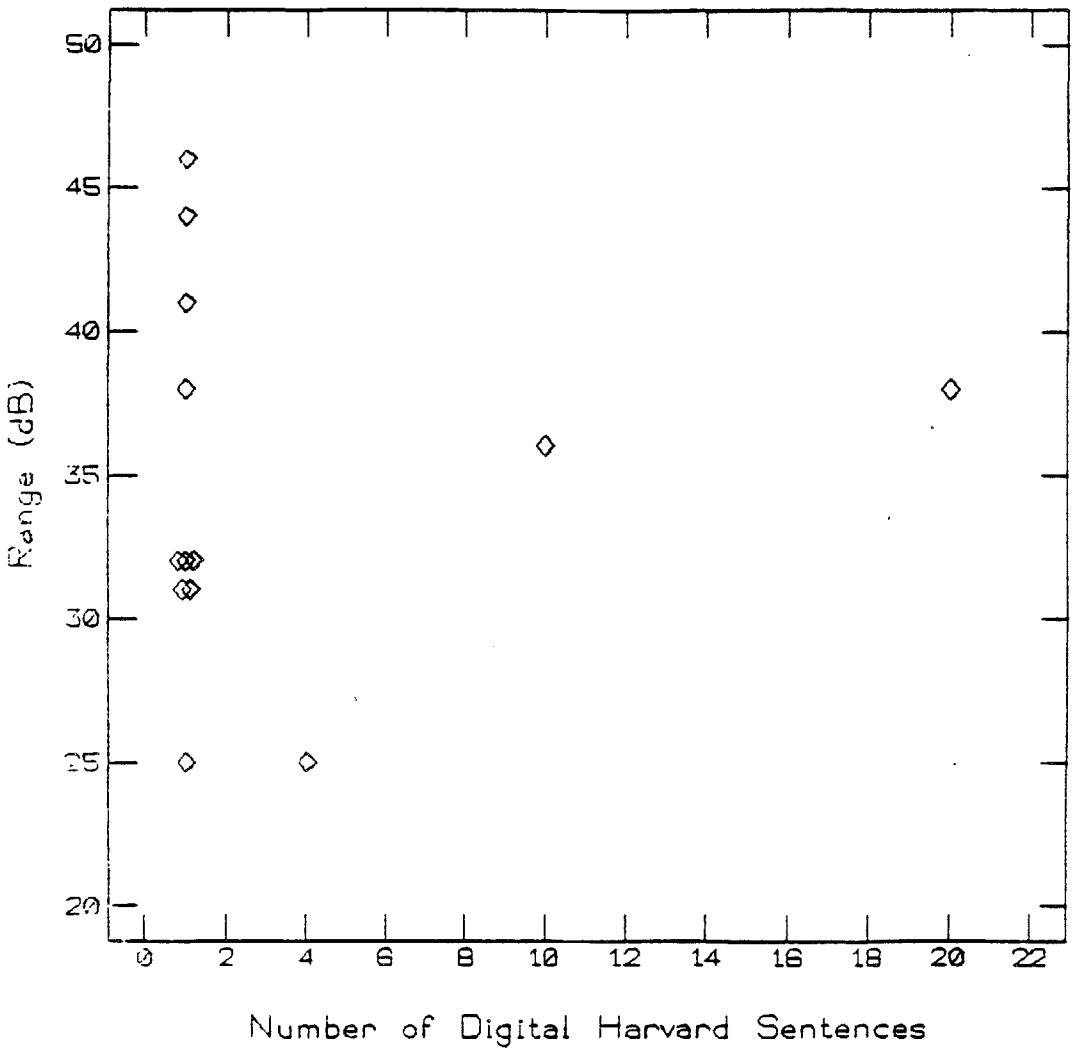


Figure 2.4 Ranges in the one-third octave band centered at 1000 Hz as a function of sentence sample size.

or 12000 samples of data are included in each histogram from which range is estimated.

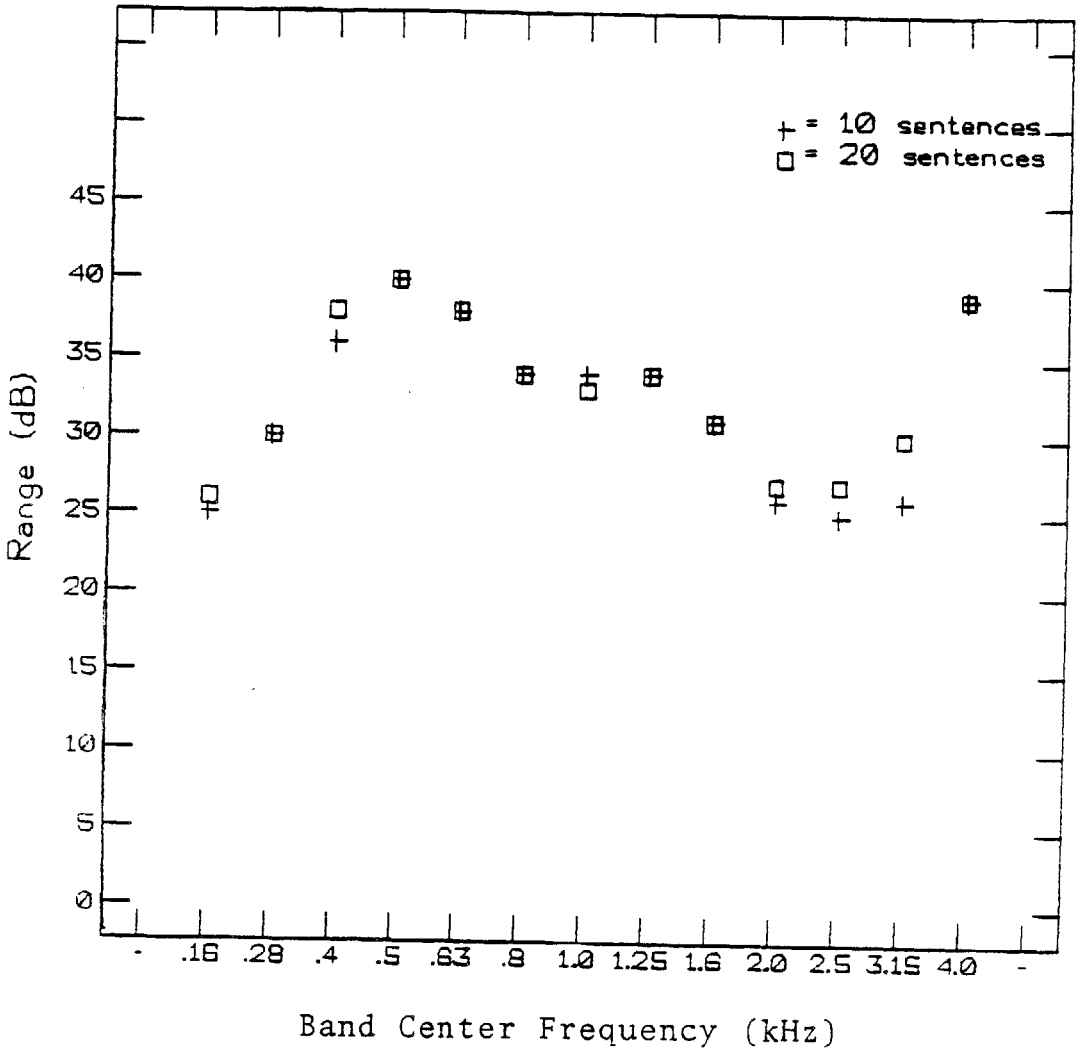


Figure 2.5 Ranges across 13 frequency channels comparing inputs of 10 and 20 digital Harvard sentences.

Chapter 3

Results

3.1 Comparison of Distributions in 16 Frequency Channels

The histograms of one-third octave bands of clipped and unmodified speech were investigated in the following experiments. First, the 10 digital sentences were clipped (pre-filter was an allpass filter) and post-filtered by the multifilter; second, the sentences were pre-filtered with an 1100-Hz highpass filter (analog), clipped, and post-filtered by the multifilter. The reference condition of speech filtered only by the multifilter was also investigated.

The amplitude ranges, shown in Figure 3.1, exhibit a dependence on band center frequency. Part of this dependence may be attributed to the variation across bands of the quantity R, the ratio of input bandwidth to the detector's RC-filter bandwidth. From the analysis presented in Section 2.5.1 (see Figure 2.3), we expect a smaller range in the channels centered at 160 Hz and 280 Hz, because R is larger than in the higher frequency channels (see Table I). Conversely, there should be relatively less smoothing of the envelope in channel 13 (centered at 4000 Hz) since the bandwidth of the RC-filter is larger in proportion to the

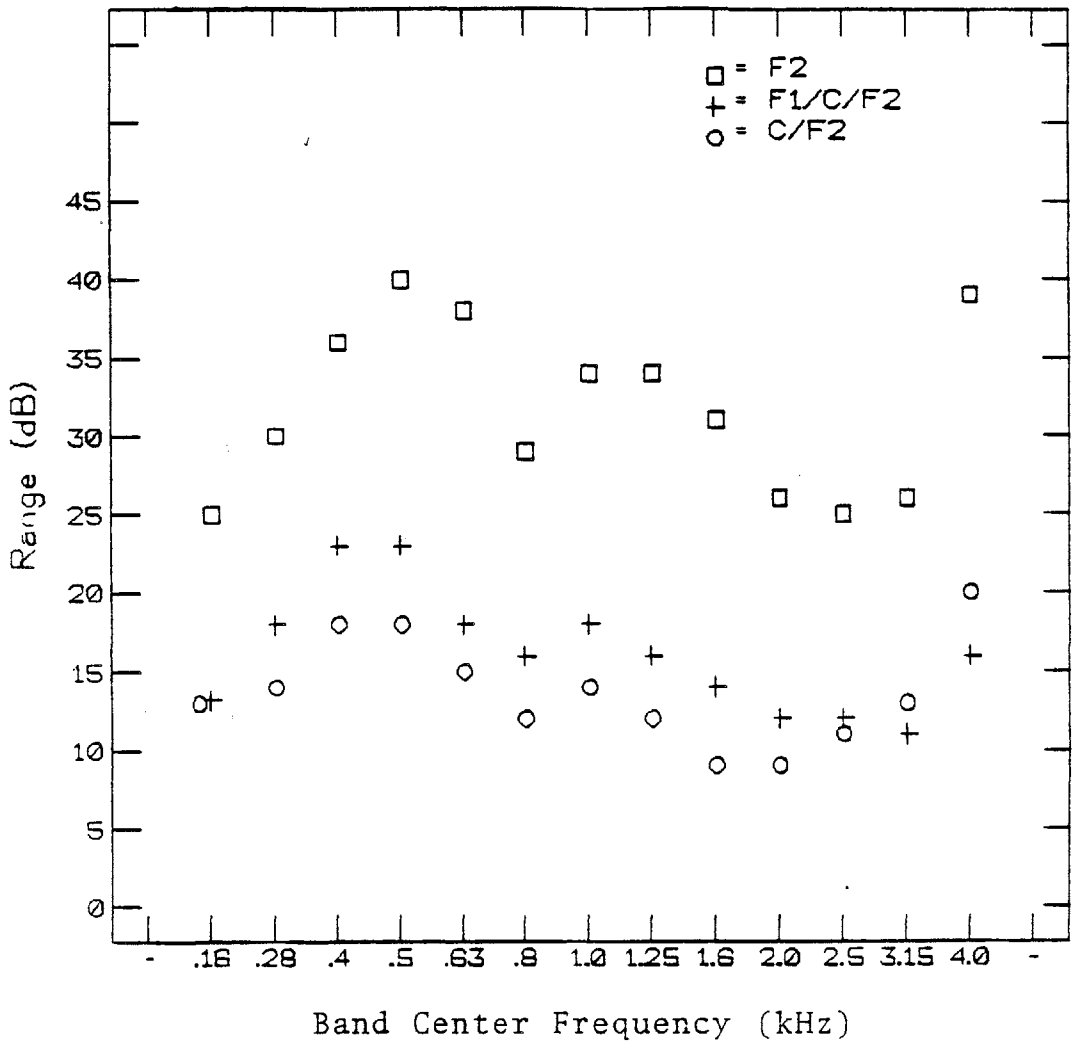


Figure 3.1 Ranges across 13 frequency bands for clipping and nonclipping conditions. Input is 10 digital Harvard sentences.

F1 = 1100 Hz highpass filter (analog)
C = Infinite peak clipper (analog)
F2 = Multifilter

signal bandwidth (R is smaller) than in the lower-frequency channels. There will also be more second-harmonic ripple in channel 13 which will tend to increase the range. (Note that results for only the first 13 channels are plotted because the digital signals were lowpass filtered at 4.5 kHz, and the three highest channels are centered at frequencies above that cutoff).

The amplitude range for speech filtered by the multifilter varies at most 15 dB across all frequency channels and averages 32 dB. The ranges agree reasonably well with the results of Krieg (1980), who used the present system, as well as with Dunn and White (1940) who used a different measurement procedure (peak pressures in 1/8 second intervals). The present measurements are slightly smaller than those reported by DeGennaro et al. (1981), who also used the same system as used here for measuring bandpass envelopes. It should be noted that while DeGennaro et al. (1981) display a 50-dB range for one talker and a single band (Figure 4 in their paper), their subsequent figures (5, 6, and 7) for at least two talkers and all sixteen bands show nearly all ranges to be between 30 and 40 dB.

Clipping reduces the range of speech by about 15 to 20 dB across all frequency bands. The range of bandpass speech that was pre-filtered with an 1100-Hz highpass filter and clipped varies 12 dB across frequency bands and averages 15.8 dB. The range of clipped speech with no pre-filtering varies by 11 dB and averages 9.5 dB across bands. This difference between highpass pre-filtering and no pre-filtering will be discussed further in Section 3.3.1.

In subsequent experiments in which a single channel is analyzed, the band centered at 1000 Hz is used. This band was chosen as representative because it is approximately in the middle of the speech frequency range, and also because the measured ranges in this band are near the respective averages across bands for the three conditions in Figure 3.1.

3.2 Comparison of Digital and Analog Implementations

3.2.1 Digital vs. Analog Sentences

In order to identify any differences in the measurements due to digital or analog sentence inputs, histograms made with tape-recorded analog sentences and digital sentences were compared. Thirty tape-recorded analog Harvard sentences and 10 digital Harvard sentences (after D/A conversion and lowpass filtering at 4.5 kHz) were processed identically with

the analog equipment. Ranges were compared for bandpass filtered speech and clipped/filtered speech. In the clipping condition, these sentences were pre-filtered with the 1100-Hz highpass filter, infinitely peak-clipped, and bandpass filtered into the sixteen channels by the multifilter. The analog sentence histograms were "edited" (as described in Section 2.3.2) to eliminate obvious noise. No editing was neccessary for the digital sentence histograms. The ranges are shown in Figure 3.2.

Except for discrepancies at 800 and 2500 Hz, the difference between the ranges for digital and analog sentences in the nonclipped condition is less than 5 dB. (Note that the three bands above 4.5 kHz are included for the analog sentences.) With the exception of channel 1, a larger range was measured with analog sentences than with digital sentences.

In the clipping condition, the difference between the ranges for digital and analog sentences is 2 to 3 dB. The smaller discrepancy here may be attributed to the clipper's "dead-zone" which was adjusted so that there was no output produced by tape hiss (when the speech was off). Because the distributions of the unclipped analog sentences are wide (30 to 40 dB), it is more likely for the speech and noise distributions to overlap in this condition than with clipped

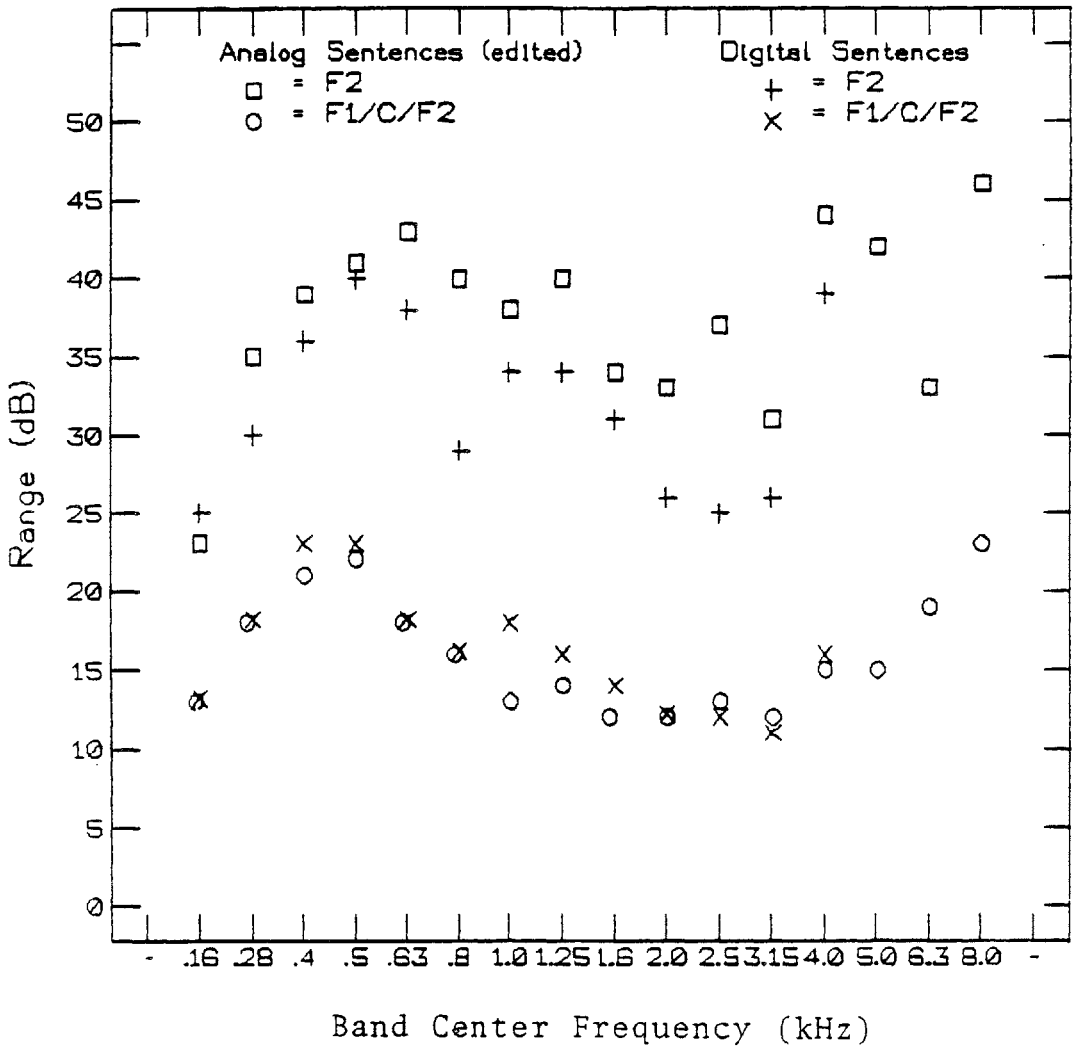


Figure 3.2 Ranges across 13 or 16 frequency bands comparing digital and analog Harvard sentences as inputs.

F1 = 1100 Hz highpass filter (analog)
C = Infinite peak clipper (analog)
F2 = Multifilter

speech. Under clipping, the mixing of noise and speech distributions is reduced because the range of speech is reduced.

It appears from these measurements that digital sentences are superior to analog sentences in terms of signal-to-noise ratio. However, for the clipping condition, analog and digital sentences produce similar results.

3.2.2 Digital vs. Analog Processing

In order to investigate the effects of a number of different pre- and post-filters, digital filters were designed. Histograms after digital and analog processing were compared. The 10 digital sentences listed in Appendix B were used as test inputs. The analog multifilter, 1100 Hz highpass filter, and clipper were used for analog processing. Counterpart digital filters were designed for digital processing. The comparison was made on the one-third octave distribution centered at 1000 Hz (channel 7).

Digital and analog processing were compared with three different pre-filters (an allpass filter, an 1100-Hz highpass filter, and a one-third octave (231 Hz) filter centered at 1000 Hz) before clipping as well as for the reference condition of no clipping.

Table II

A comparison of ranges in channel 7 (bandwidth=231 Hz, center frequency=1000 Hz) after digital and analog processing of 10 digital sentences.

F2 = 1/3 octave (231-Hz bandwidth) filter centered at 1000 Hz
F1 = 1100-Hz highpass pre-filter
C = Infinite peak clipper

	F2	C/F2	F1/C/F2	F2/C/F2
Digitally processed	36	11	13	6
Analog processed	34	14	18	7

Table II shows that the ranges differ between digital and analog processing by at most 5 dB. An interesting fact is that pre-filtering the speech with a one-third octave filter decreases the range from 36 to 6 dB. This will be discussed in Section 3.3.1.

3.3 Compression Results

3.3.1 Pre-filtering

The following experiments were conducted to determine the effect of various pre-filters on the range of clipped speech. The pre-filters used here were those employed in various studies of clipped speech. The first is an allpass filter (i.e. no pre-filtering) as used by Licklider (1948), for example. The second is an RC highpass filter with cutoff at 8000 Hz (subsequently referred to as a differentiator) as used by Licklider and Pollack (1948) and Thomas and Niederjohn (1968). The third is an 1100-Hz (12 dB/octave) highpass filter (subsequently referred to as the "optimal" filter) suggested by Thomas and Niederjohn (1968). A fourth pre-filter is the set of pre-filters described by Hildebrant (1982) that was designed to separate the formant regions of speech. The first-formant filter is a 200 Hz to 900 Hz, 4-pole Butterworth bandpass filter (spanning channels 1

through 6), the second-formant filter is a 900 Hz to 2800 Hz, 10-pole Butterworth bandpass filter (spanning channels 7 through 11), and the third-formant filter covers the range from 2800 Hz to 6000 Hz (channels 12 and 13) and is also a 10-pole Butterworth filter.

The ranges in the thirteen channels under the various configurations are shown in Figure 3.3. Clipped speech with no pre-filtering has the lowest overall range, from 9 dB to 20 dB and averaging 13.7 dB across frequency bands. Differentiating the sentences before clipping increases the range relative to no pre-filtering by about 6 dB across all bands; these ranges vary from 13 dB to 27 dB and average 19.8 dB. Pre-filtering with an optimal filter does not increase the range quite as much as differentiating; the range is 9 dB to 22 dB across all bands and averages 15.5 dB. In the low-frequency bands, speech pre-filtered with the first formant filter has the smallest range of the four types of pre-filtering (averaging 12.8 dB in the first 6 channels). The second and third formant filters produce an effect similar to the differentiator and optimal filters. The average ranges in these two formant filter regions are 16.6 dB in channels 7 through 11 and 14.5 dB in channels 12 and 13.

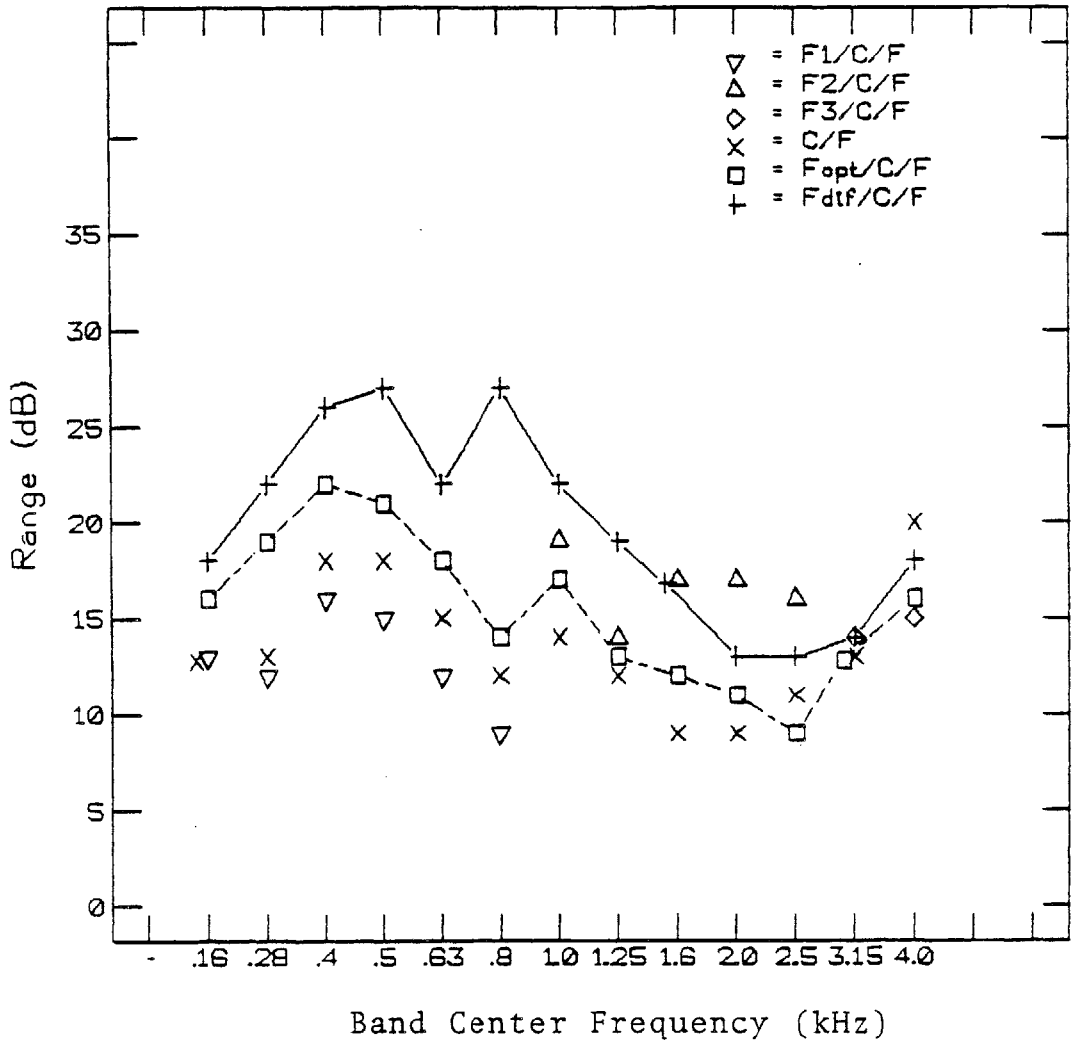


Figure 3.3 Effect of different pre-filters on ranges in 13 frequency bands.

F1 = 200-900 Hz bandpass digital filter
F2 = 900-2800 Hz bandpass digital filter
F3 = 2800-6000 Hz bandpass digital filter
Fopt = 1100 Hz highpass digital filter
Fdif = 8000 Hz highpass digital filter
C = Infinite peak clipper
F = Multifilter

The effects of varying pre-filter bandwidth and rejection slope were investigated in the following experiments. The sentences were digitally pre-filtered, clipped, and digitally post-filtered with a one-third octave bandpass filter centered at 1000 Hz, which had the same filter-skirt slope as the pre-filter also centered at 1000 Hz. The pre-filter bandwidth varied from 231 Hz to 29 Hz and the slope of the pre-filter and post-filter skirt were varied together by changing the Butterworth filter transfer function from 2-pole to 8-pole. The signal level was detected in channel 7 (RC-filter bandwidth is 80 Hz) after bypassing the multifilter.

Plots comparing the ranges of speech pre-filtered with the narrow-band filters to speech pre-filtered with the allpass, optimal, and differentiator filters are shown in Figure 3.4. Except for the case of a 2-pole Butterworth filter, the range for clipped speech that was pre-filtered into a one-third octave (231 Hz) band is 5-8 dB smaller than speech that was pre-filtered with the differentiator or optimal filters. Generally, as the bandwidth of the pre-filter decreases, the range increases. Given that the pre-filter transform has more than 2 poles, varying the pre-filter slope does not change the ranges appreciably.

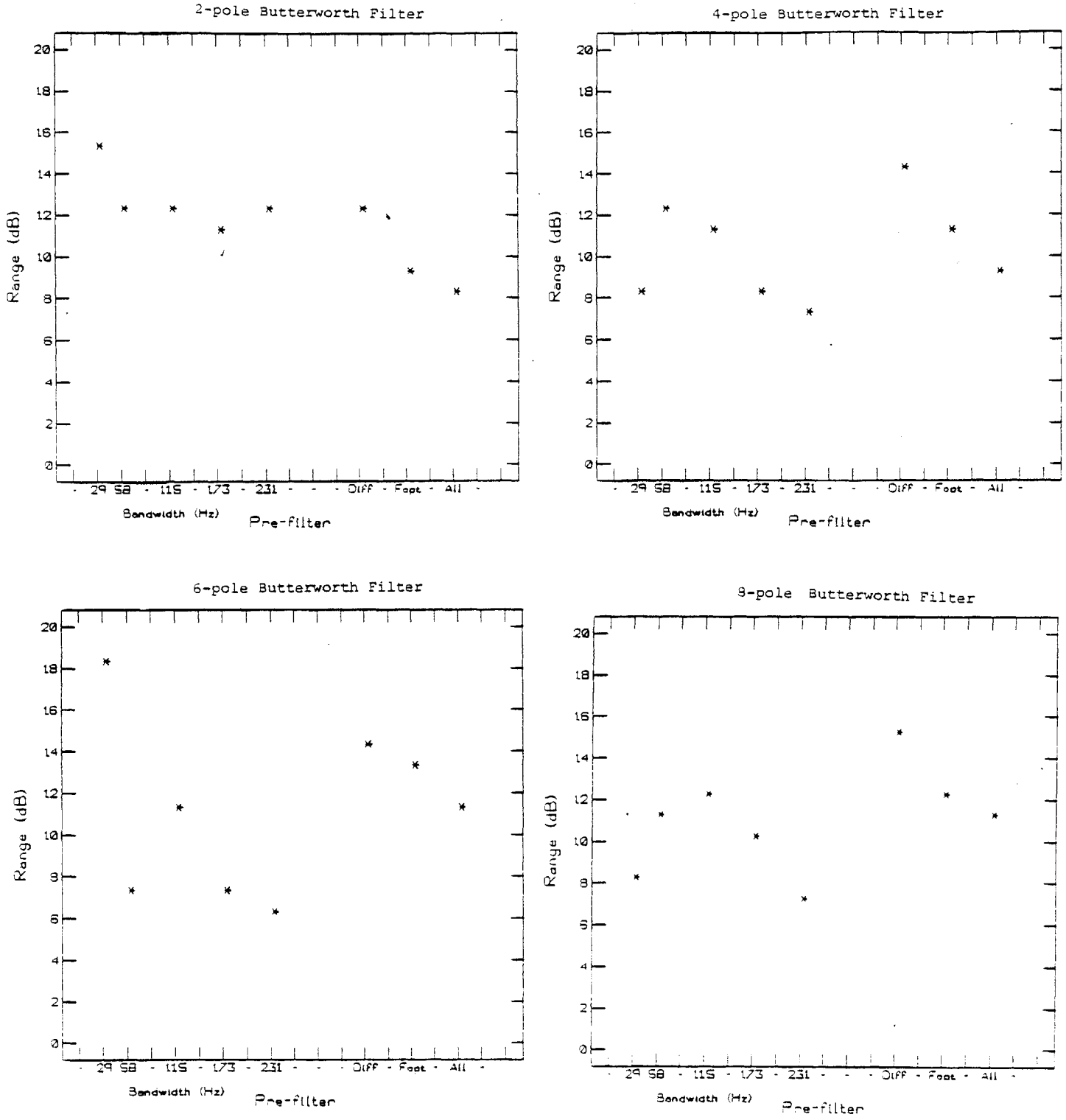


Figure 3.4 Effect of different narrowband pre-filters on ranges. Also plotted are the ranges with the differentiator (Diff), the 1100-Hz highpass filter (Fopt), and the allpass filter (All). Post-filter is 1/3-octave centered at 1000 Hz. Pre-filter and post-filter slopes are the same. For the cases of 'Diff', 'Fopt', and 'All' these slopes apply only to the post-filter.

The pre-filters with a 2-pole transfer function will produce an effect on the low frequency part of the spectrum similar to the differentiator. Since the speech spectrum is dominated by low-frequency components, the ranges are similar (upper left pannel of Figure 3.4). Other discrepancies from the general trends appear for speech that is pre-filtered with a 29 Hz, 6-pole Butterworth filter.

Clipping clearly compresses the amplitude range of speech. Digitally pre-filtering and clipping produces output ranges of 8 to 16 dB in the 1000-Hz frequency band, a reduction of more than 20 dB (Figure 3.4). Across all frequency bands the range is slightly larger, from 10 to 23 dB (Figure 3.3); that is, clipping decreases the range by 15 to 20 dB across all frequency bands.

3.3.2 Post-filtering

The effect of post-filtering the clipped speech was determined by varying the bandwidths and slopes of the post-filter. Post-filters were centered at 1000 Hz and the bandwidth was decreased from one-third octave (231 Hz) to 1/24 octave (29 Hz). The slopes of the post-filter bandpass skirts were increased by varying the Butterworth filter transfer function from 2-pole to 8-pole. Variations in

post-filtering were examined with speech that was pre-filtered with either the allpass filter, the optimal filter, or the differentiator. With the differentiator as pre-filter, only the effect of varying postfilter slopes was examined; the post-filter bandwidth was fixed at one-third octave.

Figures 3.5 and 3.6 show that for all pre-filters examined, variations in post-filter bandwidth and skirt slopes have relatively small effects on compression. Figure 3.5 presents range results with no pre-filter (i.e. an allpass filter) as a function of post-filter bandwidth and skirt slopes. The range of the clipped speech varies from 9-15 dB, with the exception of the measurement with the 29 Hz bandwidth and 8-pole Butterworth filter. Results with optimal and differentiator pre-filtering, shown in Figure 3.6, exhibit a similar spread of ranges.

3.4 Spectral Modifications

Speech sounds can be classified into two categories determined by the presence or absence of voicing. This distinction is important in the analysis of the spectra of clipped speech sounds. The spectra of the two types of speech sounds are analyzed in the following sections.

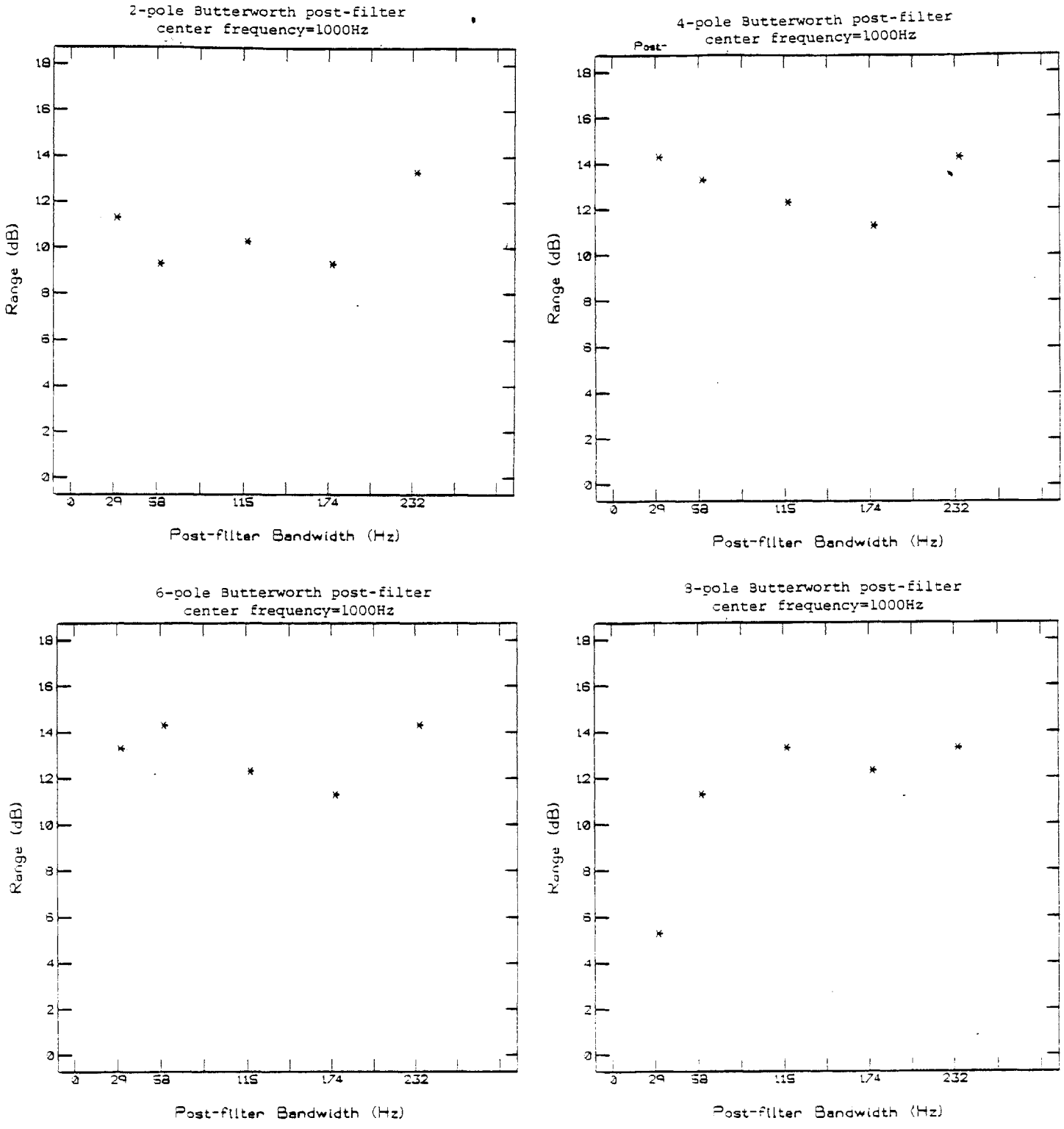


Figure 3.5 Effect of varying the post-filter bandwidth and slopes. Pre-filter is an allpass filter.

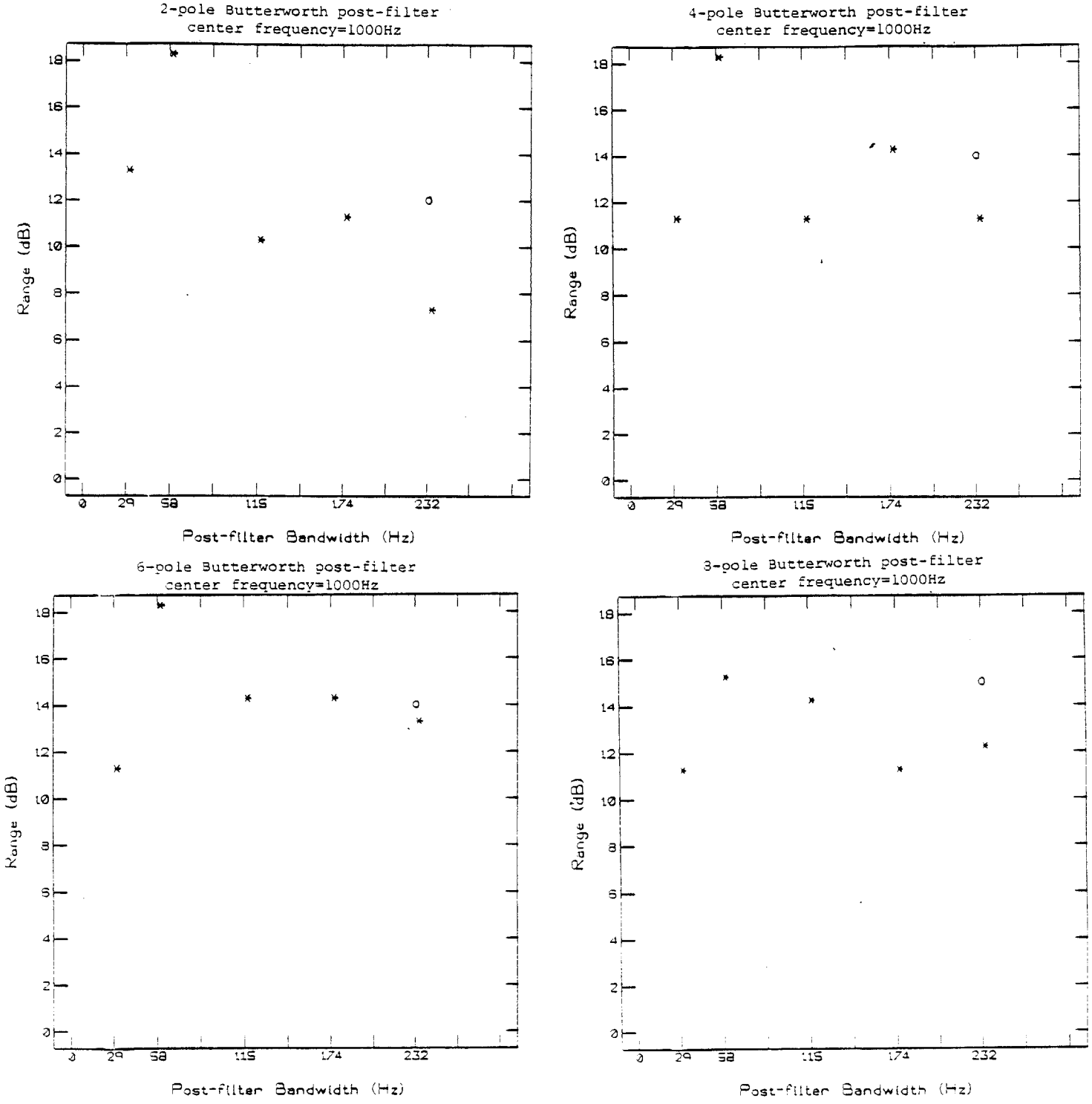


Figure 3.6 Effect of varying the post-filter bandwidth and slopes. Pre-filter is an optimal filter (*) or differentiator (o).

3.4.1 Spectra of Clipped Unvoiced Speech

It has been shown (Papoulis, 1965; Fawe, 1966) that for gaussian noise as input, the output power spectrum Gy(w) after infinite clipping can by determined from the input power spectrum Gx(w) by the "arcsin law" relating their respective autocorrelation functions, Ry(t) and Rx(t). Thus,

$$Ry(t)/Ry(0) = 2/\pi \arcsin[Rx(t)/Rx(0)]$$

This relation was verified by measuring the spectra of various noises before and after clipping. The first set of noises consisted of synthesized tokens of the consonants /f,sh,s/ and the burst portions of /p,t,k/. Clipping was done by the analog clipper and spectra were analyzed with the spectrum analyzer. The power spectrum of each synthetic consonant was also calculated from the filter transfer function from which it was generated, and the arcsin law was applied to these spectra using the FFT.

Figures 3.7 through 3.12 show both the measured spectra and the spectra calculated from the arcsin law. (Since the noise bursts were generated with a sampling rate of 10 kHz, only a 5 kHz range is shown). From the figures it is clear that the clipped spectra follow the arcsin law.

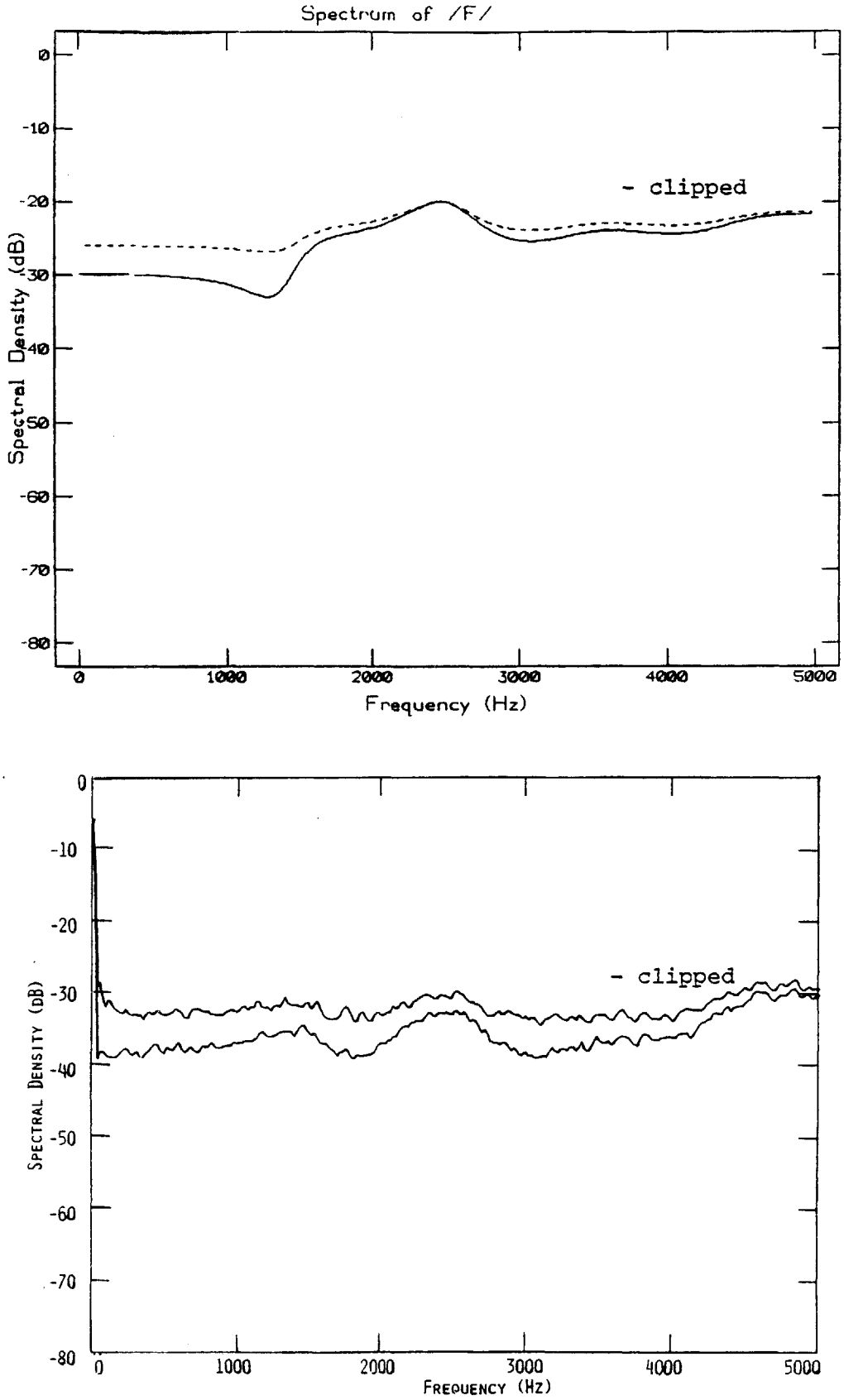
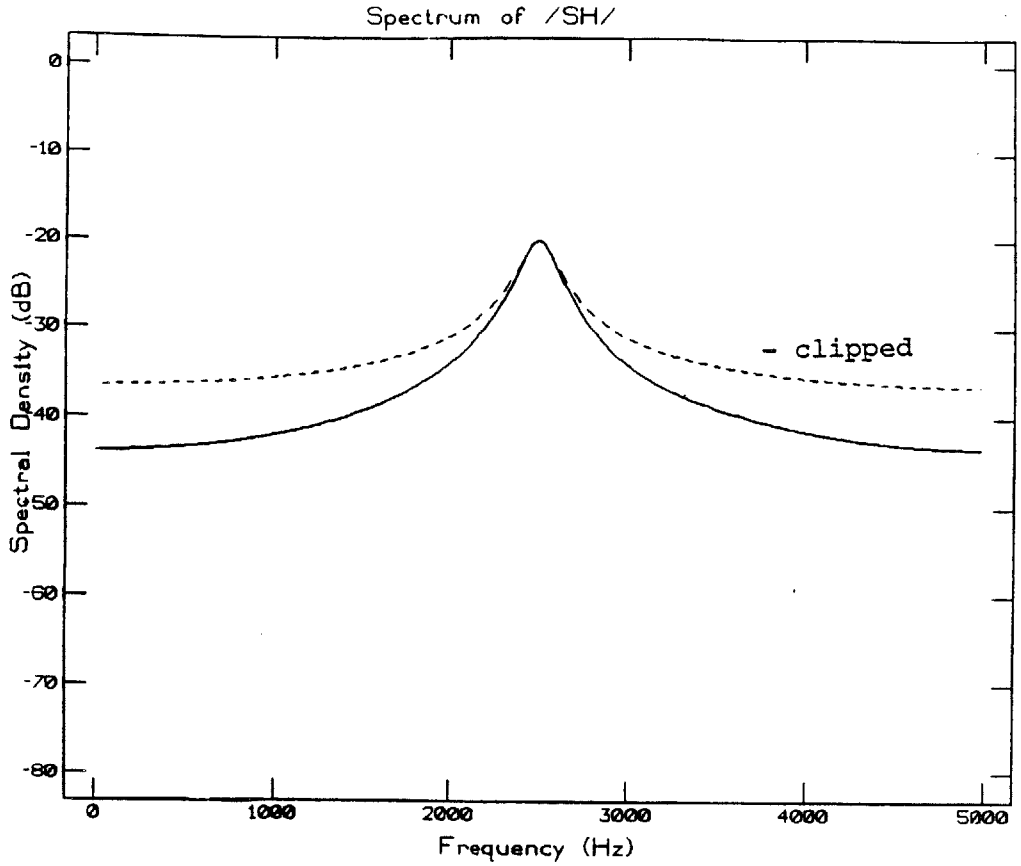


Figure 3.7 Spectra of the synthetic consonant noise burst, /f/, measured before and after clipping. In the top plot, the solid line is the ideal input spectrum and the dashed line is the spectrum calculated from the arcsin law. In the bottom plot, the lower trace is the measured input spectrum and the upper trace is the measured clipped spectrum.



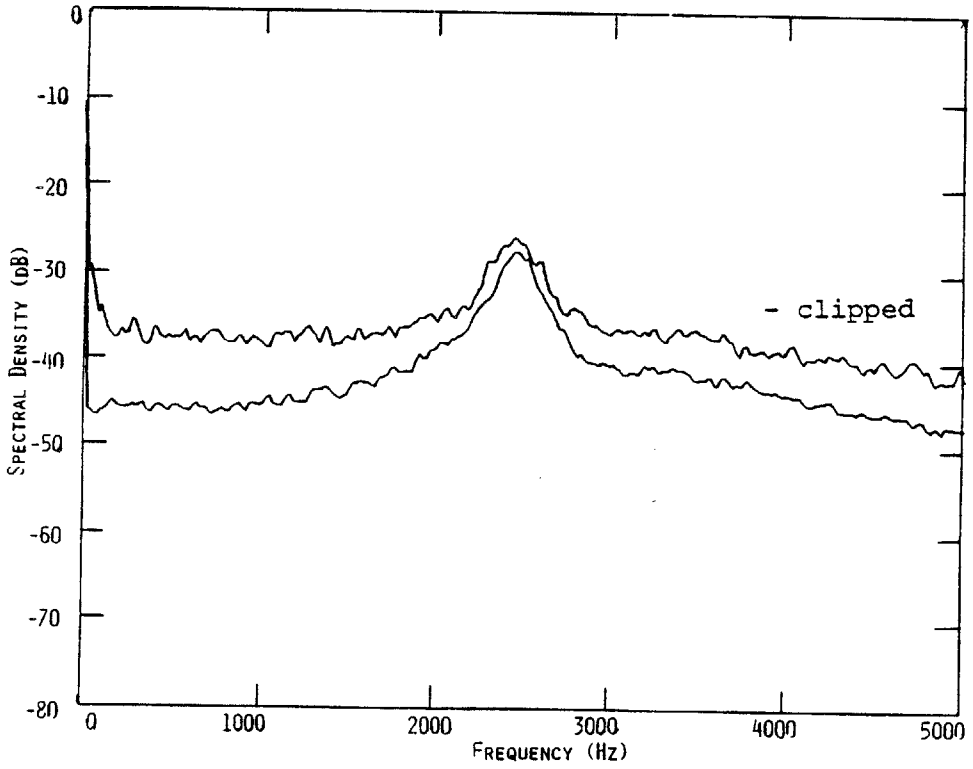
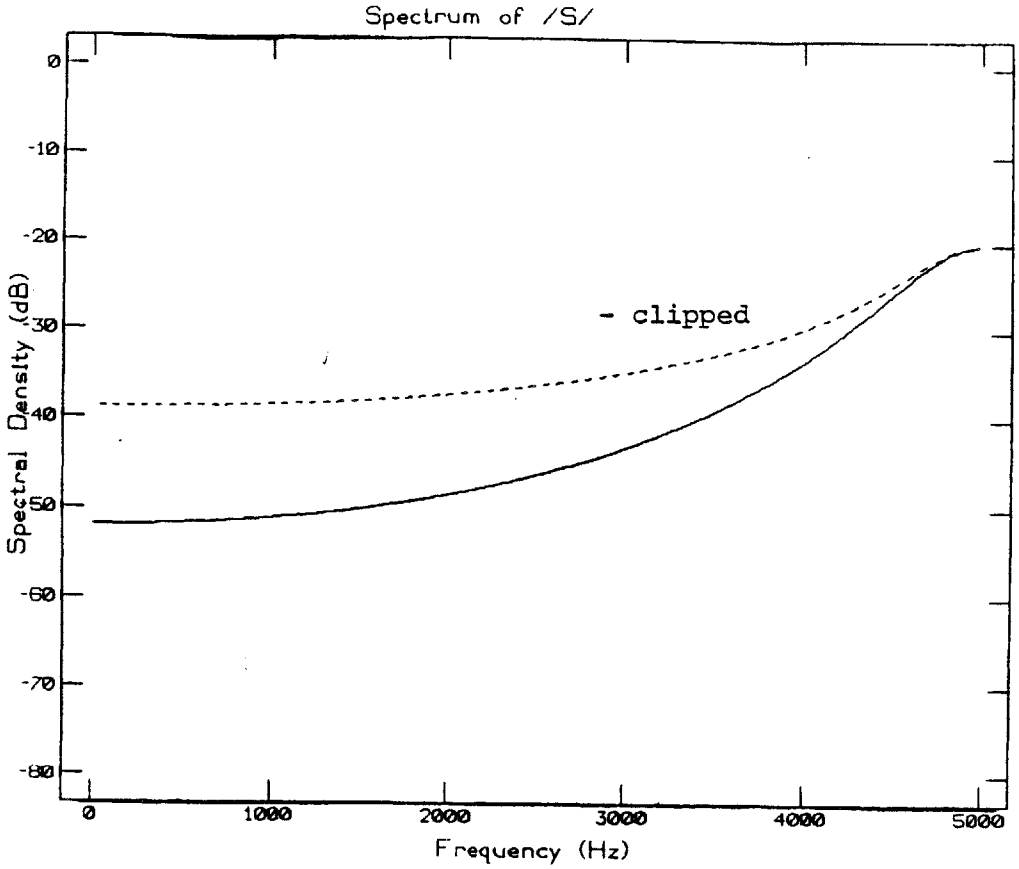


Figure 3.8 Spectra of the synthetic consonant noise burst, /sh/. Plots are the same as in Figure 3.7.



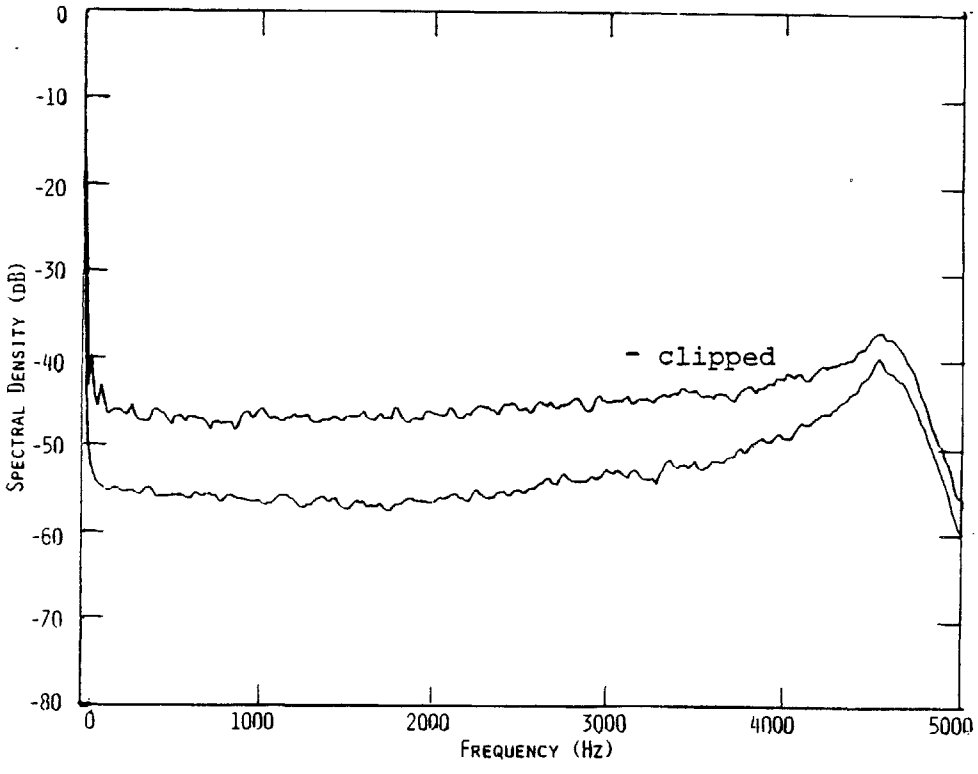
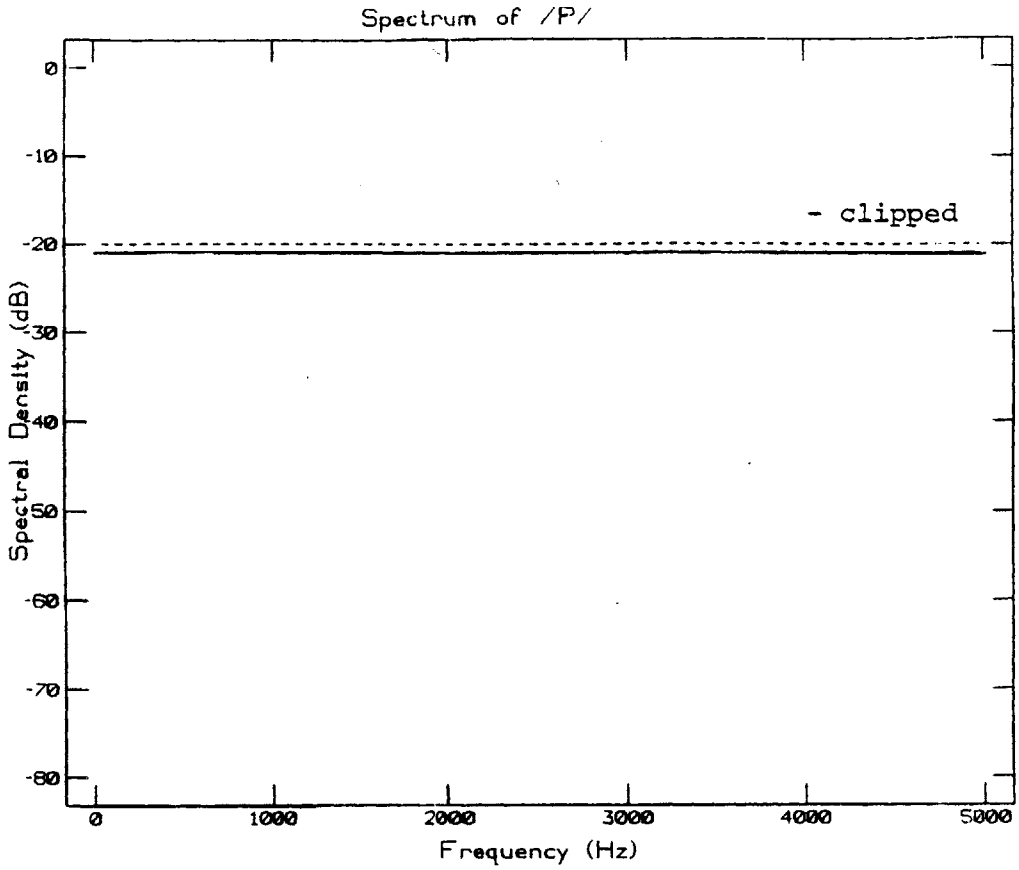


Figure 3.9 Spectra of the synthetic consonant noise burst, /s/. Plots are the same as in Figure 3.7.



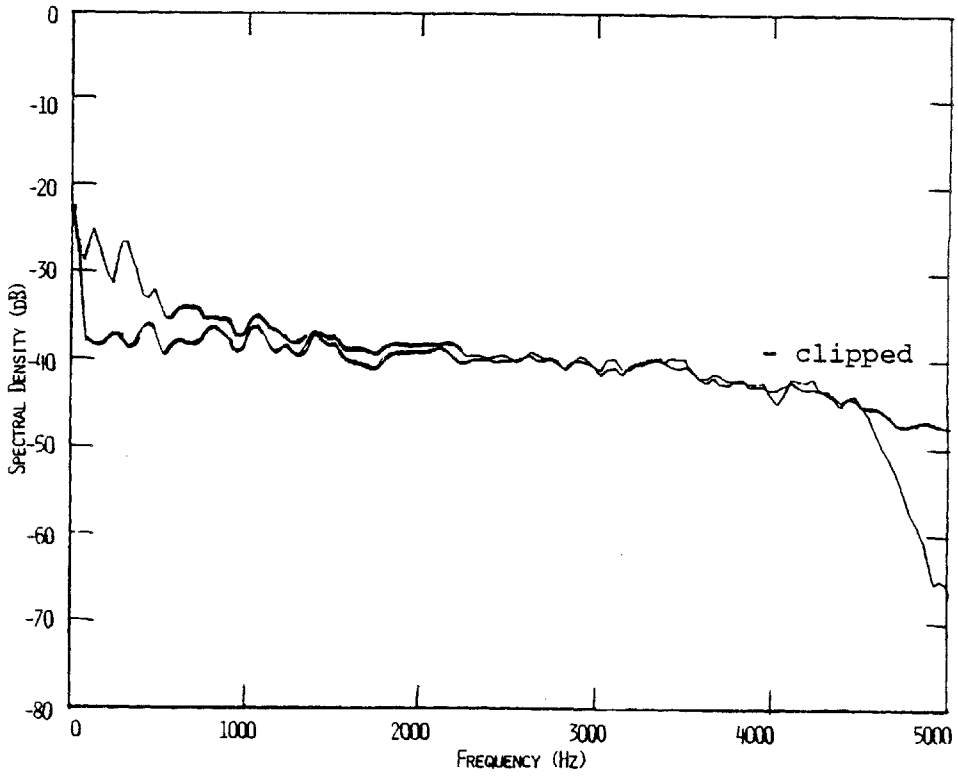
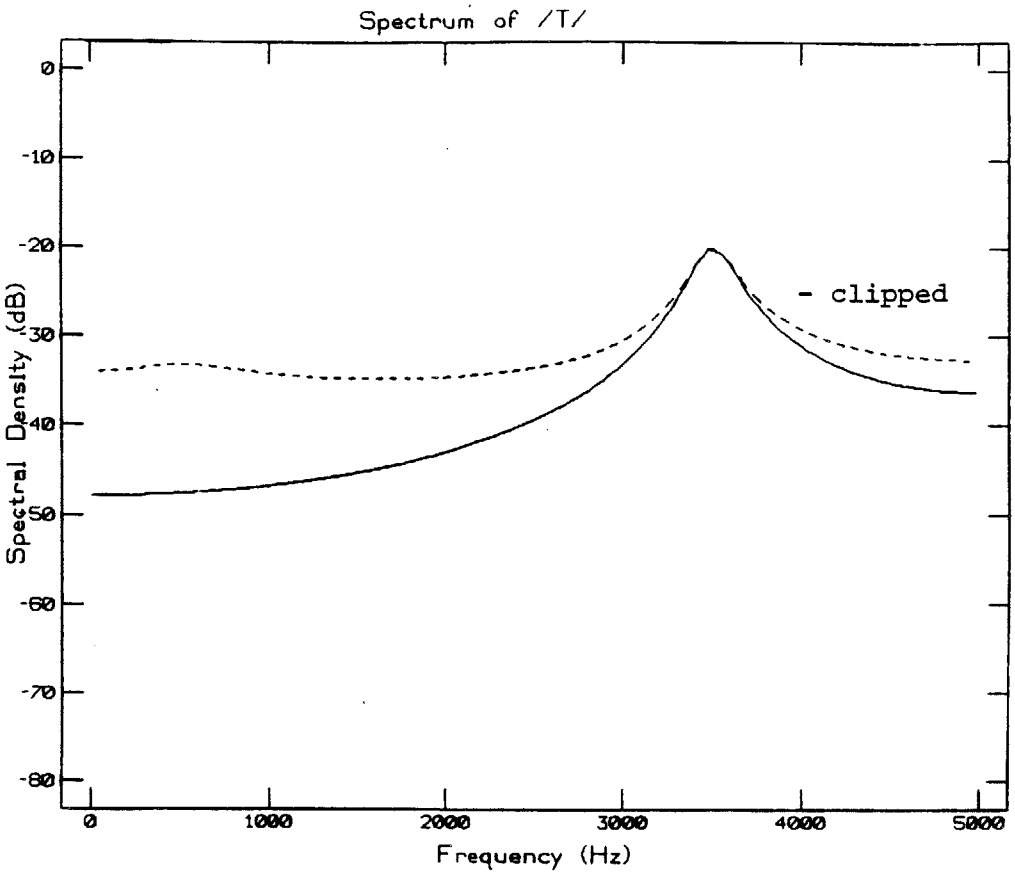


Figure 3.10 Spectra of the synthetic consonant noise burst, /p/. Plots are the same as in Figure 3.7.



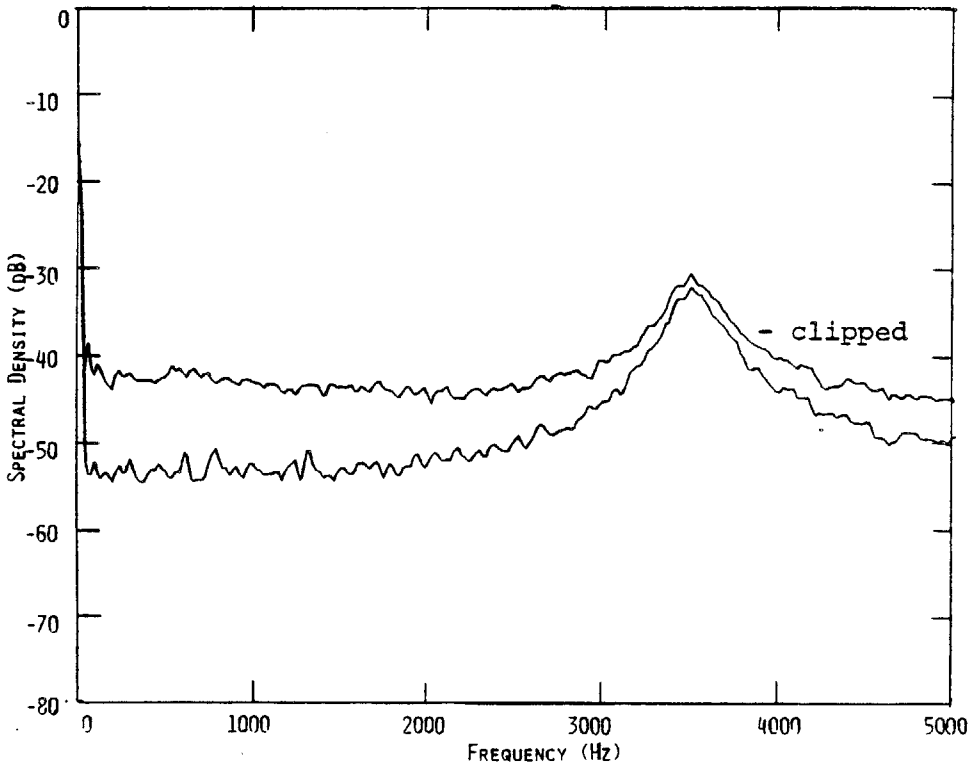
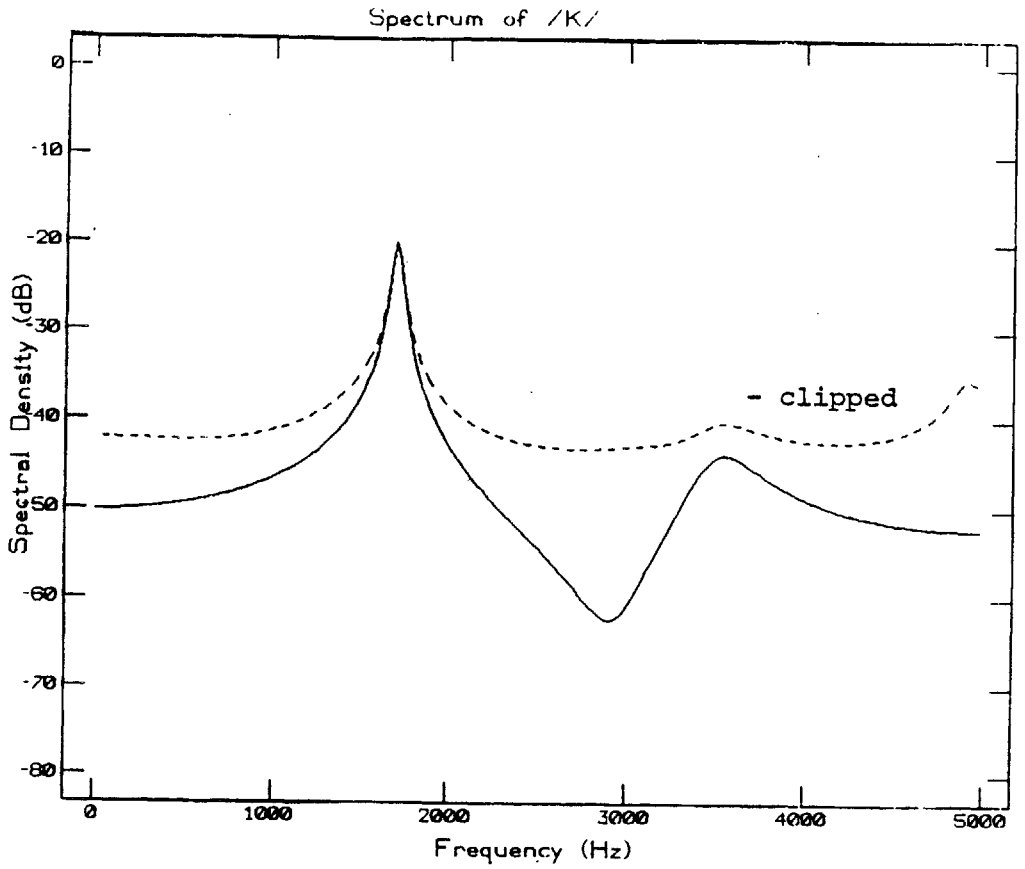


Figure 3.11 Spectra of the synthetic consonant noise burst, /t/. Plots are the same as in Figure 3.7.



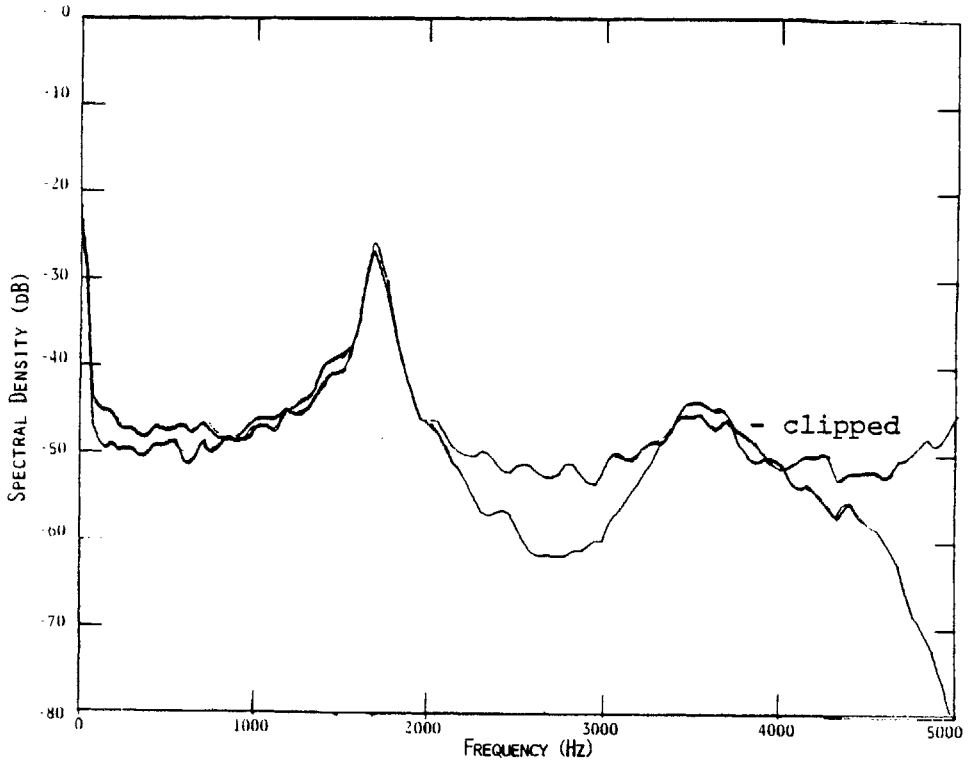
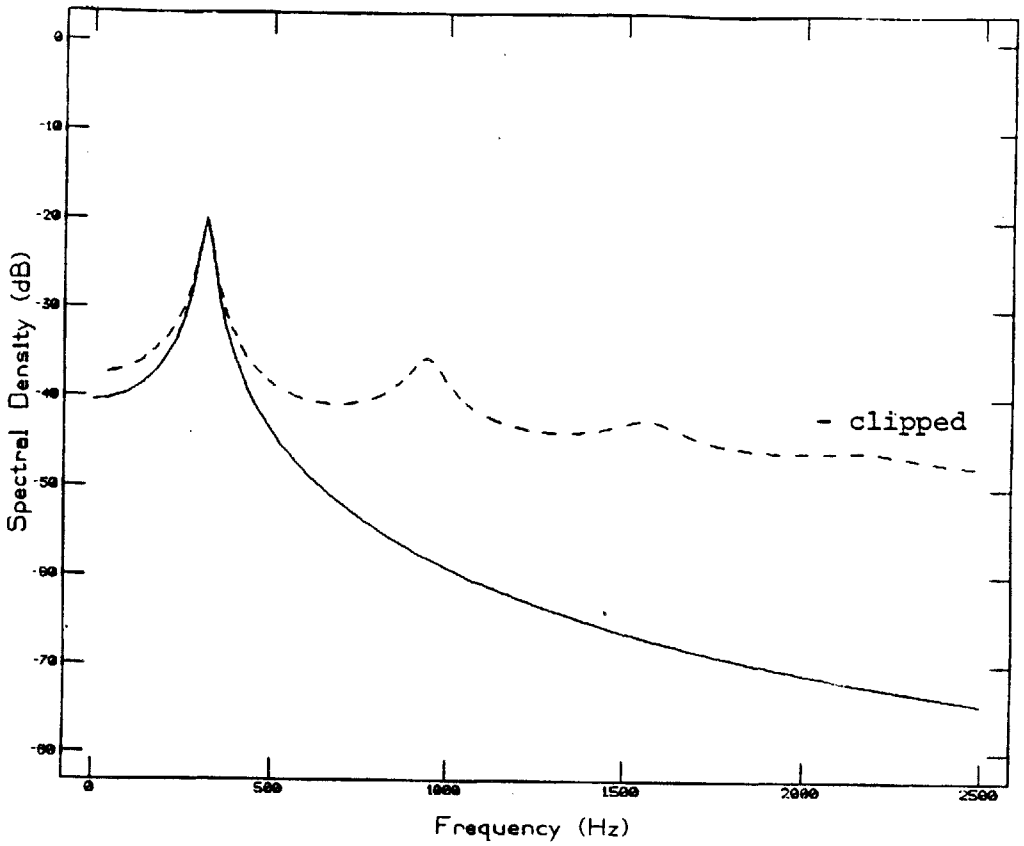


Figure 3.12 Spectra of the synthetic consonant noise burst, /k/. Plots are the same as in Figure 3.7.

In order to look at relative changes in the clipped spectrum after varying spectral frequencies and amplitudes, random noise was filtered with the 1925 General Radio 1/3 octave multifilter. The output was clipped and analyzed with the H.P. analyzer. Power spectra were also calculated by varying the parameters of the filter transfer function of the synthesized consonants and the arcsin law was applied. (Thus, the calculation was performed using a filter transfer function different from the one used empirically. This difference can be seen between the unclipped calculated and measured spectra in the figures.) Spectra with a single peak at 315 Hz and 1250 Hz are shown in Figures 3.13 and 3.14. Spectra with two peaks are shown in Figures 3.15 through 3.19.

Several observations can be made about the spectral modifications caused by clipping. First, clipping tends to fill in "valleys" in the spectrum, much as an increased noise floor would. Second, as expected from the Fourier series for a square wave, harmonics appear at odd multiples of the spectral peaks. This is especially noticeable in Figures 3.13 and 3.14 which display a single peak. Third, the amplitude relation between two spectral peaks after clipping,



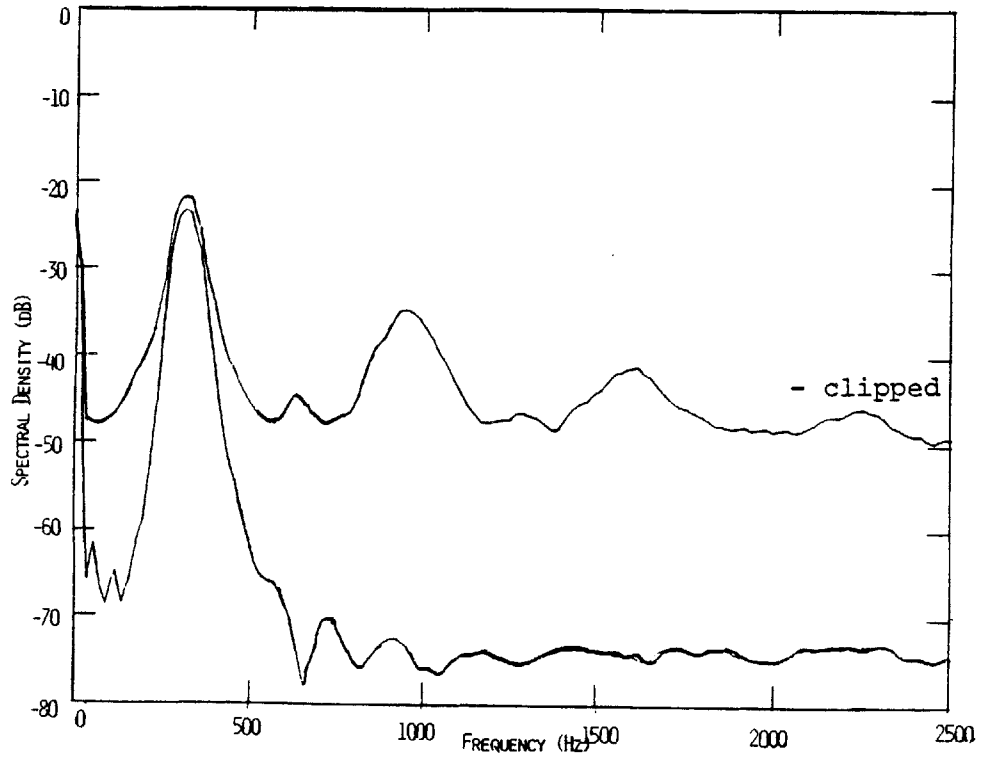
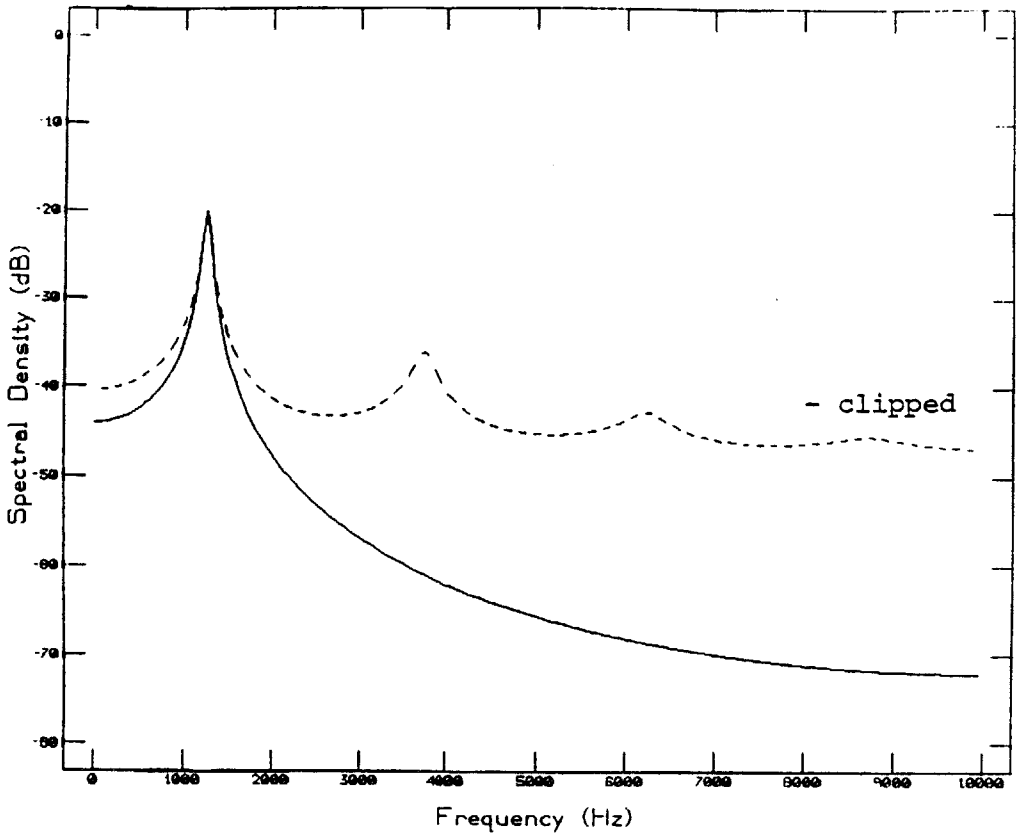


Figure 3.13 Spectra of filtered narrowband noise centered at 315 Hz measured before and after clipping. The top plot is calculated from the arcsin law using a different transfer function from that measured for the input noise. The bottom plot is measured with the spectrum analyzer. The format of this plot is the same as in Figures 3.7-3.12.



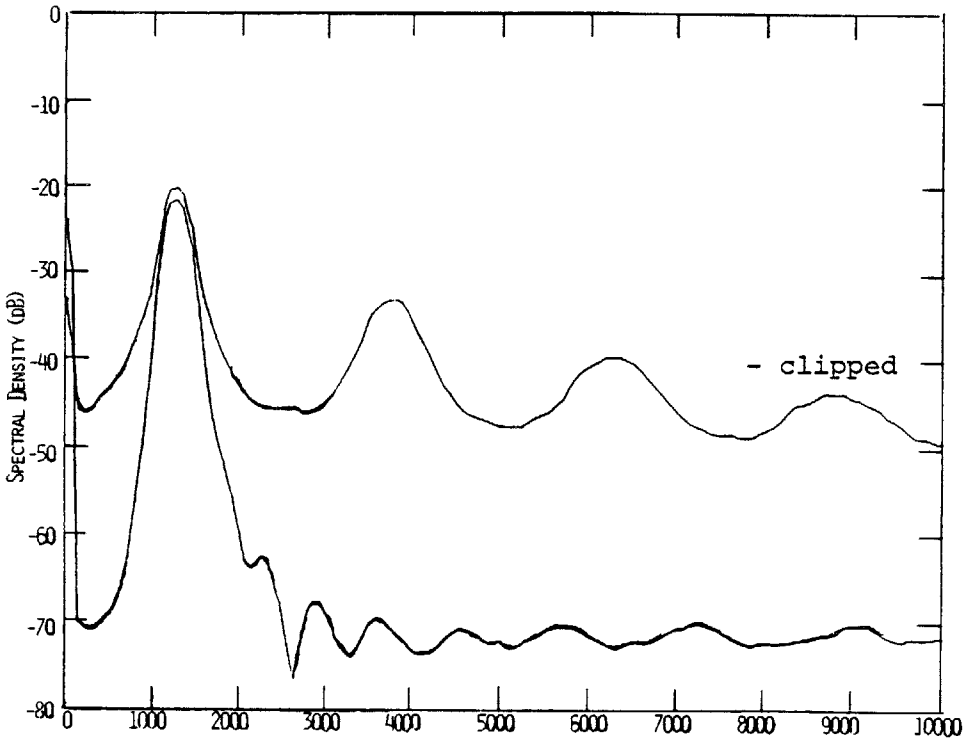
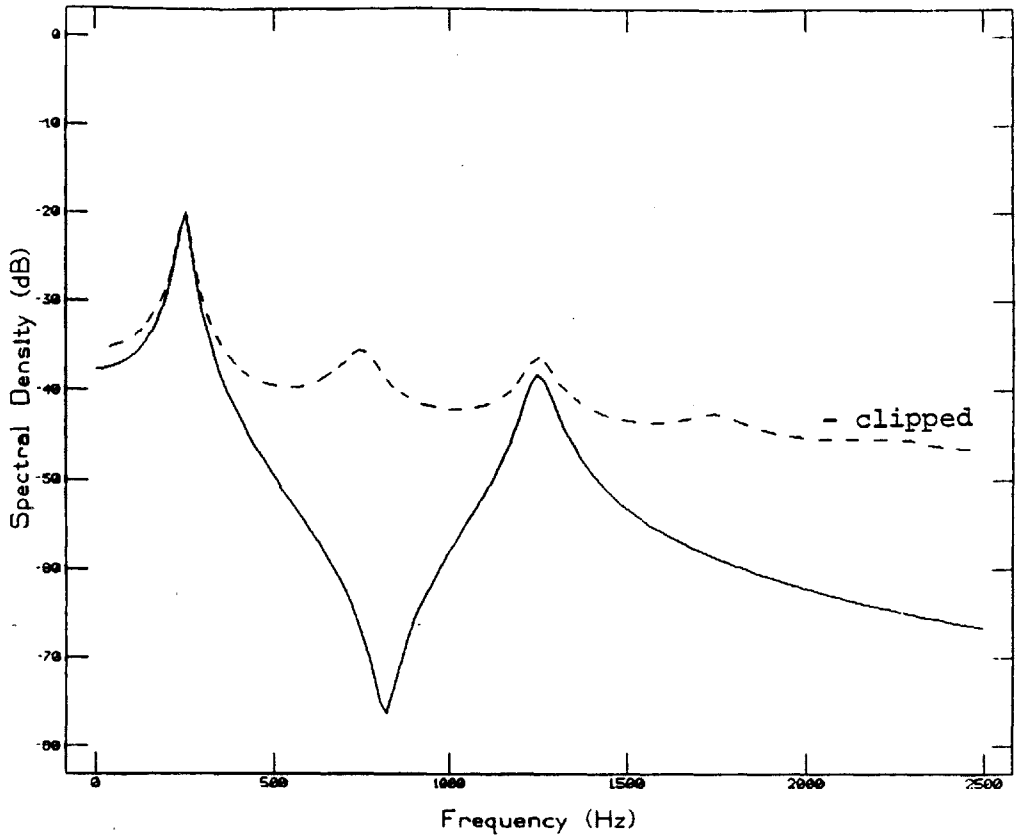


Figure 3.14 As in Figure 3.13 but with input noise centered at 1250 Hz.



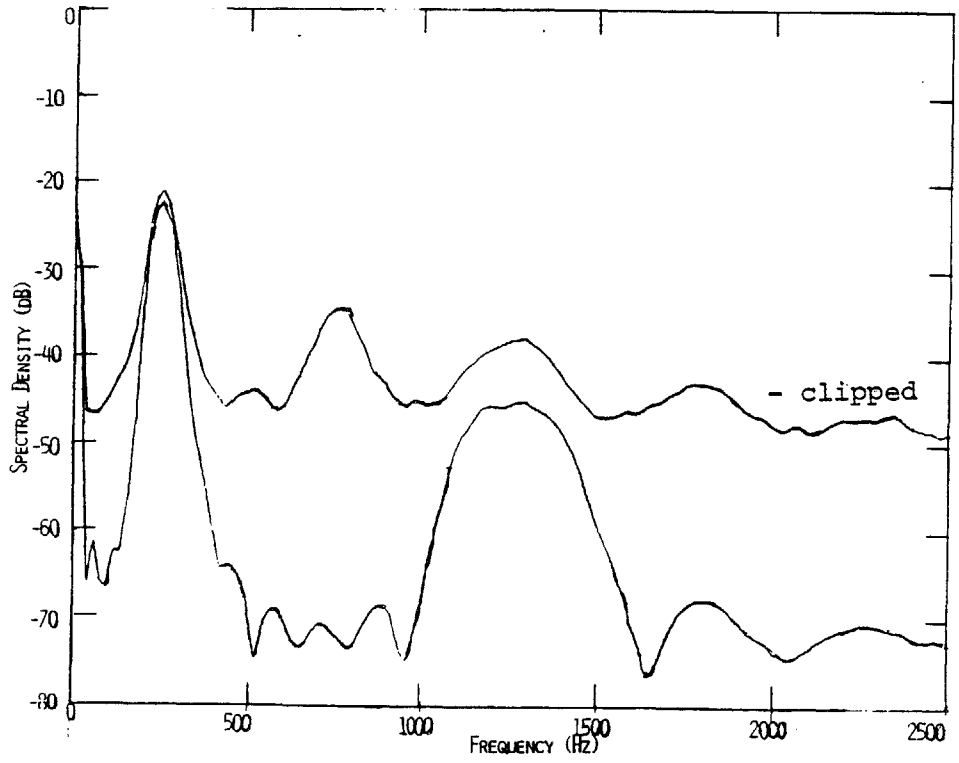
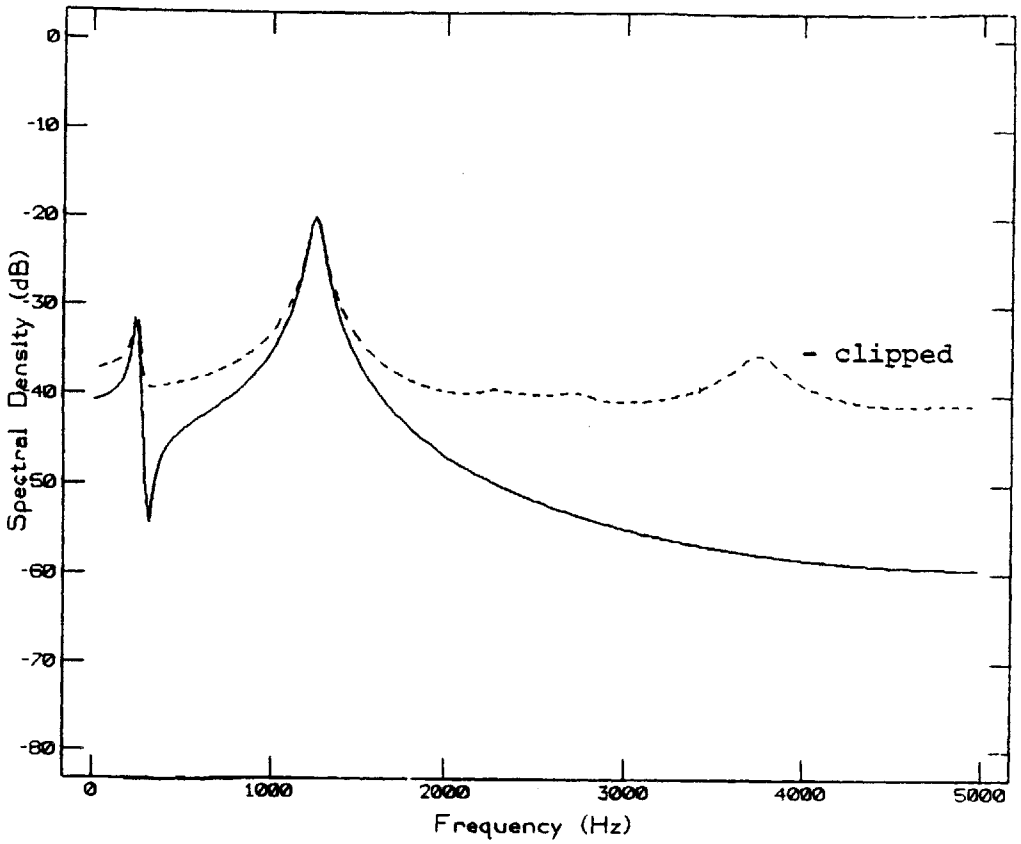


Figure 3.15 As in Figure 3.13 but with input noise peaks centered at 250 Hz and 1250 Hz.



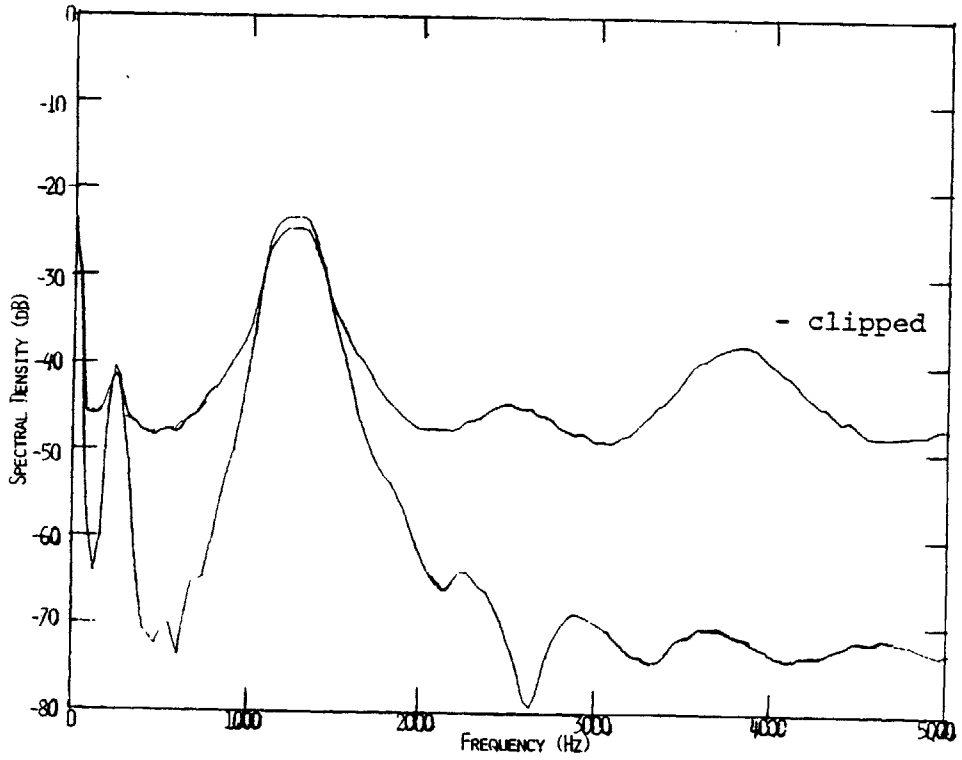
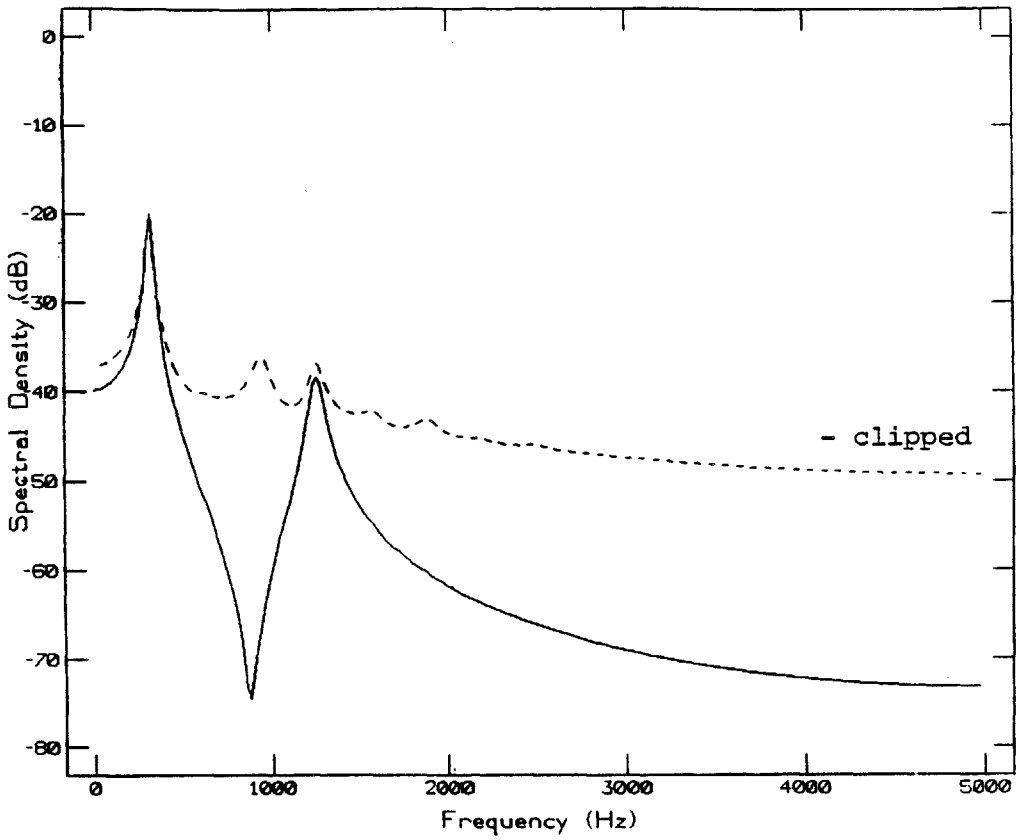


Figure 3.16 As in Figure 3.13 but with input noise peaks centered at 250 Hz and 1250 Hz.



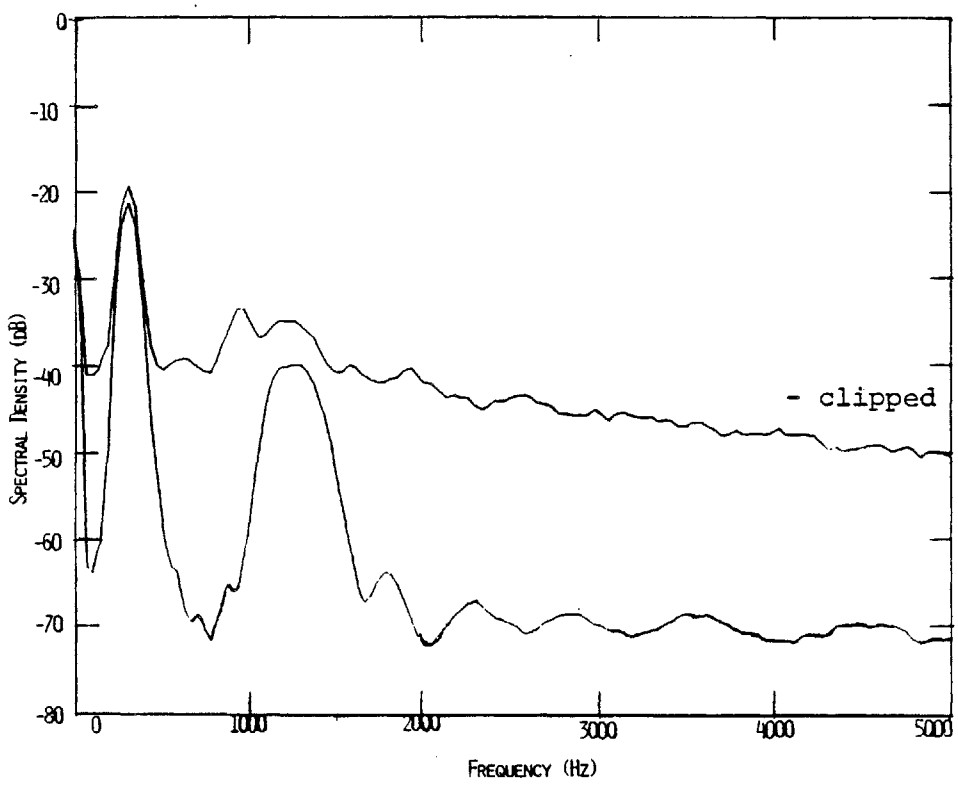
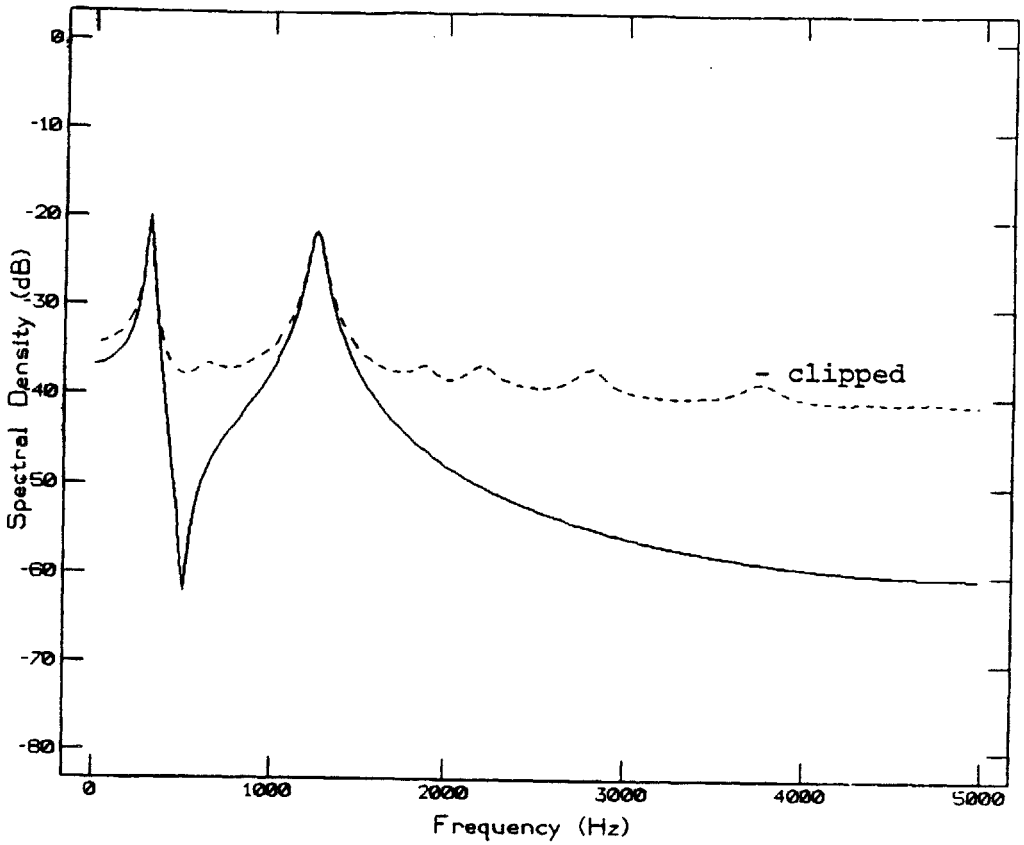


Figure 3.17 As in Figure 3.13 but with input noise peaks centered at 315 Hz and 1250 Hz.



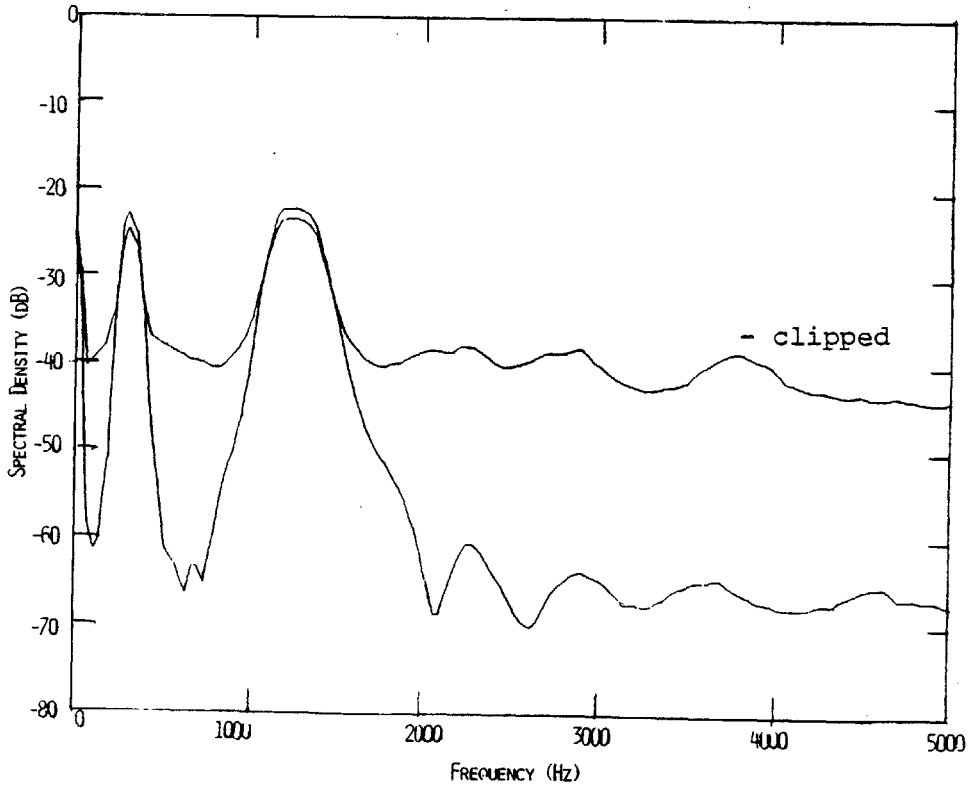
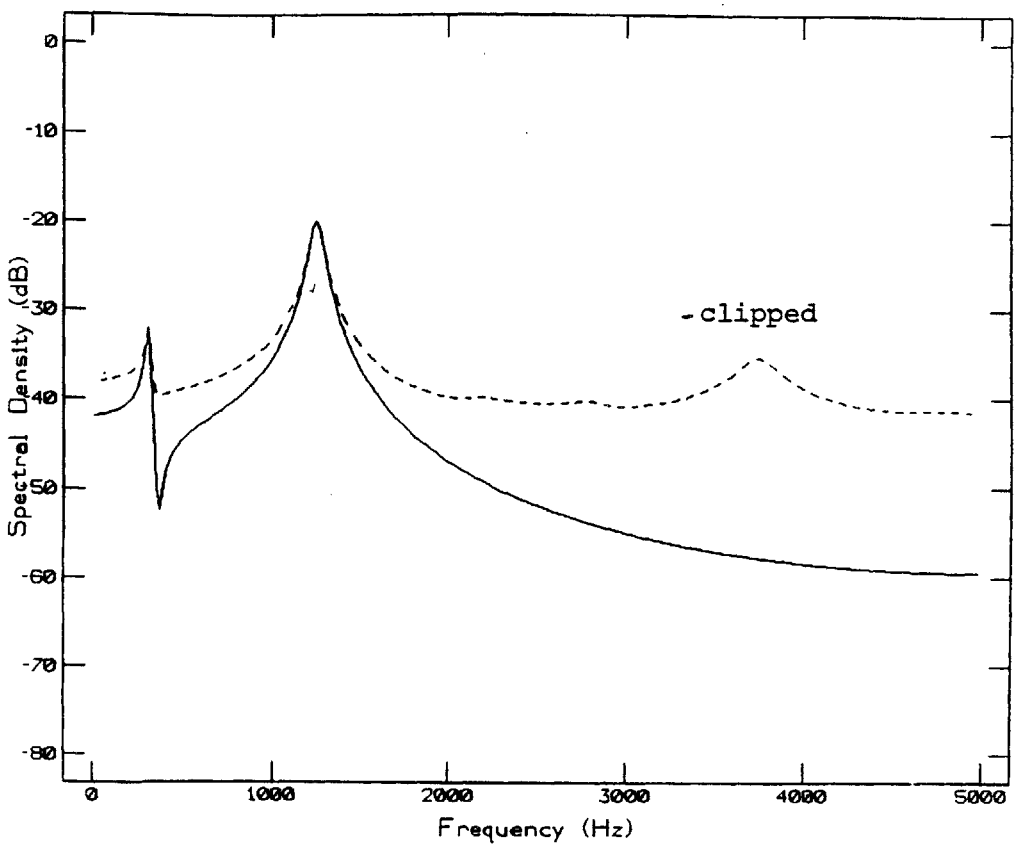


Figure 3.18 As in Figure 3.13 but with input noise peaks centered at 315 Hz and 1250 Hz.



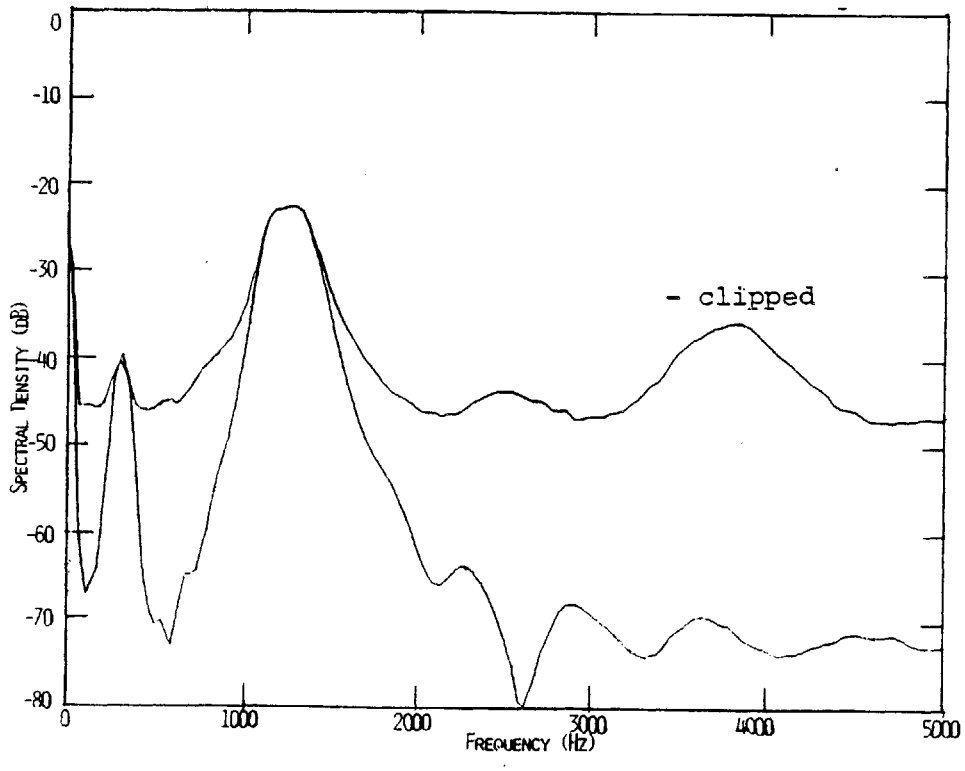


Figure 3.19 As in Figure 3.13 but with input noise peaks centered at 315 Hz and 1250 Hz.

when neither is obscurred by distortion components, is nearly the same as the input (Figures 3.16, 3.18, 3.19). Low-level peaks can be obscurred by the harmonics of higher-level peaks, especially if the frequency of the low-level peak is greater than the frequency of the higher-level peak (Figures 3.15 vs. 3.16 and 3.17 vs. 3.19). The low-level peak can also be enhanced by the harmonics of higher-level peaks (Figure 3.15).

3.4.2 Phase Dependence of Clipped Voiced Sounds

For the case of voiced sounds, which consist of harmonic components, it is expected that the arcsin law will not be applicable in general, although it may possibly provide a useful approximation. In this section the problem in predicting the clipped spectrum of harmonic sounds is demonstrated.

The problem can be seen by constructing a signal that is the sum of harmonically related sine waves and varying the phases of the components (Figure 3.20). The three frequency components were chosen in a one-third octave band with center

frequency of 1000 Hz.

Any prediction of the clipped spectrum based solely on the input auto-correlation function would be independent of the input phases. It is clear that clipped spectra are dependent on the phases of the individual frequency components in the input (panels A-C of Figure 3.21). The prediction from the arcsin law is shown is Panel D of Figure 3.21. It will be noted that the gross spectral shape is rather consistent across the three clipped spectra and follows the shape predicted by the arcsin law. However, the validity of the arcsin law approximation in other harmonic cases is unknown. Given that fine details of the spectrum are lost as a result of "critical-band" filtering by the ear, an approximation as good as that shown in Figure 3.21 would probably be adequate. Determining the goodness of the arcsin approximation for harmonic signals is a worthwhile but lengthy task which was not pursued here because of time constraints.

If the tones in the signal are not harmonically related, phase shifts of components can be viewed as a shift in the time orgin, and there should be no effect of phase shifts on the clipped spectrum. However, this case is not relevant to speech signals, since the components of voiced sounds are harmonically related.

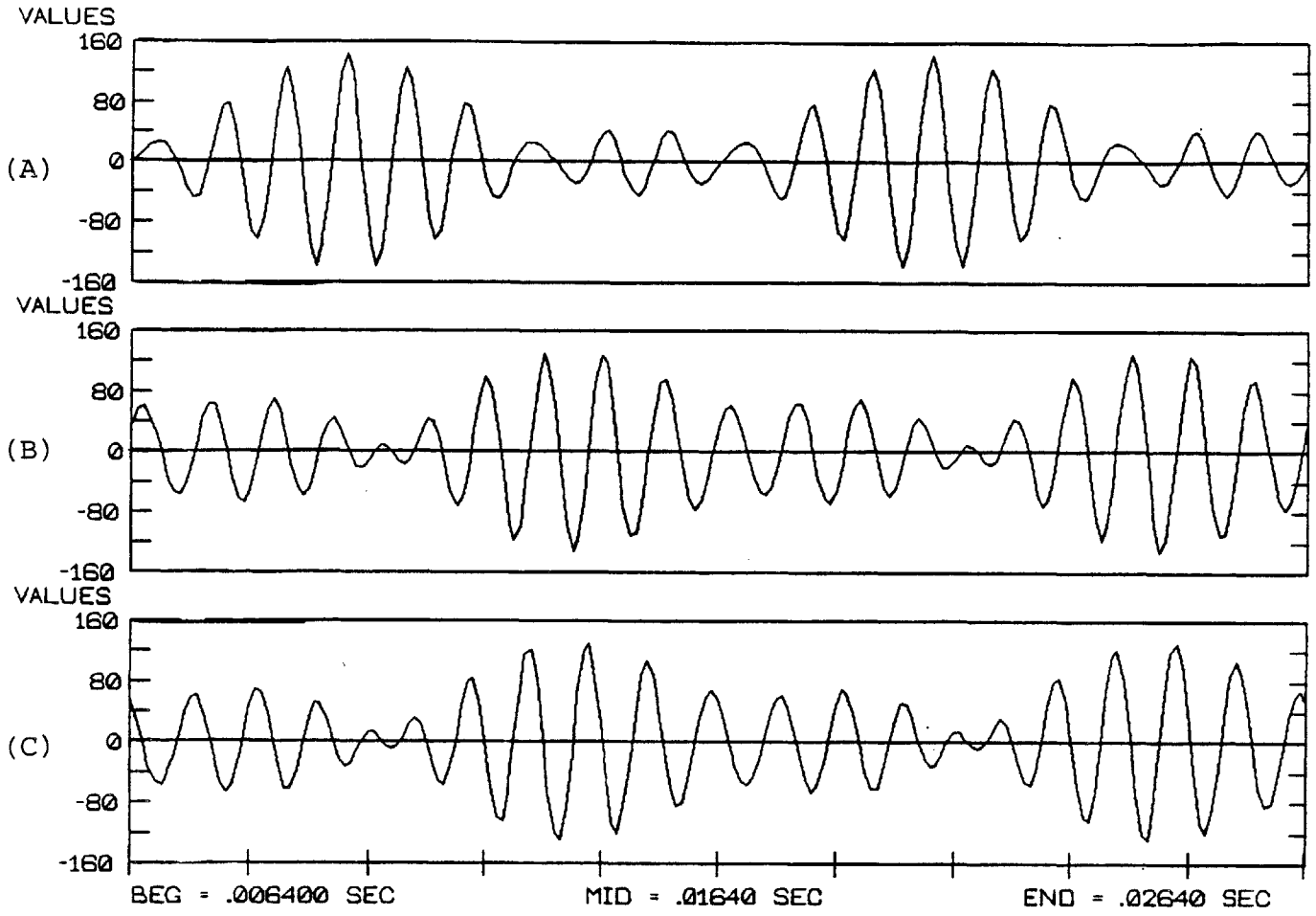


Figure 3.20 Differing time-waves caused by phase shifting. The signal was generated by summing three harmonically related tones with frequencies of 900, 1000, and 1100 Hz, and relative amplitudes of 0, -2, and -5 dB respectively. The three tones (900, 1000, and 1100 Hz) in Panel A have zero phase. The phases of the three tones were respectively 0, 180, and 90 degrees in Panel B. The phases were 90, 270, and 180 degrees in Panel C.

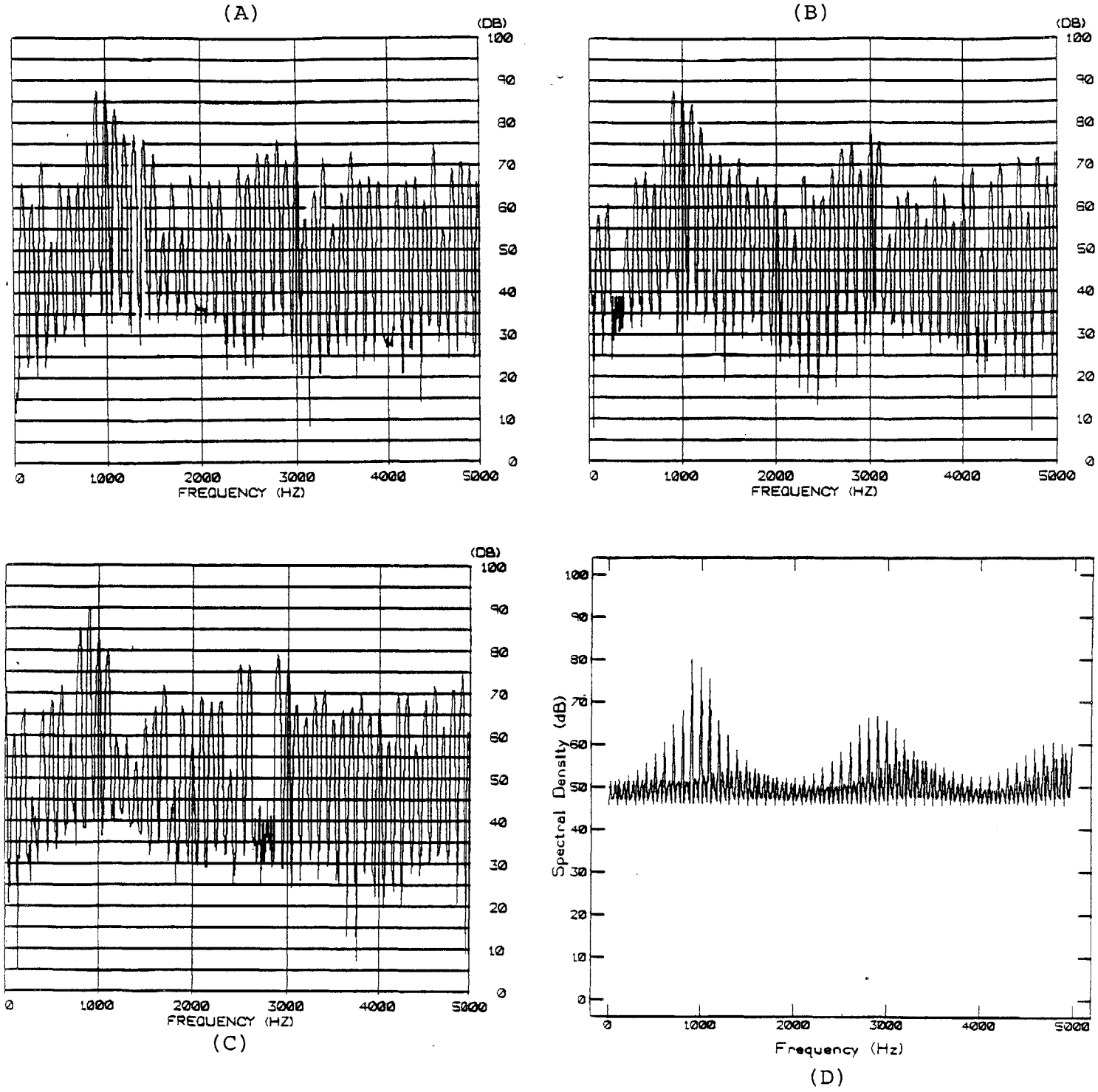


Figure 3.21 The phase dependence of harmonically related tones. Panels A, B, and C show the clipped magnitude spectra with the three signals in Figure 3.20 as input. Panel D shows the prediction of the arcsin law applied to the input.

3.4.3 Spectra of Clipped Vowels

In order to observe the effects of clipping on vowel spectra, measurements were made with steady-state synthetic vowels. The spectra of the vowels were evaluated with the ILS spectral analysis package.

The vowels were synthesized using a model in which the glottal source is passed through a filter designed to model the vocal tract (Rabiner and Schafer, 1978). Parameters of the model were varied to change the frequencies and bandwidths of the formants. In the present implementation, an impulse was used as a glottal pulse waveform in order to remove the negative spectral tilt associated with the more natural waveform. It is easier to evaluate the clipped spectra after eliminating this low-frequency emphasis.

The vowels were synthesized with a fundamental frequency of 125 Hz. Thus, the spectrum of each vowel is a sequence of harmonics spaced 125 Hz apart with spectral-envelope peaks at specified formant frequencies.

The spectra of clipped synthetic vowels are presented in Figures 3.22 through 3.39. There are some major differences between the clipped spectra of vowels and noise. A clipped vowel spectrum is a sequence of harmonics with the same

inter-component spacing of 125 Hz as the unmodified spectrum, but with components extending to higher frequencies. As with noise, harmonics appear at odd multiples of the spectral peaks. However, there is no counterpart to the "noise floor" seen with clipped noise.

Clipped harmonics of the fundamental frequency components can emphasize or de-emphasize formant peaks. For example, in Figures 3.22 through 3.25, the formant is at 1000 Hz, a multiple of the pitch period, and the clipped spectrum has peaks only at odd harmonics of this formant. In another example where the formant of the vowel is higher in frequency and not exactly at a harmonic of the pitch period (Figures 3.26-3.29, the clipped spectrum contains other spectral peaks at low frequencies. However, the original formant is preserved as an absolute maximum in the spectral envelope. The extent to which the formant is retained in the clipped spectrum, even when the input spectral peak is very shallow, is quite remarkable (Figures 3.25 and 3.29).

The clipped spectra of vowels with two formants have peaks at not only odd multiples of the formants but at other frequencies as well (Figure 3.30). Again, clipping seems to emphasize individual formants that are difficult to see in the unmodified spectrum (figure 3.31).

In Figures 3.32-3.34, the two formants (at 1000 Hz and 3000 Hz) are not only harmonically related to the pitch, but the higher frequency formant is an odd multiple of the lower frequency formant. After clipping, narrow spectral peaks appear at only odd harmonics of the formants. Even as the bandwidths of the formants are varied, no new peaks appear in the spectrum (as seen in Figures 3.30 and 3.31), and the spectrum remains unchanged except for the spectral height of the peaks.

When two formants are not harmonics of the fundamental pitch frequency, the formants are preserved but the clipped spectrum contains other spectral peaks (Figures 3.35-3.39). Similarly to noise, lower-amplitude formants in the spectrum are obscurred after clipping. (Figure 3.37).

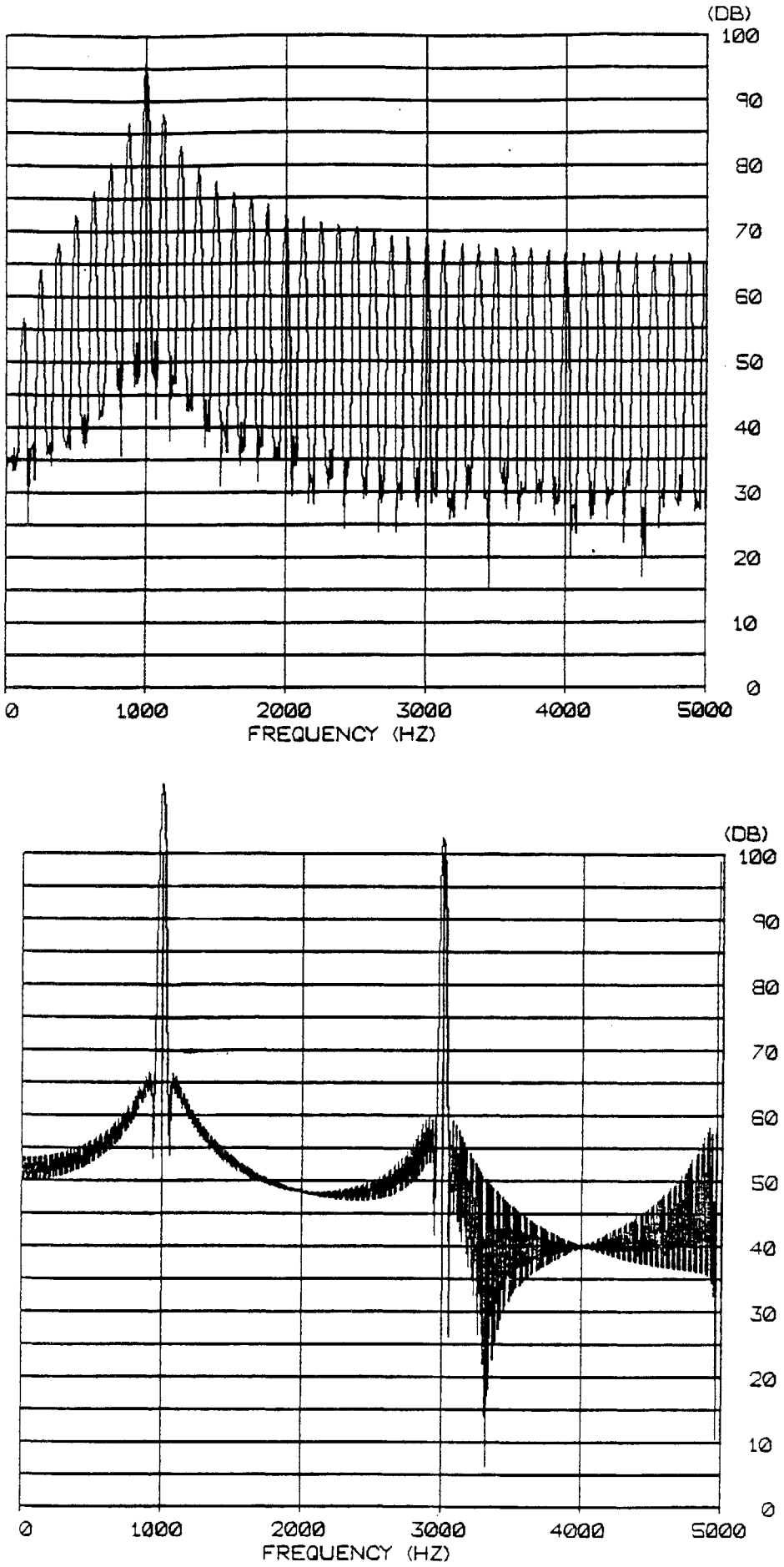
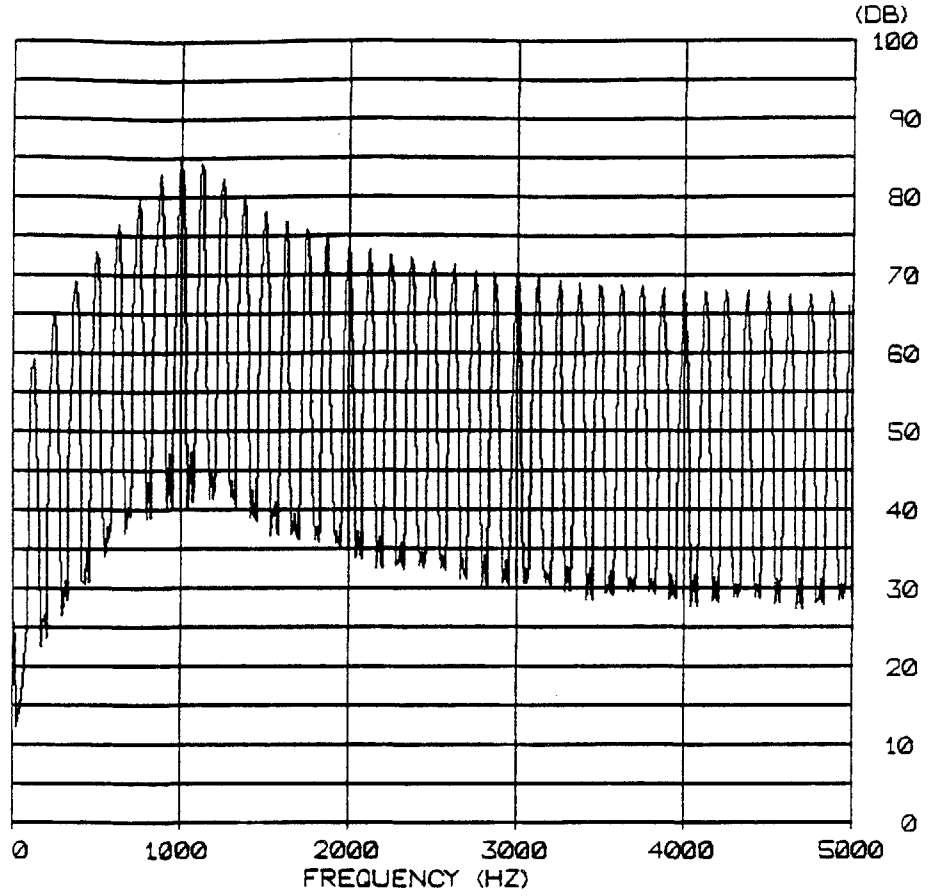


Figure 3.22 Spectra of a synthetic steady-state vowel measured before (top plot) and after (bottom plot) clipping. The formant frequency is 1000 Hz and bandwidth is 50 Hz.



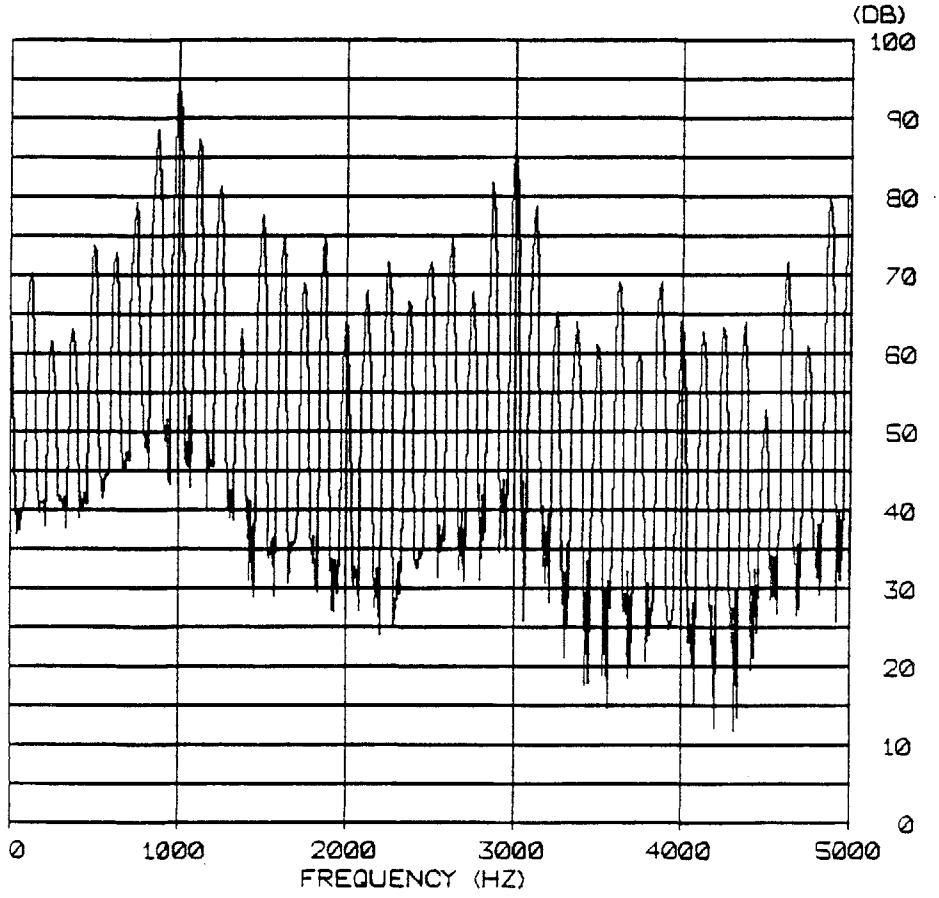


Figure 3.23 As in Figure 3.22 with a 1000 Hz formant frequency and a 200 Hz bandwidth.

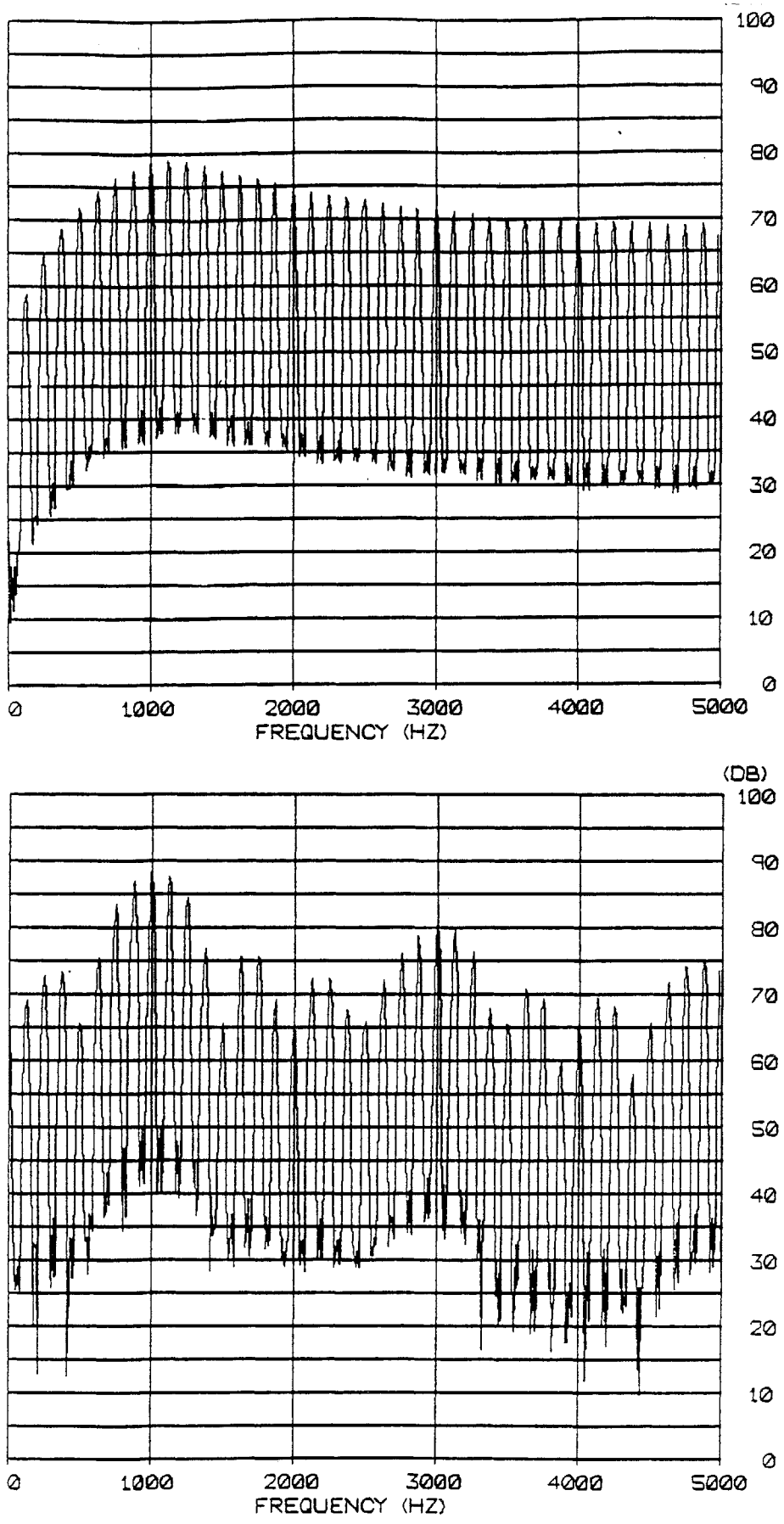


Figure 3.24 As in Figure 3.22 with a 1000 Hz formant frequency and a 500 Hz bandwidth.

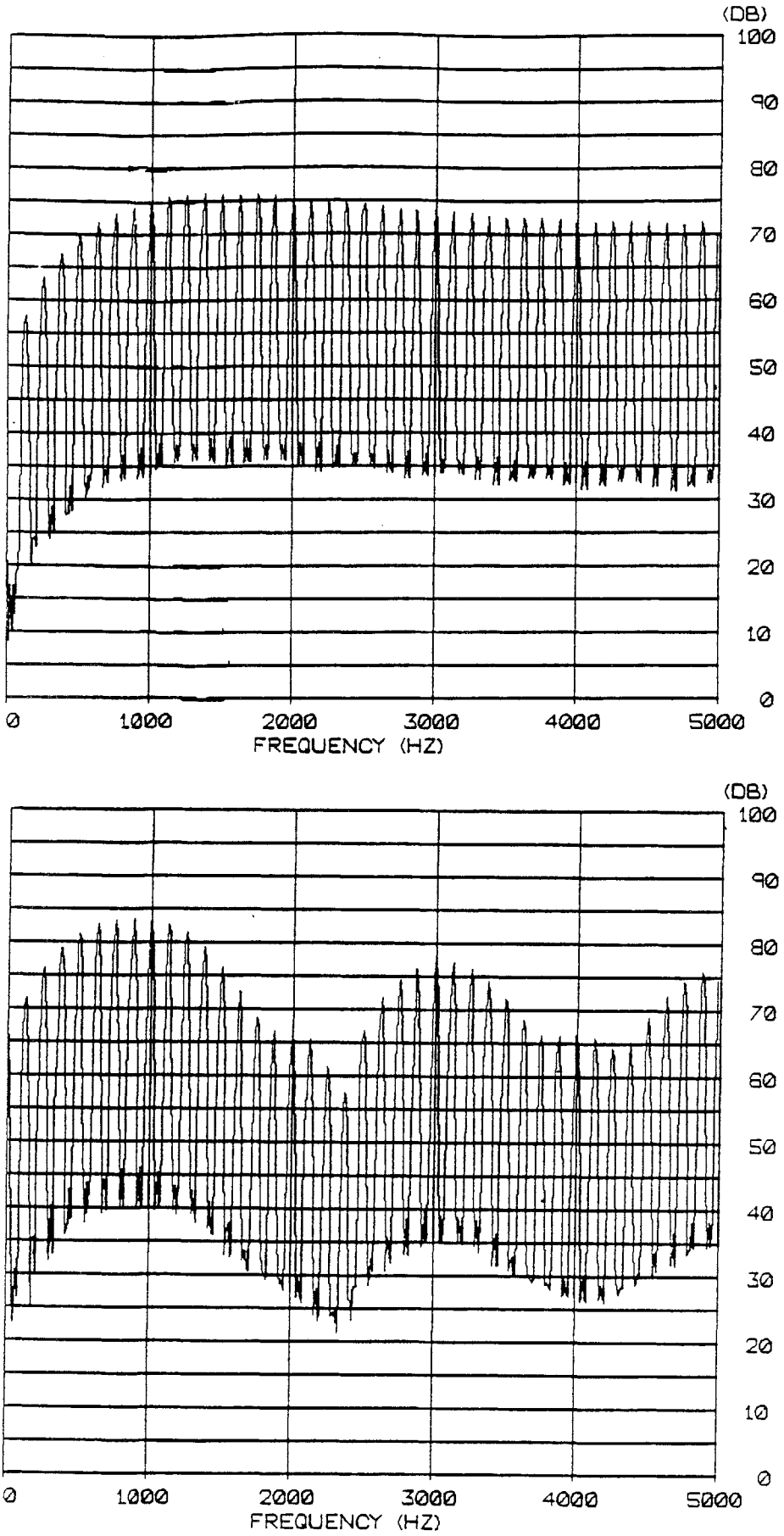


Figure 3.25 As in Figure 3.22 with a 1000 Hz formant frequency and a 1000 Hz bandwidth.

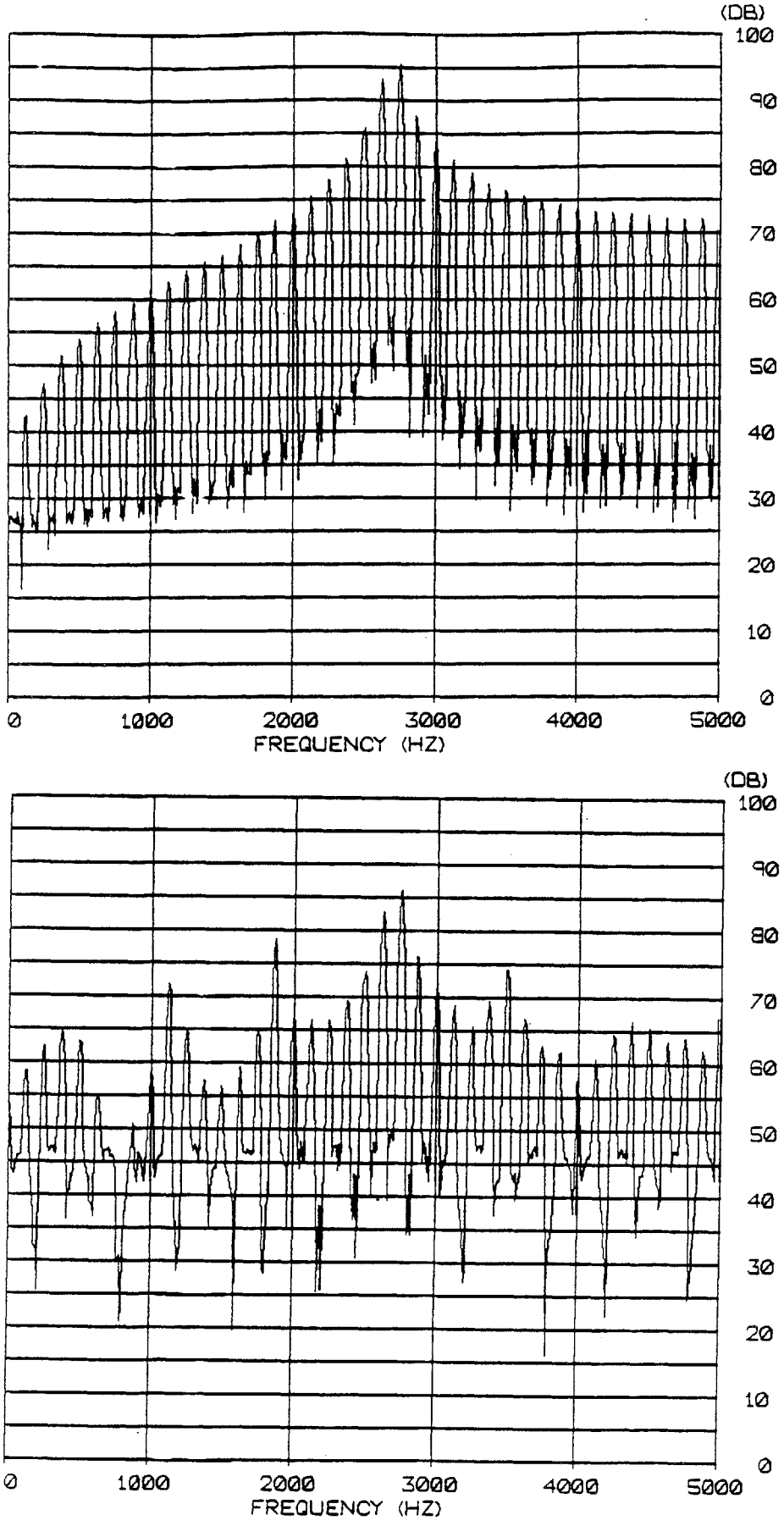
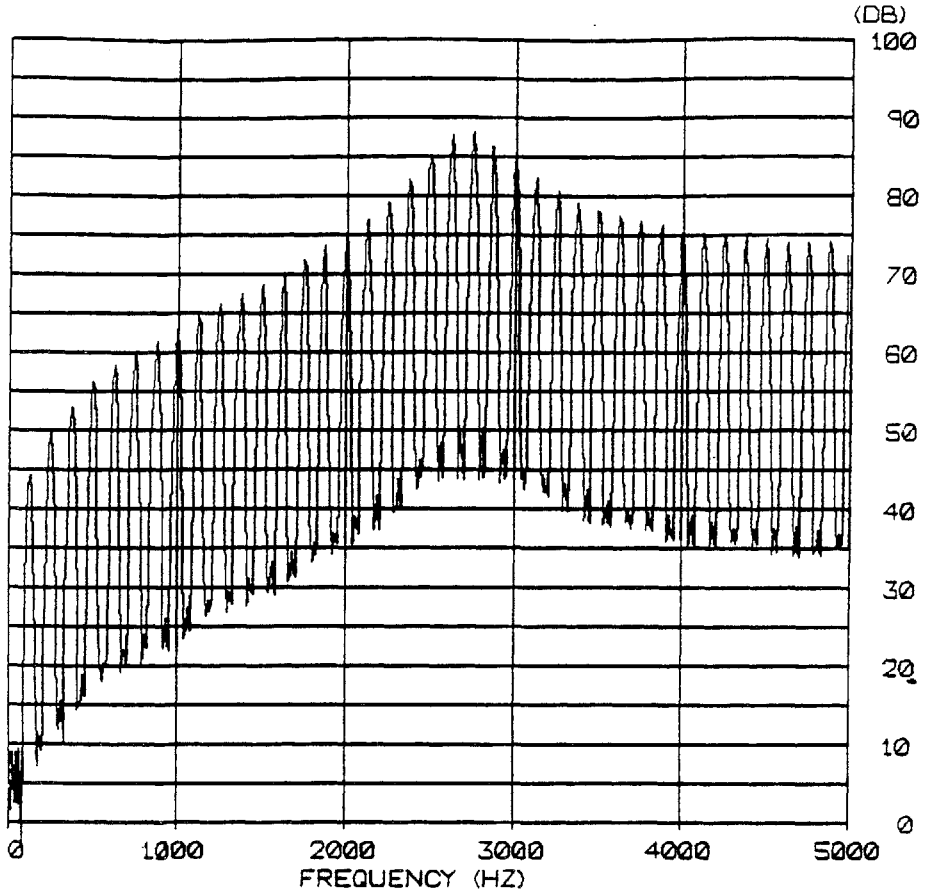


Figure 3.26 As in Figure 3.22 with a 2700 Hz formant frequency and a 50 Hz bandwidth.



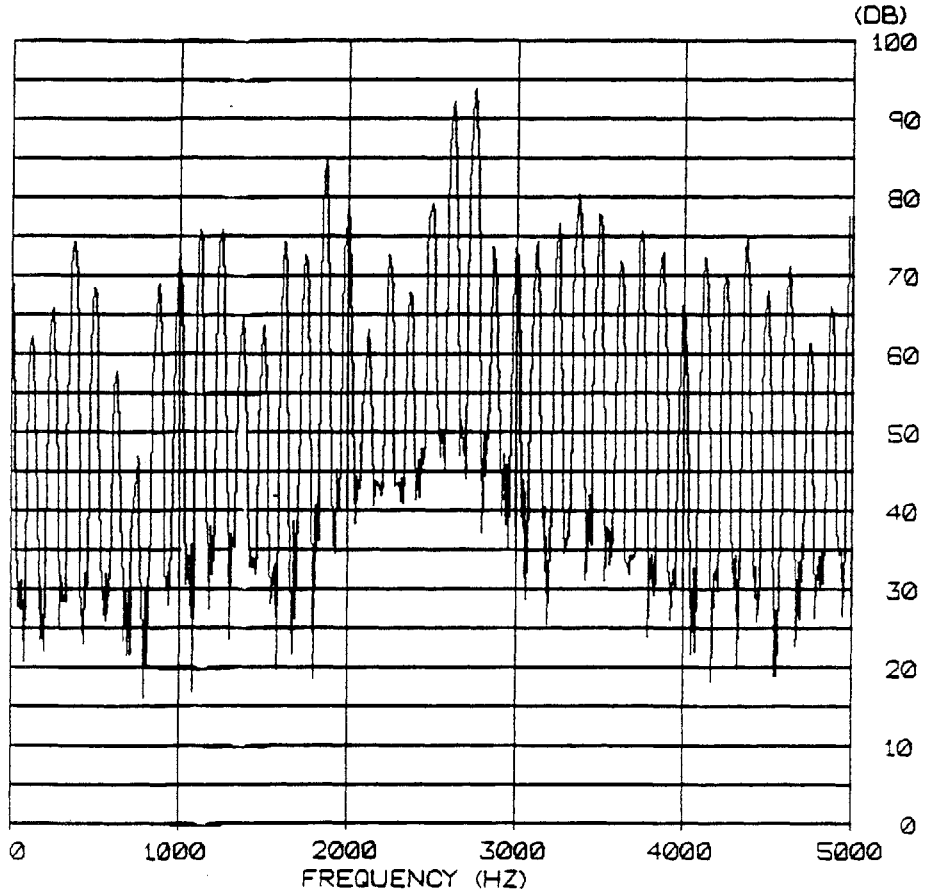
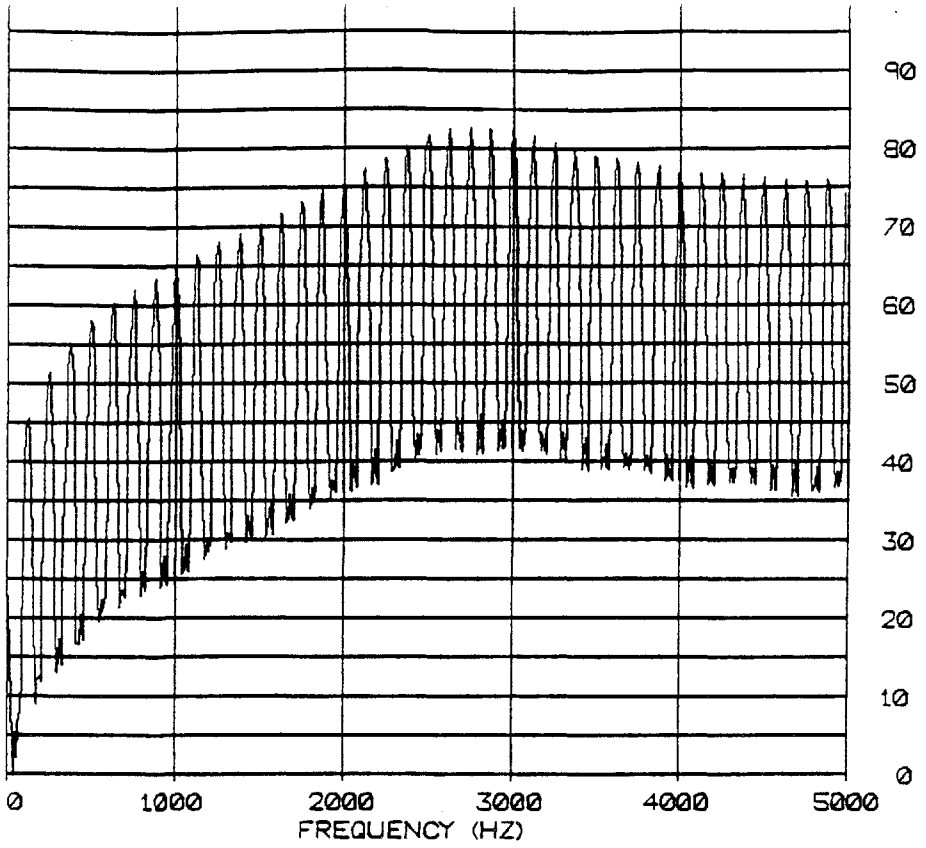


Figure 3.27 As in Figure 3.22 with a 2700 Hz formant frequency and a 100 Hz bandwidth.



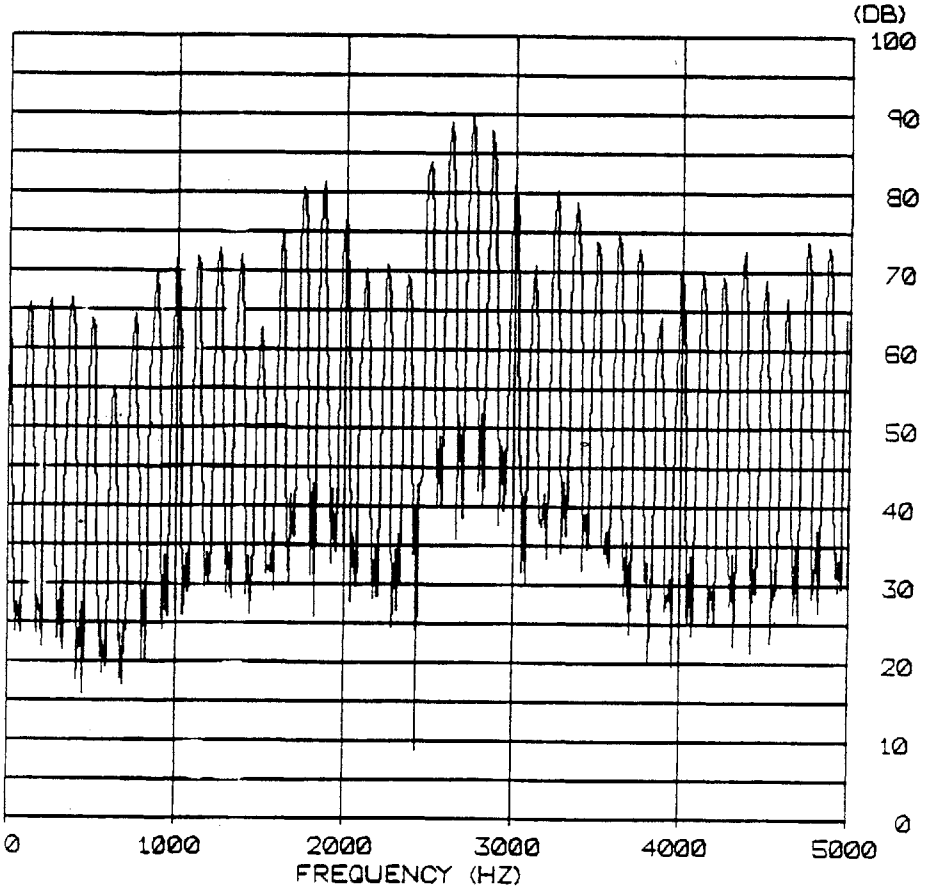


Figure 3.28 As in Figure 3.22 with a 2700 Hz formant frequency and a 500 Hz bandwidth.

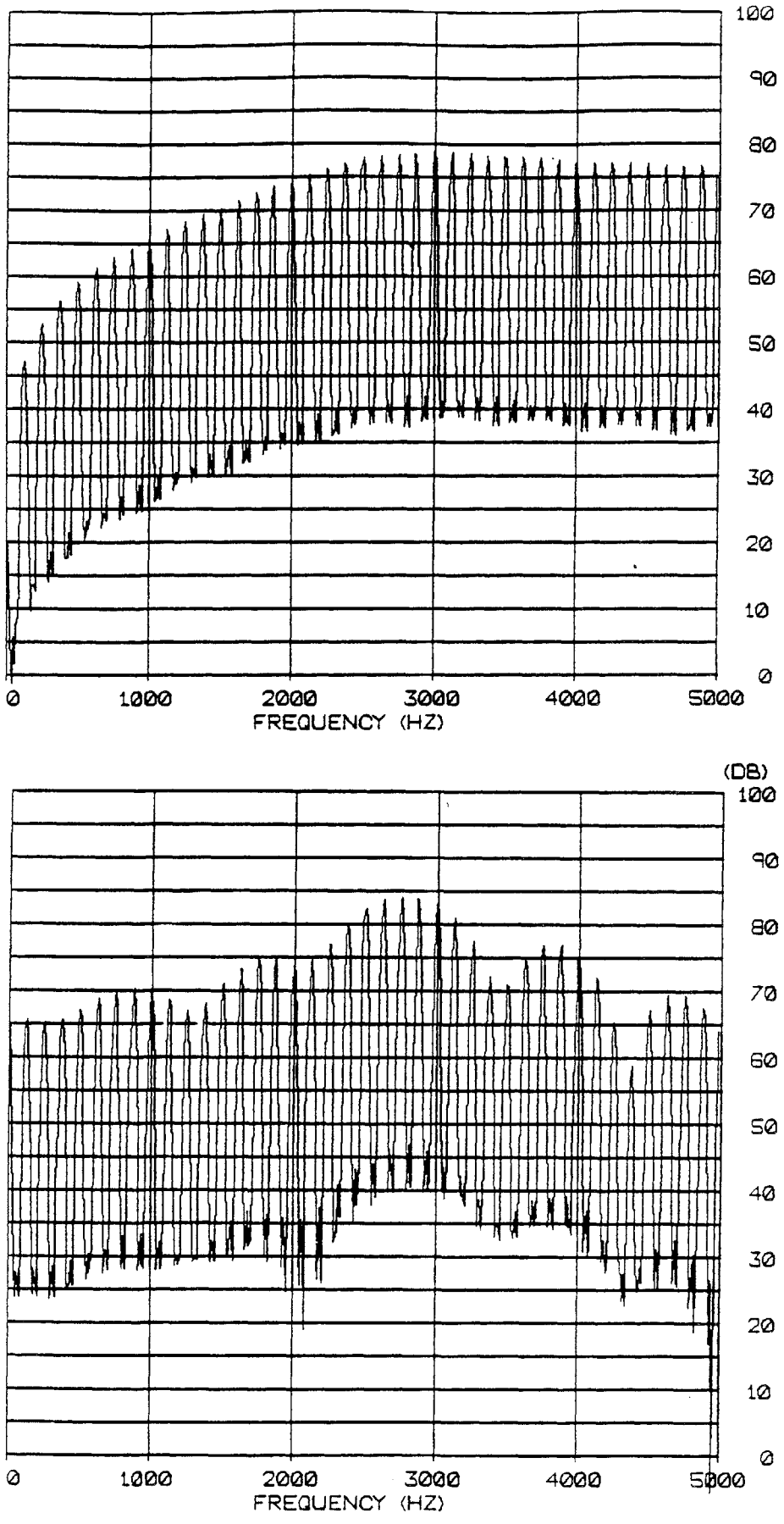


Figure 3.29 As in Figure 3.22 with a 2700 Hz formant frequency and a 1000 Hz bandwidth.

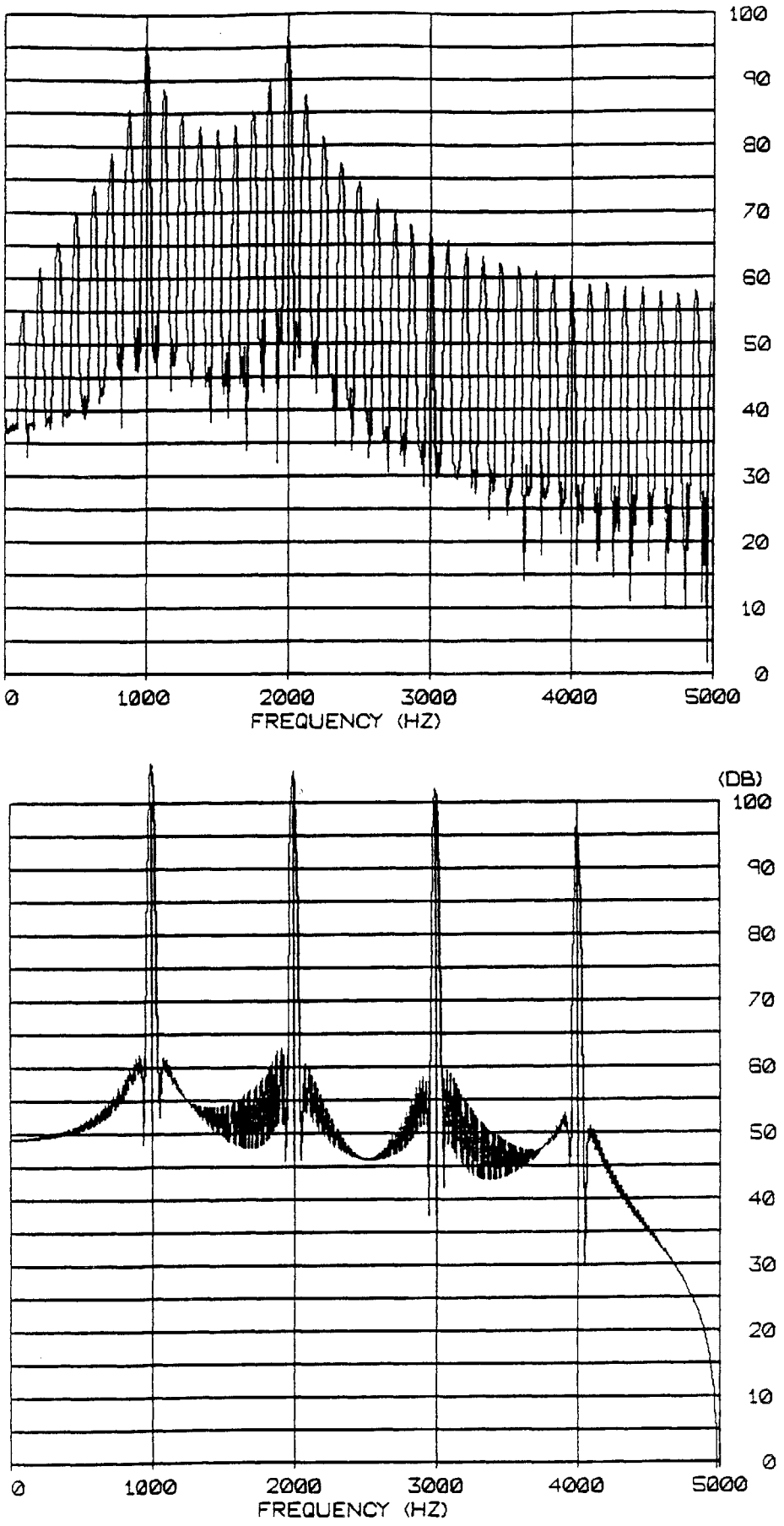


Figure 3.30 As in Figure 3.22. First formant frequency is 1000 Hz with 50 Hz bandwidth. Second formant frequency is 2000 Hz with 50 Hz bandwidth.

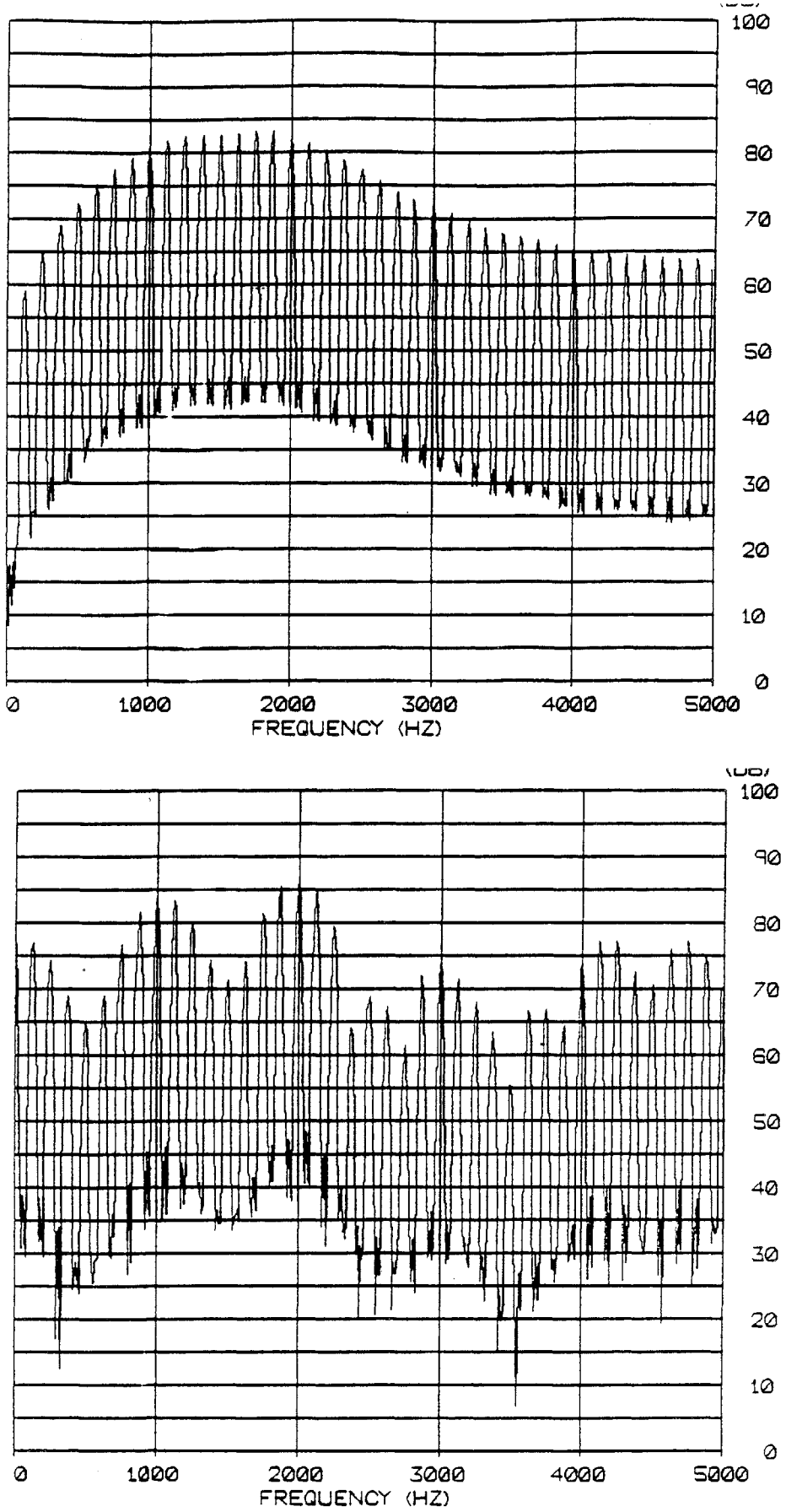


Figure 3.31 As in Figure 3.22. First formant frequency is 1000 Hz with 500 Hz bandwidth. Second formant frequency is 2000 Hz with 500 Hz bandwidth.

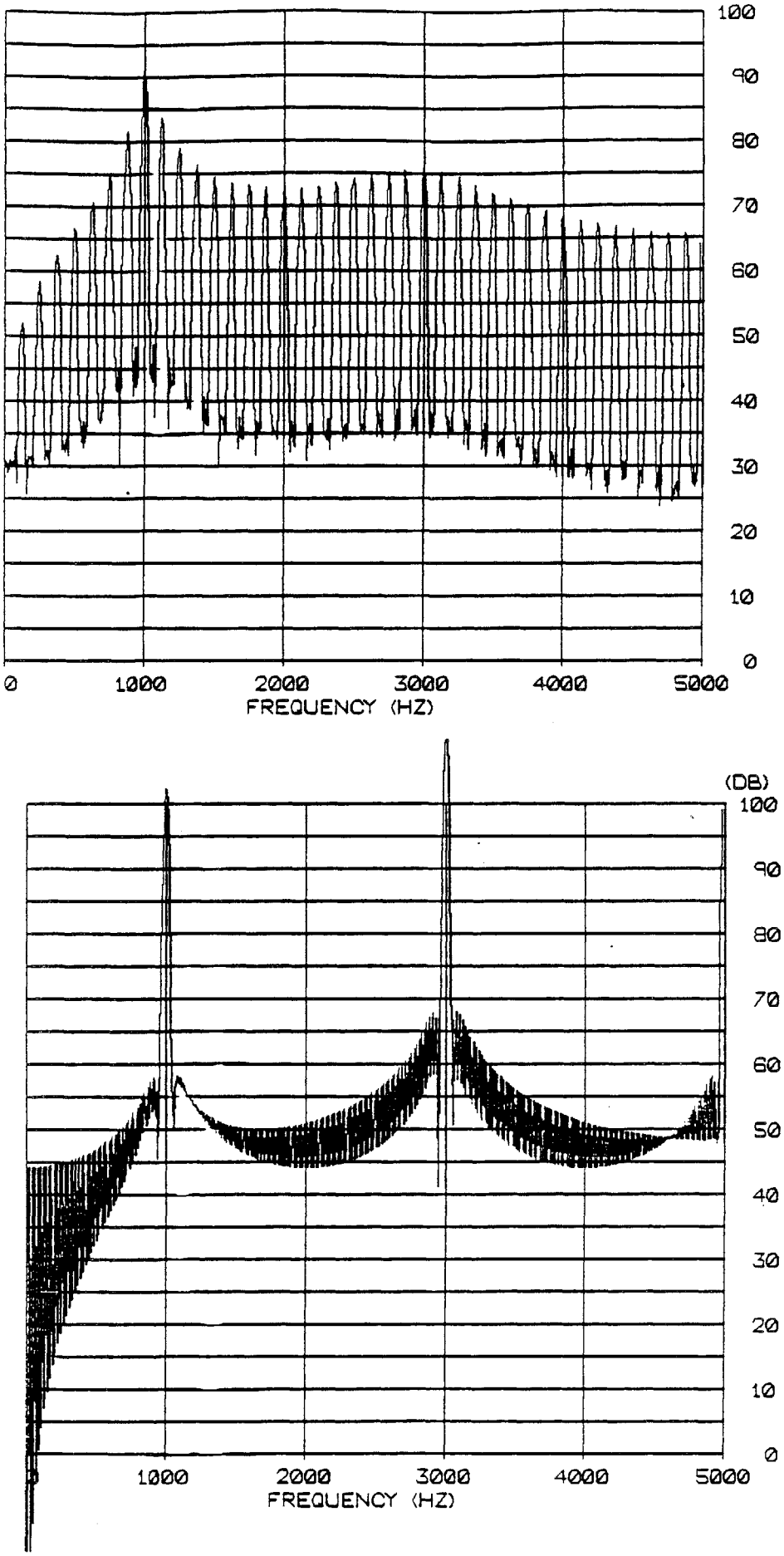
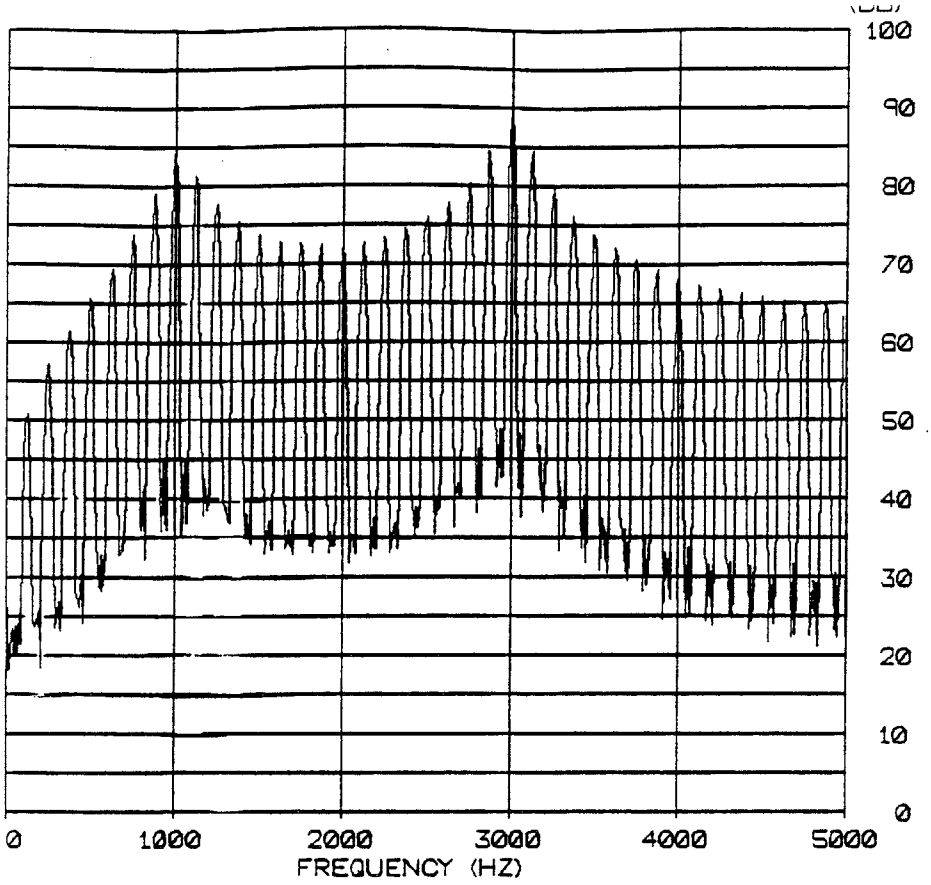


Figure 3.32 As in Figure 3.22. First formant frequency is 1000 Hz with 50 Hz bandwidth. Sencond formant frequency is 3000 Hz with 500 Hz bandwidth.



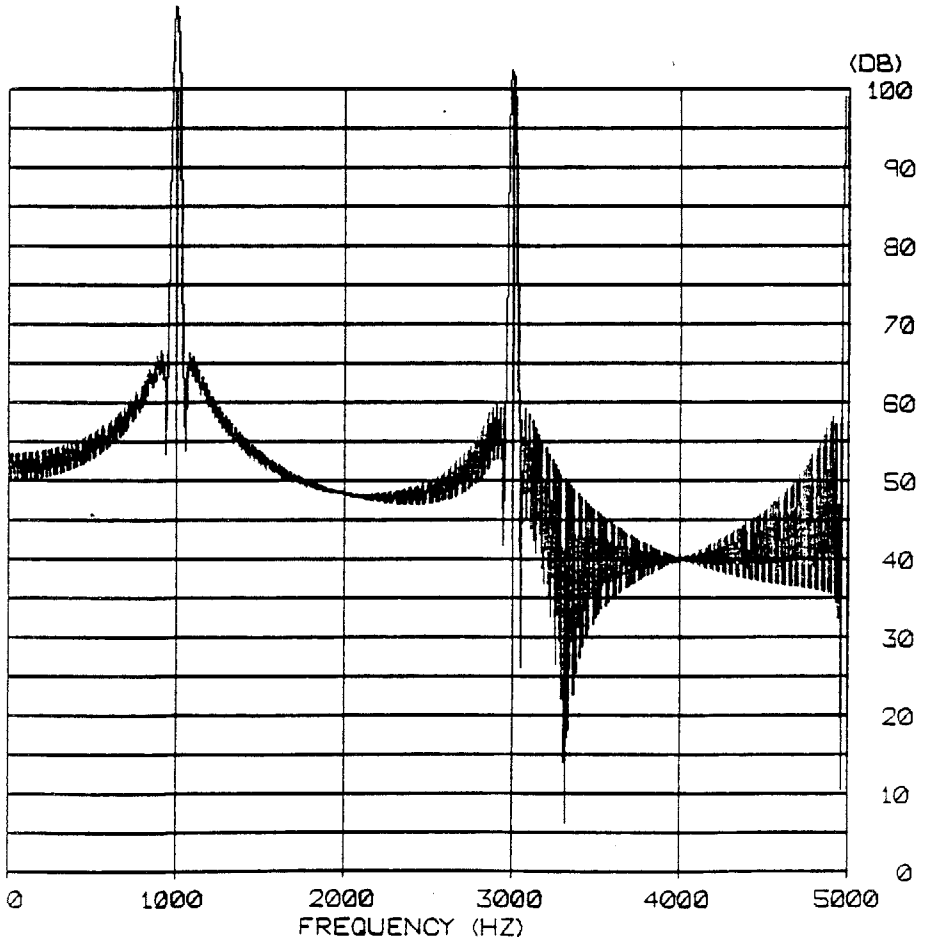
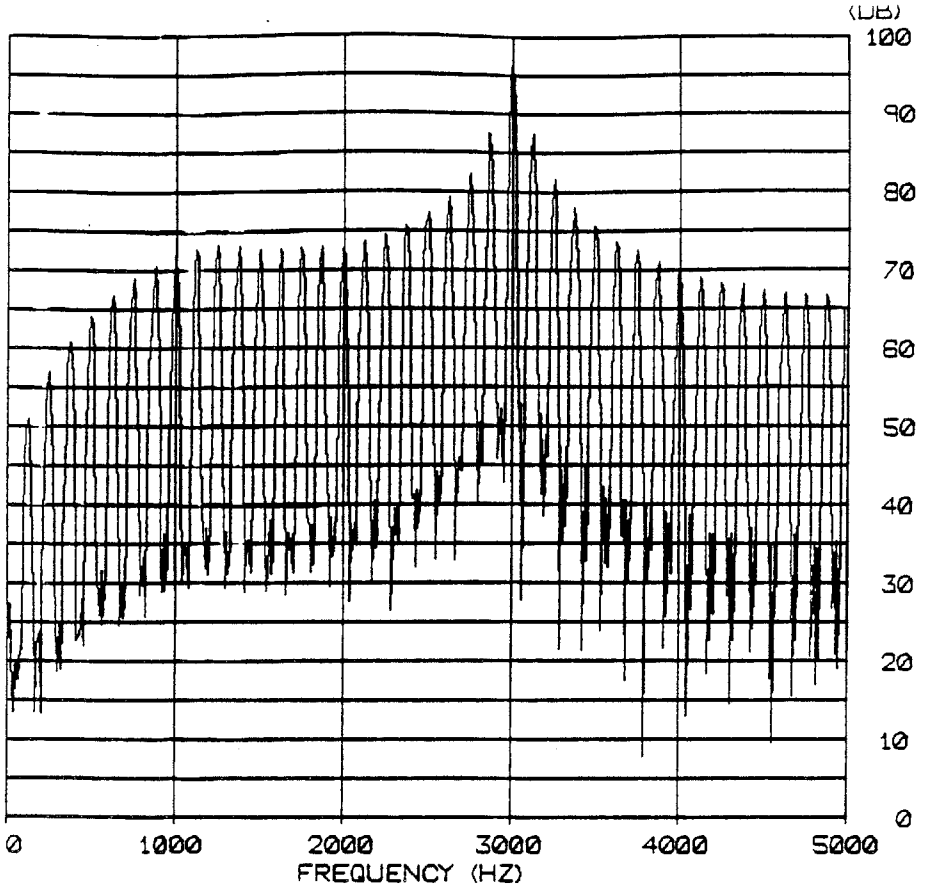


Figure 3.33 As in Figure 3.22. First formant frequency is 1000 Hz with 100 Hz bandwidth. Second formant frequency is 3000 Hz with 100 Hz bandwidth.



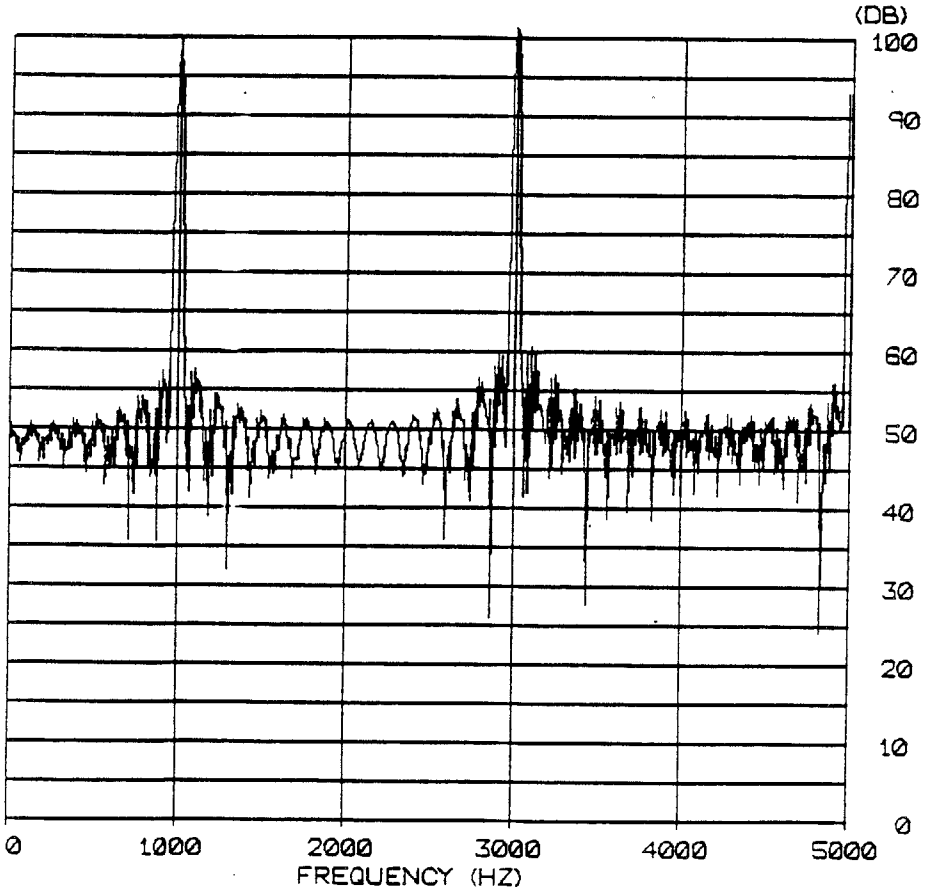


Figure 3.34 As in Figure 3.22. First formant frequency is 1000 Hz with 500 Hz bandwidth. Second formant frequency is 3000 Hz with 50 Hz bandwidth.

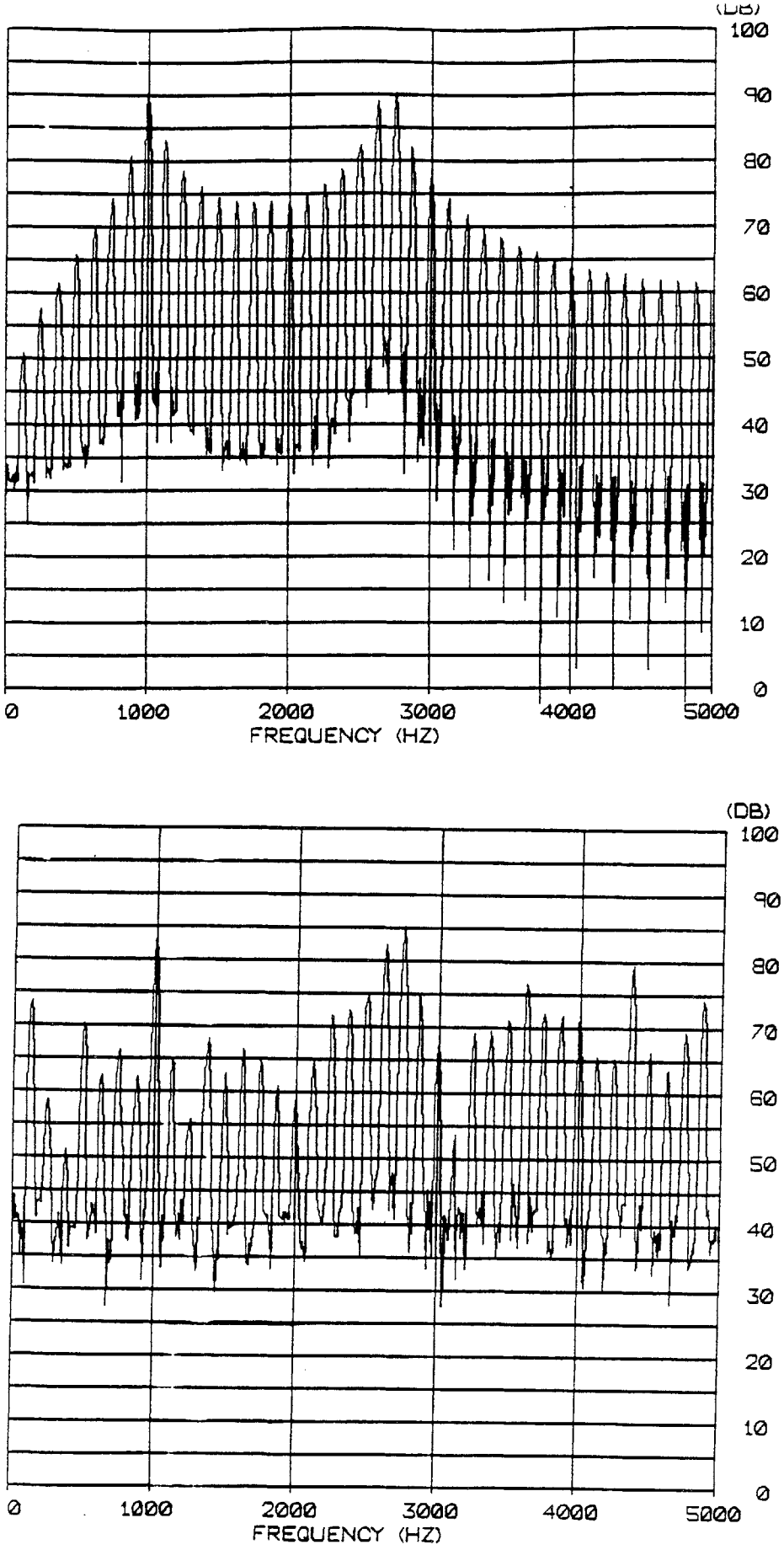
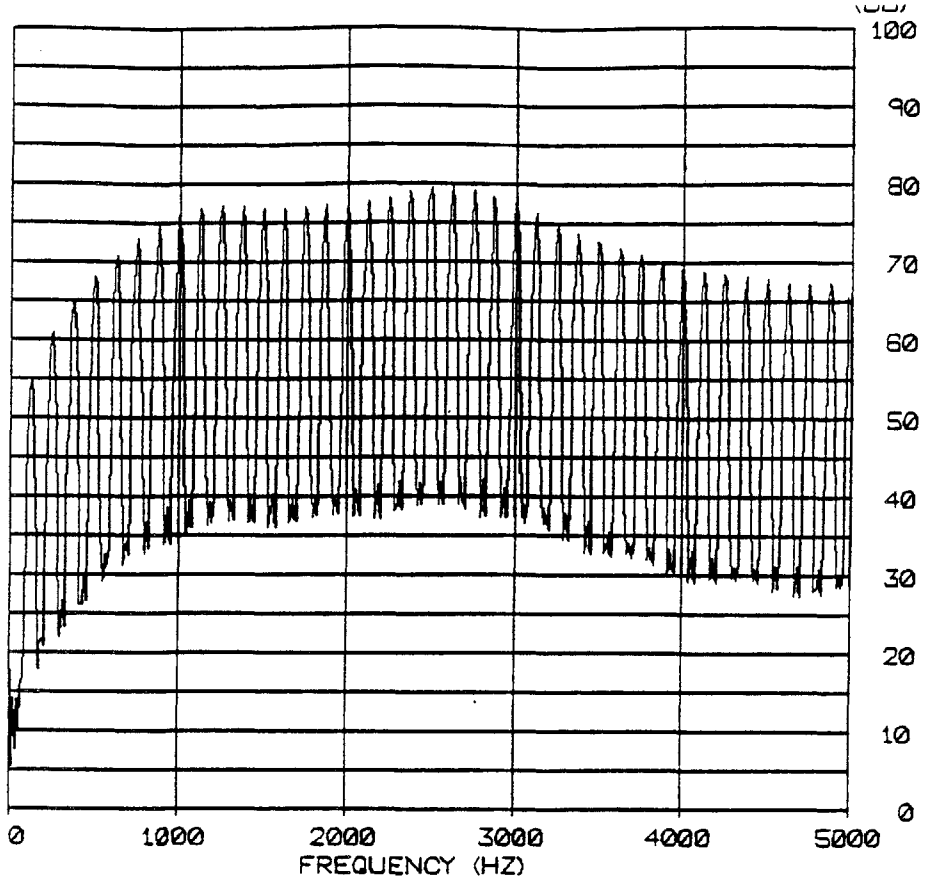


Figure 3.35 As in Figure 3.22. First formant frequency is 1000 Hz with 50 Hz bandwidth. Second formant frequency is 2700 Hz with 50 Hz bandwidth.



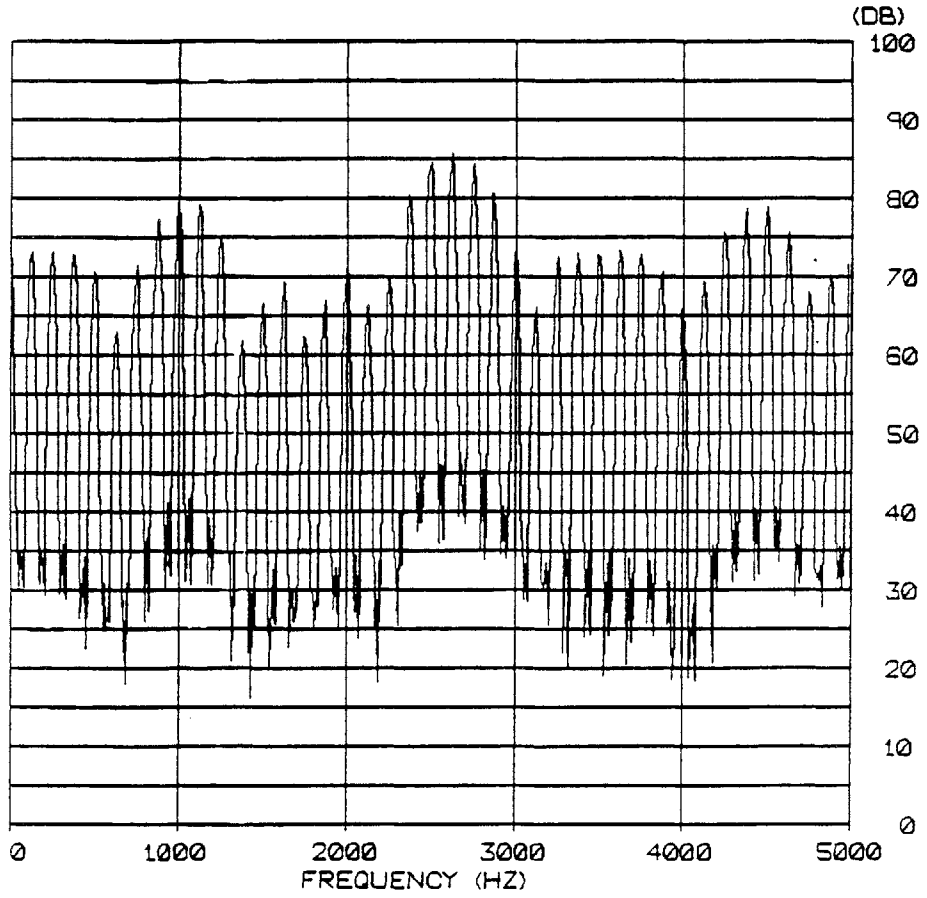
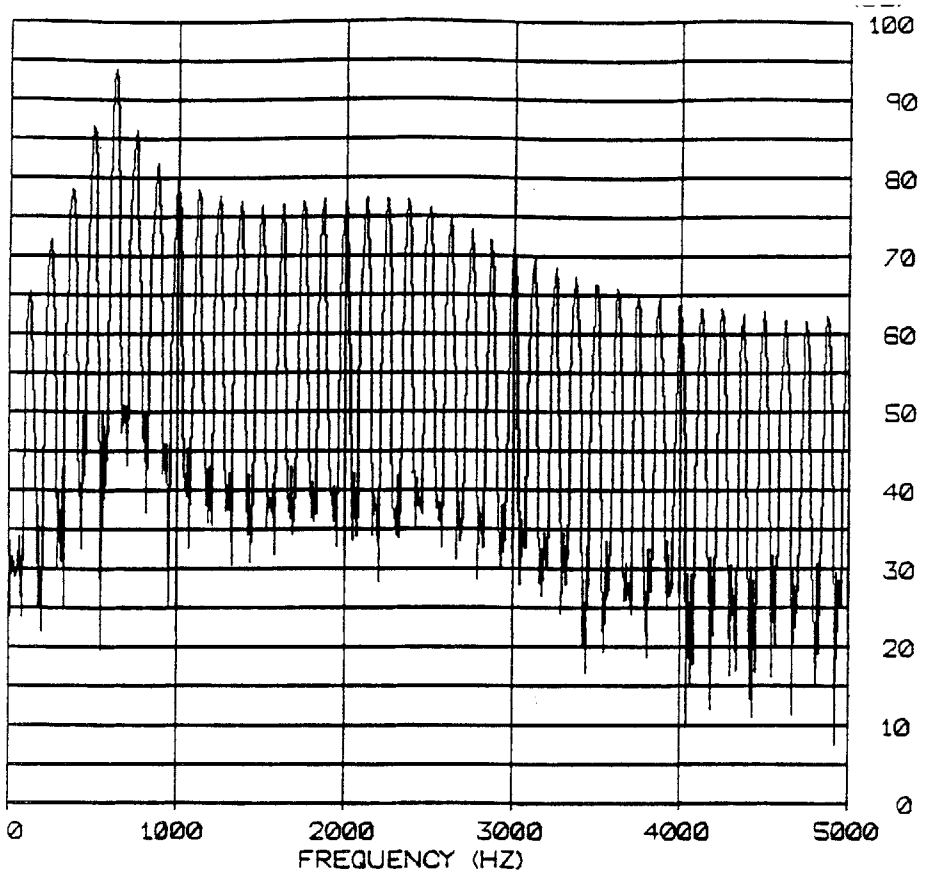


Figure 3.36 As in Figure 3.22. First formant frequency is 1000 Hz with 500 Hz bandwidth. Second formant frequency is 2700 Hz with 500 Hz bandwidth.



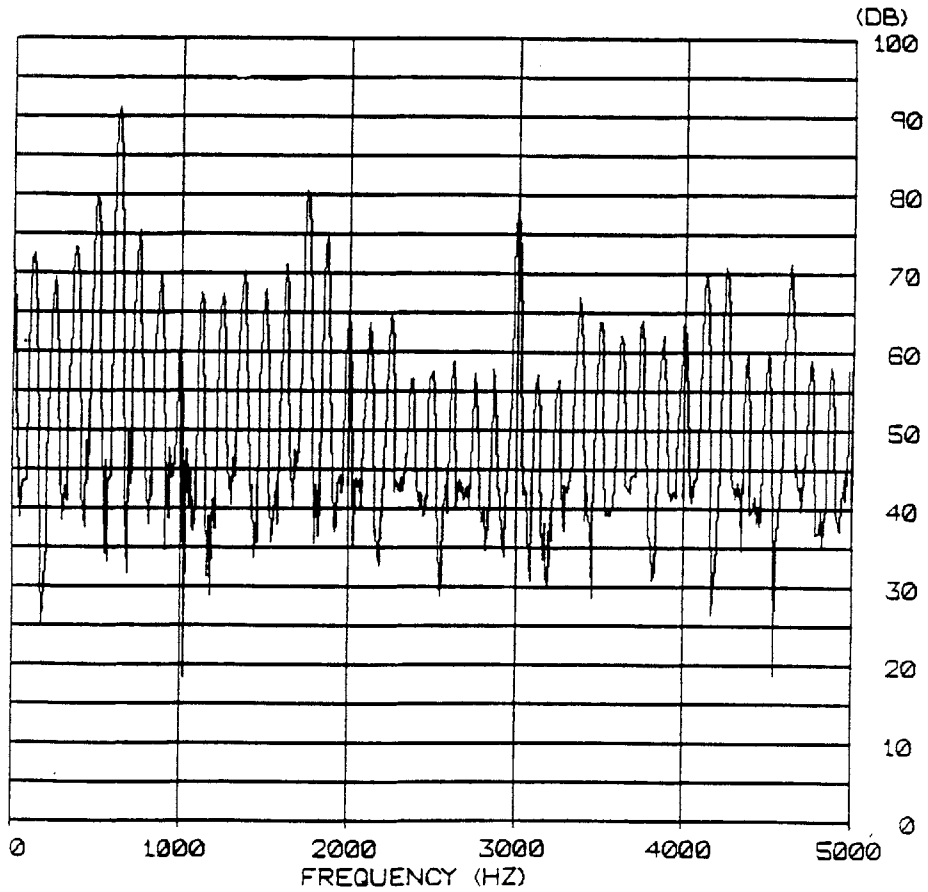


Figure 3.37 As in Figure 3.22. First formant frequency is 600 Hz with 50 Hz bandwidth. Sencond formant frequency is 2300 Hz with 500 Hz bandwidth.

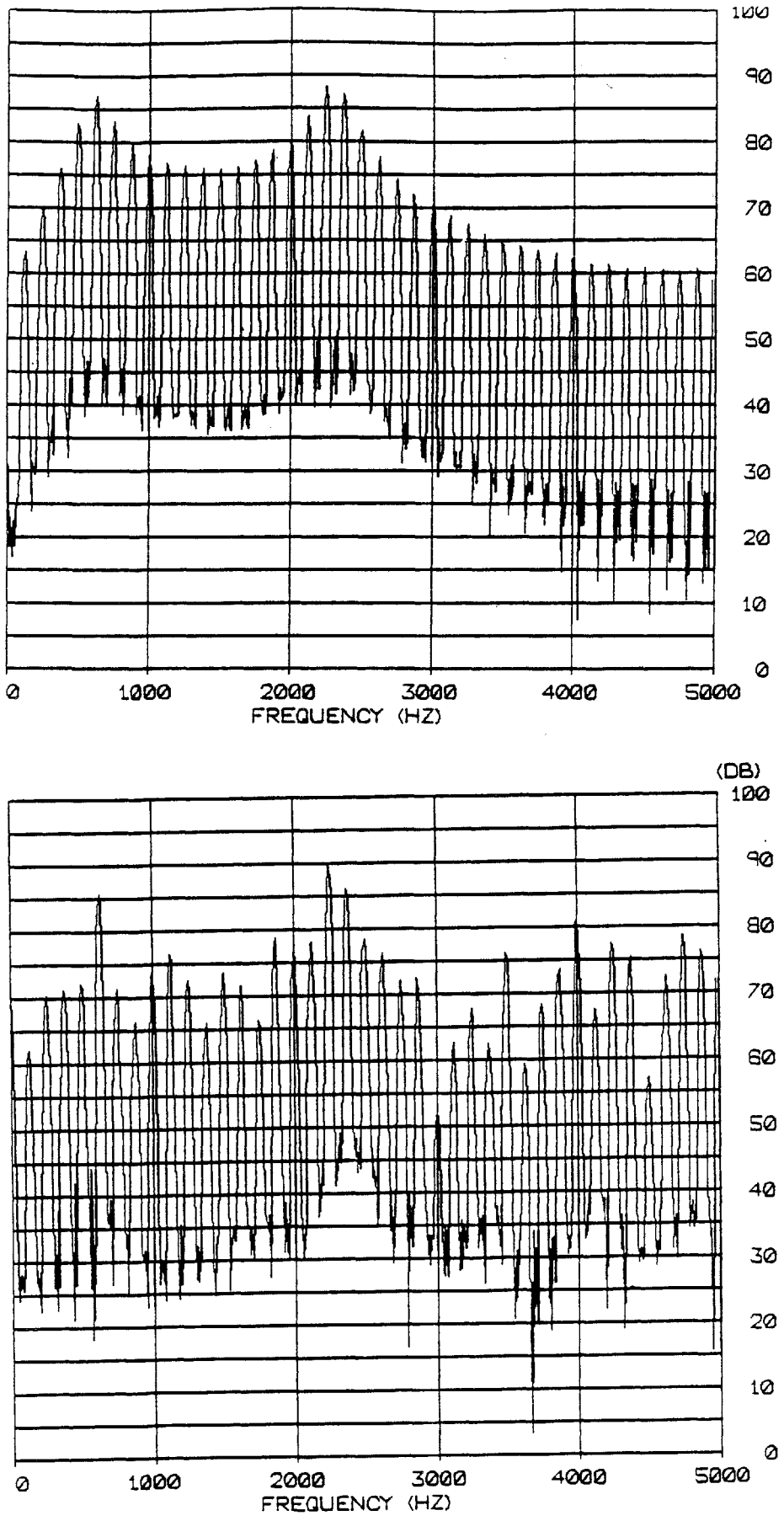
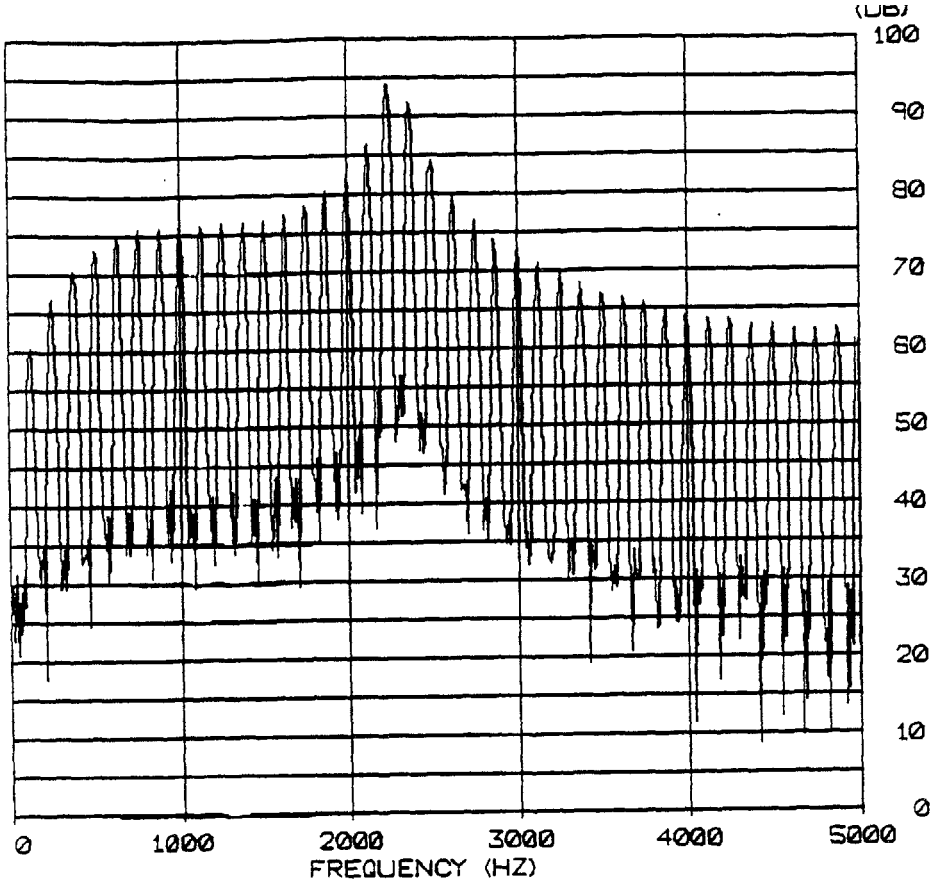


Figure 3.38 As in Figure 3.22. First formant frequency is 600 Hz with 100 Hz bandwidth. Second formant frequency is 2300 Hz with 100 Hz bandwidth.



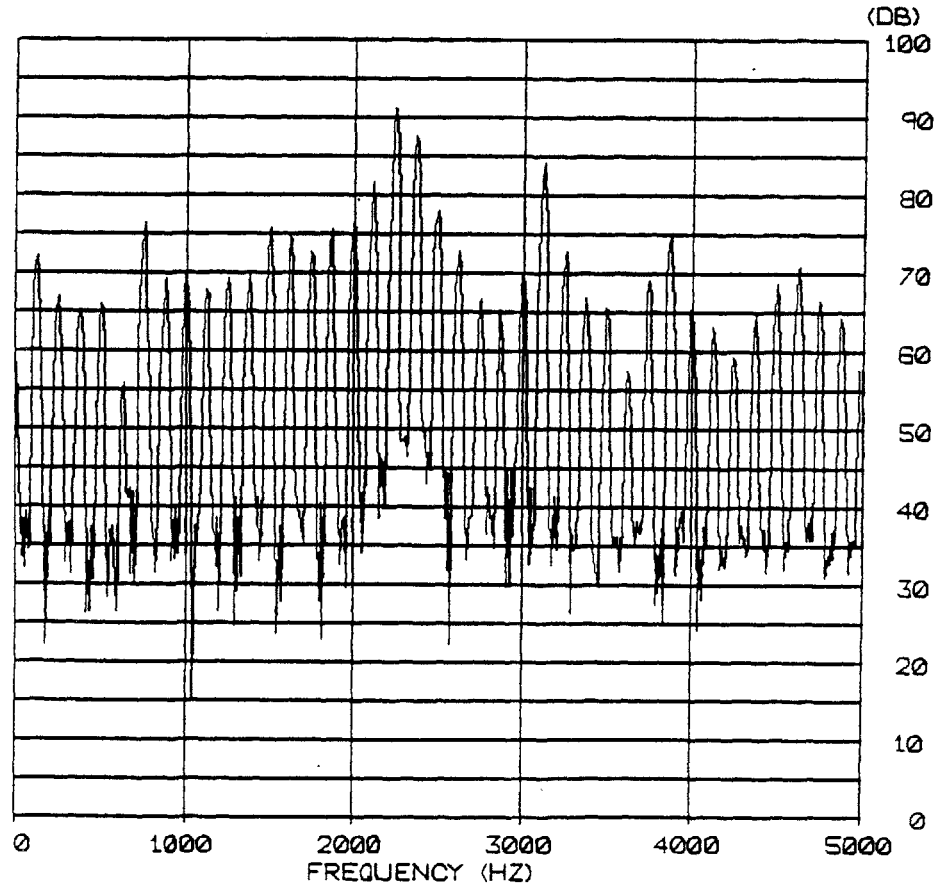


Figure 3.39 As in Figure 3.22. FIrst formant frequency is 600 Hz with 500 Hz bandwidth. Second formant frequency is 2300 Hz with 50 Hz bandwidth.

Chapter 4

Discussion

The present study was designed to assess infinite peak clipping as a means of reducing the amplitude variations in speech levels. The study was based on the following two assumptions about hearing and the perception of speech. The first is that the ear analyzes sound into "critical-band" filters, the outputs of which are relevant to the perception of loudness. The second assumption is that the primary cues to intelligibility lie in the spectral (magnitude) patterns of speech. The investigation was designed to answer questions based on these assumptions: what are the effects of post-filtering speech after infinite-peak clipping, and how can spectral distortion be minimized? The goal was to determine if an amplitude compression system based on infinite peak clipping could be designed which would give maximum compression and minimum spectral distortion, thereby maximizing intelligibility for individuals with sensorineural hearing loss.

4.1 Compression Results

The general result from the study of compression was the insensitivity of the amplitude range of clipped speech to the characteristics of the filters that preceded and followed the clipper. Regardless of which pre-filters and post-filters were used, clipping reduced the range of speech by at least 15 to 20 dB. Output ranges were on the order of 10 to 20 dB.

The general trends of the effects of pre- and post-filtering clipped speech can be understood qualitatively with concepts discussed by DeGennaro et al. (1981). The fluctuations in a bandpass speech envelope can be modeled as consisting of slow and rapid components. The rapid components result from the narrowband filtering (the rate of fluctuations are on the order of the bandwidth) and account for 10-15 dB of the measured range. The slow component (with fluctuations on the order of a few per second) results from variation in the overall intensity and in spectral distributions across different speech sounds. The slow component accounts for about 30 dB of the range. To a first approximation, the effect of infinite peak clipping on the range measured after narrow post-filtering is to remove the slow component. An example is a third-octave band of noise that is slowly modulated over a large decibel range, then clipped and post-filtered by a third-octave filter with the

same center frequency. The zero-crossing pattern is unaffected by the slow modulation, so the clipped noise will exhibit the same envelope variations as if there were no slow modulations.

The preceding analysis applies to the case where the post-filter is in the spectral region that is dominant at the input to the clipper. If this dominant region changes over time, variations in the clipped spectral distributions will contribute to the measured output range.

This discussion aids in interpreting the finding that with no pre-filtering the 16 band ranges were slightly smaller than with a highpass pre-filter (the optimal filter or differentiator), or a bandpass pre-filter (formant filters). With no pre-filtering, the input speech waveform (and hence the sequence of zero-crossings) is dominated by the low-frequency first formant. The harmonics in the higher frequency regions will be relatively constant across the various speech sounds after clipping. If the first-formant is reduced in amplitude by highpass filtering, the input zero-crossing sequence will exhibit greater variability, resulting in greater variation in the clipped spectrum across the different speech sounds.

It was also observed that ranges were reduced further if the pre-filter bandwidth was narrowed to one-third octave or less. This effect can be understood by considering the case of a very small pre-filter bandwidth. The input to the clipper will approximate a tone with slowly varying amplitude and frequency modulation. Since the clipped signal will be frequency modulated with constant amplitude, the output of the third-octave filter will also be a tone with slow frequency modulation and an amplitude modulation resulting from FM-to-AM conversion by the post-filter. Thus, the narrower the input band, the less frequency modulation and consequently the less amplitude modulation there will be after post-filtering.

The finding that the clipped range is not strongly dependent on the bandwidth or slope of the post-filter inplies that the exact characteristics of the ear's critical-band filters need not be known in order to make meaningful measurements of range.

It is interesting to speculate on whether there is a minimimum range of speech levels that can be achieved after (auditory) post-filtering by any compression system. When one considers that the envelope ranges of narrowband noise and harmonic complexes are on the order of 10-15 dB, one must wonder how the variations in the speech envelope could be

further reduced. It would seem that further reduction of these envelope variations could be achieved only in special cases.

4.2 Comparison with AGC Systems

Krieg (1980) studied multiband amplitude compression achieved by independent AGC in each band. The input speech was CVC lists spoken by a male talker, and the compression ratios in the 16 frequency bands were 3 (with a deviation of approximately 0.5 in channels 11 and 16). The input third-octave ranges were 25 to 42 dB, averaging 36.8 dB across all 16 channels (Krieg, 1980, Figure 9). After compression, the ranges were 14 to 38 dB, averaging 24.3 dB across all frequency bands. The actual amount of compression was approximately 12 dB (Krieg, 1980, Figures 11 and 15) and the effective compression ratio was approximately 1.5.

Infinite peak clipping, in the present study, reduced the range of speech by approximately 20 dB, 8 dB more than the AGC system used by Krieg (1980). Assuming that greater intelligibility results from presenting a greater range of speech within the residual hearing area of the listener, then clipping should allow superior intelligibility results when the listener's dynamic range is less than about 20 dB.

4.3 Spectral Modifications

The spectral pattern of clipped unvoiced sounds can be predicted from the arcsin law (Section 3.4). After clipping, spectral peaks will appear at odd multiples of the original spectral peaks. Clipping spreads the spectrum, filling in the low-level region much as a noise floor would. Clipped spectra with two formants have harmonic peaks at not only odd multiples of the formants but at combination frequencies. The lower-frequency formant, when stronger, tends to obscure the higher-frequency formant.

The clipped spectra of voiced sounds are dependent on the phases of the input frequency components, so the arcsin law cannot be aplied. Except for the fact that the clipped harmonic spectrum contains frequency components with the same fundamental spacing as the unmodified spectrum (i.e. there is no counterpart to the "noise floor" of unvoiced sounds), the general results are similar with voiced and unvoiced sounds. Spectral peaks which appear at odd multiples of the formants and combination frequencies after clipping will obscure the low-level formants.

A remarkable aspect of clipping vowels is the degree to which spectral peaks can be enhanced. This finding suggests that infinite peak clipping could be exploited in schemes for

formant detection and tracking.

4.4 Implications for Hearing Aid Design

As stated above, the present study sought to determine the effects of infinite peak clipping in order to minimize both output ranges in critical-band-like filters and spectral distortions. It was found that output ranges are relatively insensitive to post-filtering and slightly dependent on pre-filtering. If minimizing output ranges were the only criterion, there would be either no pre-filtering or narrowband pre-filtering. However, pre-filters like Thomas and Niederjohn's (1968) 1100-Hz highpass filter and Hildebrant's (1982) formant filters produced ranges only a few dB larger than no pre-filter and are known to produce greater intelligibility. The differentiator pre-filter gave the largest ranges of the pre-filters investigated.

The present results on spectral distortions will help to determine a choice of pre-filter. Even though the narrowband filters produce the greatest amount of compression, the relative spectral qualities among frequency bands would be lost in such a system. For multiple-formant sounds it would seem that a pre-filtering scheme that separates the formants, like Hildebrant's (1982), would be beneficial for preserving the formants (without losing the relationship between bands)

and avoiding interactions in the clipper. However, for single-peak sounds, this scheme may result in spurious output peaks, or a less well-defined peak than in the input. In this respect, a single-channel clipper with highpass pre-filtering might prove superior.

Of course, the particular loss of the individual should also be considered. For example if the individual depends only on first formant cues, and has virtually no second formant hearing, then pre-filtering with an optimal filter will hinder intelligibility by decreasing the saliency of the cues.

4.5 Recomendations for Future Work

Future work should test intelligibiltiy with a 3-band clipper and optimal filter/clipper systems on hearing-impaired listeners with small dynamic ranges. Ideally, results could be compared with performance on other amplitude compression systems using the same subjects. Examination of the types of speech errors for both hearing-impaired and normal subjects should help to determine the next course of action.

REFERENCES

Braida, L.D., et al. (1979). "Hearing Aids--A review of Past Research on Linear Amplification, Amplitude Compression, and Frequency Lowering," ASHA Monograph No. 19

Coln, M.C.W. (1979). "A Computer Controlled Multiband Amplitude Compressor," Master's Thesis, Mass. Inst. Tech., Cambridge, MA.

De Gennaro, S., Braida, L.D., Durlach, N.I. (1981). "A Statictical Analysis of Third-Octave Speech Amplitude Distributions," Paper presented at the 101st meeting of the Accoustical Society of America, Ottawa, Ontario, Canada

De Gennaro, S. Krieg, K.R., Braida L.D., Durlach, N.I. (1981). "Third-Octave Analysis of Multichannel Amplitude Compressed Speech," Proceedings of IEEE Inter. Conf. on Acoust., Speech and Signal Processing, Atlanta, Georgia

De Gennaro, S. (1982). "An Analytical Study of Syllabic Compression for Severely Impaired Listeners," PhD Thesis, Mass. Inst. Tech., Cambridge, MA.

Dunn, H.K., White, S.D. (1940). "Statistical Measurements on Conversational Speech," JASA 11, pp278-288

Fawe, A.L. (1966). "Interpretation of Infinitely Clipped Speech Properties," IEEE Trans. on Audio and Electroacoustics 14, pp178-183

Green, D.M., Swets, J.A. (1966). Signal Detection Theory and Psychophysics, (John Wiley and Sons, New York)

Guidarelli, G. (1981). "Intelligibility and Naturalness Improvement of Infinitely Clipped Speech," Acoustics Letters 4, No.7

Hildebrant, E.M. (1982). "An Electronic Device to Reduce the Dynamic Range of Speech," Bachelor's Thesis, Mass. Inst. of Tech., Cambridge, MA

IEEE (1969). "Recommended practice for speech quality measurements." IEEE Trans. Audio Electroacoust. 17, pp225-246

Kac, M., Siegert, A.J.F. (1947). "On the Theory of Noise in Radio Receivers with Square Law Detectors," J. Applied Physics 18, pp383-397

Krieg, K.R. (1980). "Third Octave Band Level Distributions of Amplitude Compressed Speech," Bachelor's Thesis, Mass. Inst. of Tech., Cambridge, MA.

Licklider, J.C.R. (1946). "Effects of Amplitude Distortion upon the Intelligibility of Speech," J. Acoust. Soc. Am. 18, pp429-434

Licklider, J.C.R., Bindra, D., Pollack, I. (1948). "The Intelligibility of Rectangular Speech-Waves," Am. J. Psychology 61, pp1-20

Licklider, J.C.R., Pollack, I. (1948). "Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech," JASA 20, pp42-51

Papoulis, A. (1965). Probability, Random Variables and Stochastic Processes, (McGraw-Hill, New York)

Pollack, I. (1952). "On the Effect of Frequency and Amplitude Distortion on the Intelligibility of Speech in Noise," JASA 24, pp538-540

Rabiner, L.R., Schafer, R.W. (1978). Digital Processing of Speech Signals, (Prentice-Hall, New Jersey)

Thomas, I.B., Niederjohn, R.J. (1970). "The Intelligibility of Filtered-Clipped Speech in Noise," J. Audio Engin. Soc. 18, pp299-303

Thomas, I.B., Sparks, D.W. (1971). "Discrimination of Filtered/Clipped Speech by Hearing-Impaired Subjects," JASA 49, pp1881-1887

Appendix A

Probability Densities of Test Signals

Narrow-band Noise

The probability density for the random variable, V, generated by squaring and then lowpass-filtering a band of noise centered at a high frequency was calculated by Kac and Siegert (1947) to be

$$P(V) = \sum_{s=1}^{\infty} \frac{\exp(-V/\lambda s)}{\lambda s \prod_{i \neq s} (1 - \lambda i/\lambda s)} \qquad V>0 \qquad (A1)$$

where λi is the ith eigenvalue in the solution of an integral equation. For the specific case where the input noise spectrum is that of a "simple-tuned" circuit and the lowpass filter is a simple RC-lowpass filter, the eigenvalues are found from the zeroes of the (R-1)th order Bessel function, where R is the ratio of the input noise bandwidth to the lowpass filter's -3 dB bandwidth.

$$J_{R-1}[2(R/\lambda i)^{1/2}] = 0. \qquad (A2)$$

Probability density functions were computed for values of R between 0 and 1 (Figure A1), and values between 1 and 8 (Figure A2). The case of R=0 corresponds to that of a Rayleigh random variable squared, or an exponentially-distributed variable. The range was computed to be the decibel difference between the the two values, V10 and V90, at which the cumulative distribution function is equal to 0.10 and 0.90, respectively. The range values are graphed in Figure 2.2.

Tones

An input sine wave with unit amplitude and cyclic frequency $f=w/2\pi$ is considered in this section. Squaring and lowpass filtering of this signal results in the variable

$$V = \frac{1 - |H(2f)| / \cos(4\pi ft)}{2} \qquad (A3)$$

where

$$H(f) = 1/\sqrt{1+R^2}$$

and R is the ratio of f to the -3 dB bandwidth of the RC-lowpass filter.

Because of symmetry in the time-wave, the calculations can be limited to the range for $\alpha = 4\pi ft$ of $0 \leq \alpha \leq \pi$. The cumulative distribution function for V is found by integrating the probability density function on α (which is assumed uniformly distributed) between the limits determined by the inverse function,

$$\alpha = F^{-1}(V) = \cos^{-1}\left[\frac{(1-2V)}{|H(2f)|} \right] . \qquad (A4)$$

The cumulative distribution function is derived below.

$$Pr(V \leq Vo) = Pr(\alpha < F^{-1}(Vo))$$

$$= \int_0^{\cos^{-1}\left[\frac{(1-2V_o)}{|H(2f)|}\right]} d\alpha/\pi$$

$$= \frac{\arccos\left[(1-2V)/|H(2f)|\right]}{\pi} \qquad (A5)$$

The two values, V10 and V90, at which the cumulative distribution is equal to 0.10 and 0.90, were calculated from equation A5. The range, which is the decibel difference of these two values, is graphed as a function of R in Figure 2.3.

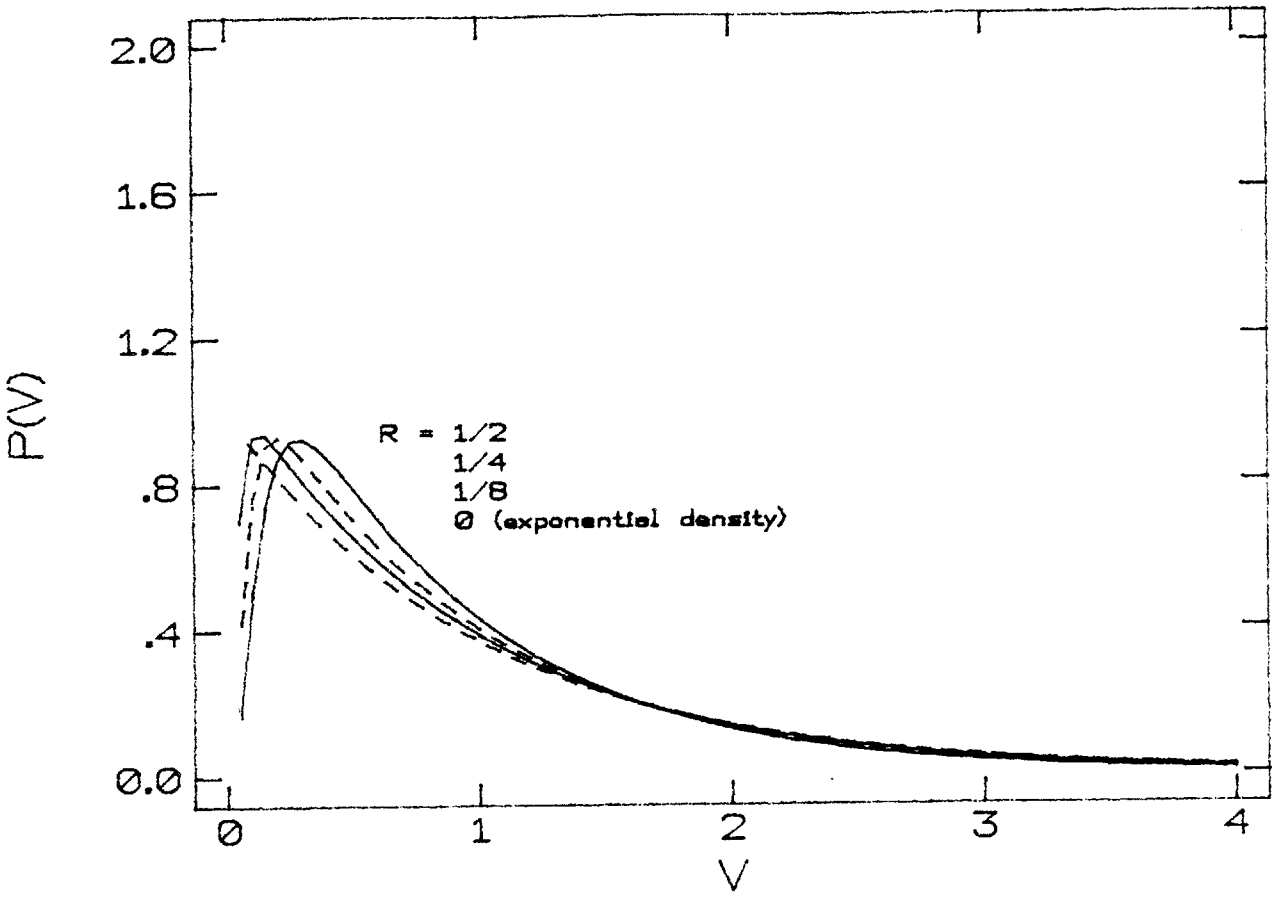


Figure A1 Probability density functions for narrowband noise after squaring and lowpass filtering for R less than 1. R is the ratio of the input noise bandwidth to the lowpass filter's -3dB cutoff.

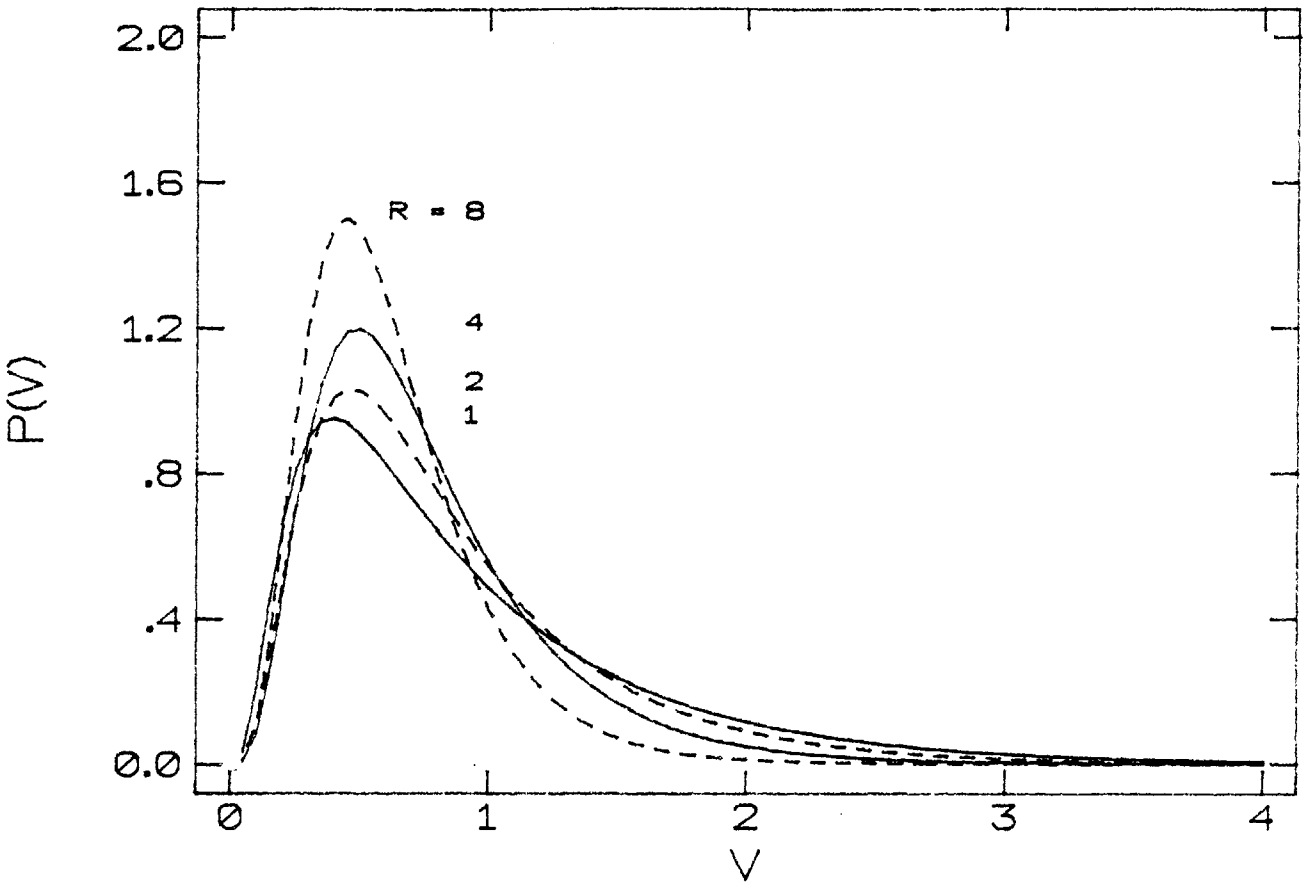


Figure A2 Probability density functions for narrowband noise after squaring and lowpass filtering for R greater than or equal to 1.

APPENDIX B

10 Harvard sentences used as input for experiments. Speaker is male.

1. A large size in stockings is hard to sell.
2. The birch canoe slid on the smooth planks.
3. Glue the sheet to the dark, blue background.
4. It's easy to tell the depth of a well.
5. These days a chiken leg is a rare dish.
6. Rice is often served in round bowls.
7. The juice of lemons makes fine punch.
8. The box was thrown beside the parked truck.
9. The hogs were fed chopped corn and garbage.
10. Four hours of steady work faced us.