Paper presented at XVIIIth ANNUAL CONFERENCE OF INDIAN ASSOCIATION OF BIOMEDICAL SCIENTISTS (IABMS)
New Delhi, India, Oct. 22-24, 1997

Splitting of Speech Signal by Critical Band Filtering for Bilateral Sensorineural Hearing Impairment

D.S. Chaudhari[1] *and P.C. Pandey*[2]
[1]School of Biomedical Engineering, Indian Institute of Technology, Bombay
Powai, Mumbai - 400 076. E-mail: deven@ee.iitb.ernet.in
[2]Department of Electrical Engineering, Indian Institute of Technology, Bombay
Powai, Mumbai - 400 076.
Phone (+91-22) 576-7445 Fax (+91-22) 579-3707 E-mail: pcpandey@ee.iitb.ernet.in

ABSTRACT

Sensorineural hearing loss causes reduction in frequency resolving capacity of the ear due to spread of masking of frequency components by adjacent frequency components. We have carried out experimental evaluation of splitting speech into two complementary signals on the basis of frequency and presenting it binaurally for increasing the speech intelligibility. In this scheme, input speech signal is filtered into two signals by using a bank of critical band filters where odd numbered critical bands are presented to one ear and even numbered ones to the other. Thus, the effect of spectral masking in the cochlea on speech information reaching the auditory cortex is reduced.

The process of splitting speech was done off-line, and listening tests were performed using twelve English consonants in vowel-consonant-vowel context with /a/ as in `father', on four normal hearing subjects with simulated sensorineural hearing loss. Simulation of sensorineural impairment in subjects with normal hearing was done by addition of white masking noise to the speech signal at different SNRs. The test results were analyzed by comparing the recognition scores for various unprocessed speech to those with processed speech. Under adverse masking noise condition, the recognition score is found to increase by 10 percent. The average response time over four subjects is also found improved by this scheme.

Keywords: Hearing impairment, masking, filter bank, dichotic presentation

INTRODUCTION

People who suffer from significant hearing impairment can be helped with modern medical and surgical procedures and instrumentation. The conventional hearing aids provide the amplification of speech. Many such aids are also provided with filtering and amplitude compression. However, these aids don't improve the speech presentation for person suffering from sensorineural impairment (arising due to loss of hair cells in the inner ear and auditory nerve).

The sensorineural impairments are characterized by high frequency hearing loss, increase in threshold of hearing, compression in dynamic range, severity of temporal masking, and loss of spectral resolution due to spread of masking. Various attempts have been made to solve these problems. Some of the techniques currently being investigated are based on signal processing schemes such as spectral transposition, speech enhancement by properties of " clear speech" . These are likely to improve the performance of hearing aids for persons with residual hearing as well as that of other sensory aids like cochlear prosthesis and vibro-tactile aids used by profound hearing impaired.

It is known that the ability of humans to binaurally receive and perceptually combine signals from two ears improves recognition of speech under adverse listening conditions (Moore, 1982). There is possibility that splitting speech into two complementary parts on the basis of frequency and presenting these dichotically might increase speech intelligibility. The hearing aid based on this principle, can be helpful to bilateral sensorineural hearing impaired people with some residual hearing.

The objective of this investigation is to split the speech in two signals, with complementary spectra, for binaural dichotic presentation as a possible solution to problem of spectral masking. The study was carried out by processing digitized speech, and listening tests were conducted using normal hearing subjects with simulated sensorineural hearing loss.

I. METHOD

A. Subjects and stimuli

Four normal hearing subjects (three male and one female) in the age group of 21 to 30 years participated in the experiments. The subjects had puretone threshold of better than 20 dB HL at audiometric test frequencies from 125 Hz to 10 kHz. The subjects were from different parts of India. They had no difficulty in clearly recognizing the test stimuli.

In order to minimize the contribution of linguistic factors and maximize the contribution of acoustic factors nonsense syllables were used for stimuli. Twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ were used in vowel-consonants-vowel (VCV) context with the vowel /a/ as in `father'.

B. Apparatus

Speech signal acquisition was done with 12-bit resolution at 10 k samples/sec. The set-up for acquisition and analysis of speech is as shown in Figure 1. The signal from the microphone goes to an amplifier, attenuator, frequency weighting filter, and buffer amplifier (microphone B & K 4176 alogwith sound level meter B & K 2235). The output from the buffer amplifier is low pass filtered (f_c = 4.8 kHz), and given to ADC of TI-TMS 320C25 based DSP board interfaced to PC. The acquired speech segment can be output using DAC of the DSP board. The spectrographic analysis (Thomas et al., 1994; and Baragi 1996) of acquired speech is done using this set-up where FFT analysis is performed on DSP board and display is done on PC using its graphics adapter. The processing of selected speech segments is done off line on the PC. The processed sounds are again studied using spectrographic analysis.

In our listening tests experiments, we need two output channels for binaural dichotic presentation, which are not available on DSP board. For listening test experiments, a computerised test administration as shown in Figure 2 is used. It includes a terminal connected to a asynchronous serial port (RS-232C) and a PC based data acquisition card (PCL-208 - from Dynalog MicroSystem Limited, Bombay) which has two output ports. The subject terminal (placed in acoustically isolated chamber) was used for displaying the response choices on its screen and for obtaining subject responses from its keyboard. For presentation, the stimuli were outputted at a rate of 10 k samples/sec through two D/A ports of data acquisition card. The D/A outputs were passed through a pair of smoothing low pass filters and a pair of power amplifiers. The presentation sound pressure level in dB SPL was adjusted using artifi-

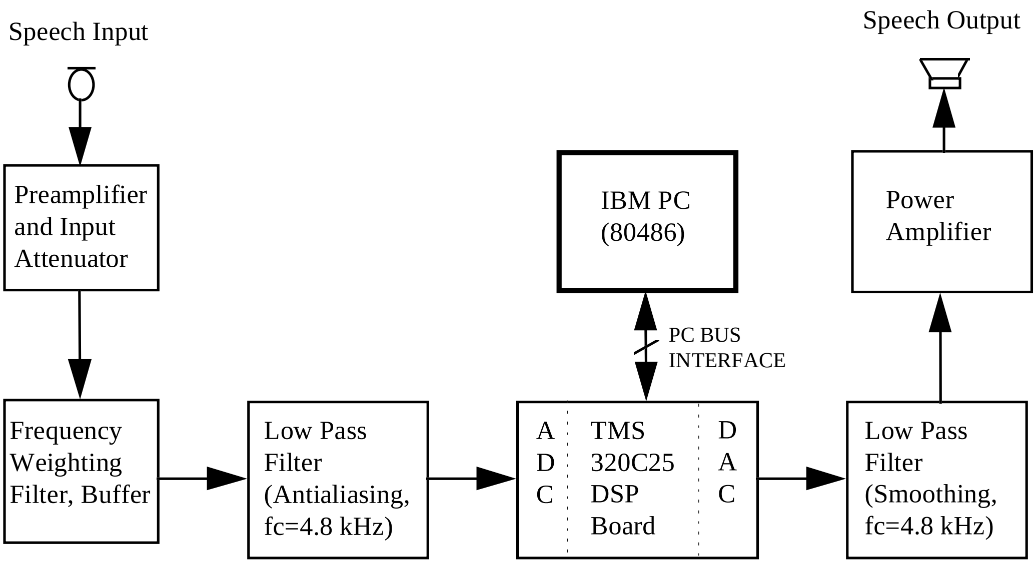

FIGURE 1. Experimental setup for acquisition and analysis of speech segment.

cial ear (B & K, 4153). The stimuli were presented to the ears through a pair of Telephonics TDH-39P headphones.

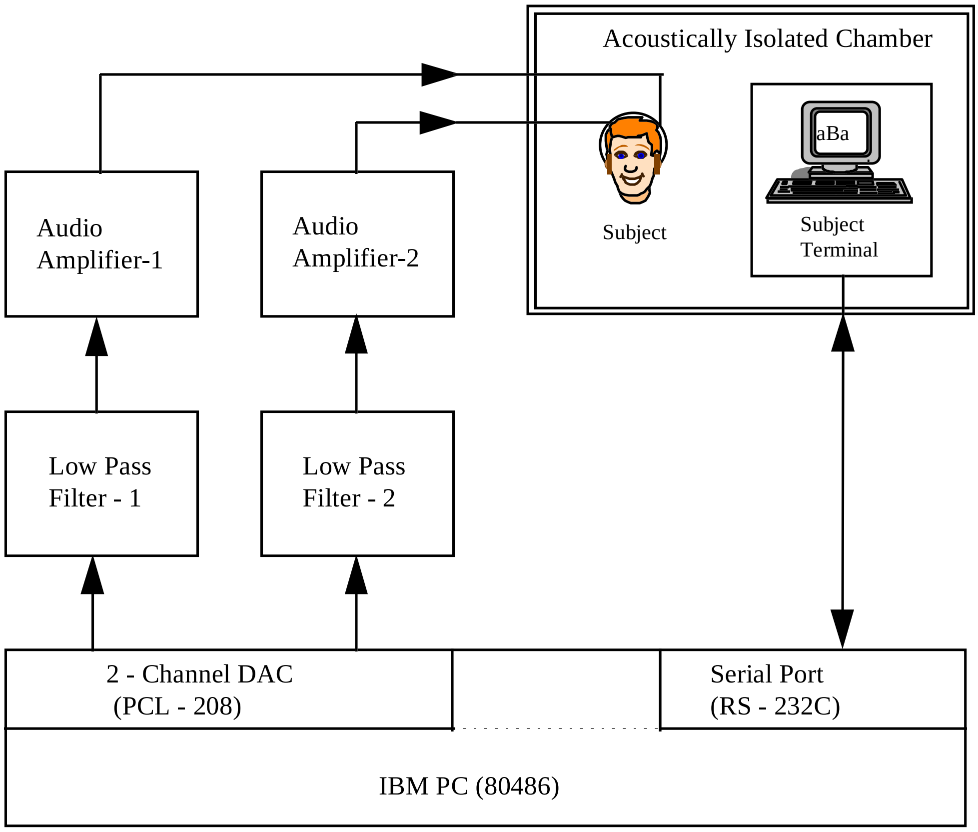

FIGURE 2. Experimental setup for listening tests.

C. Presentation Level and Simulation of Sensorineural Hearing Loss

The effect of presentation level on consonant identification for normal hearing subjects have been studied by several investigators (Simon, 1978; Dorman and Dougherty, 1981). They noticed that identification performance declines at very low (<35 dB SPL) and very high (>90 dB SPL) presentation levels. In one of the studies, the stimuli were presented at 75, 80, or 85 dB SPL (Leek and Summers 1996). In our study, presentations were done at the almost comfortable listening level for the subject, which ranged from 75 to 85 dB SPL.

On the basis of various studies reported on in the literature (Fletcher, 1952; Lochner and Burger, 1961; DeGennaro et al., 1981; Jesteadt, 1997), it appears that broadband noise can be used for simulating various aspects of sensorineural hearing loss in normal hearing subjects for speech reception. We have used Gaussian noise bandlimited to the band of the speech signal as masking noise and varied the signal-to-noise ratio for varying the severity of simulated loss.

D. Procedures

1. Processing of speech signal

The objective of this investigation was to study possible solution to the problem of spectral masking in case of sensorineural hearing impaired subjects. In view of this, acquired speech was split based on critical multiband filtering. For this investigation, critical bandwidth was chosen as per the auditory filter bandwidths (Zwicker, 1961). The speech was filtered and divided into two parts in such a way that frequency components lying within a critical

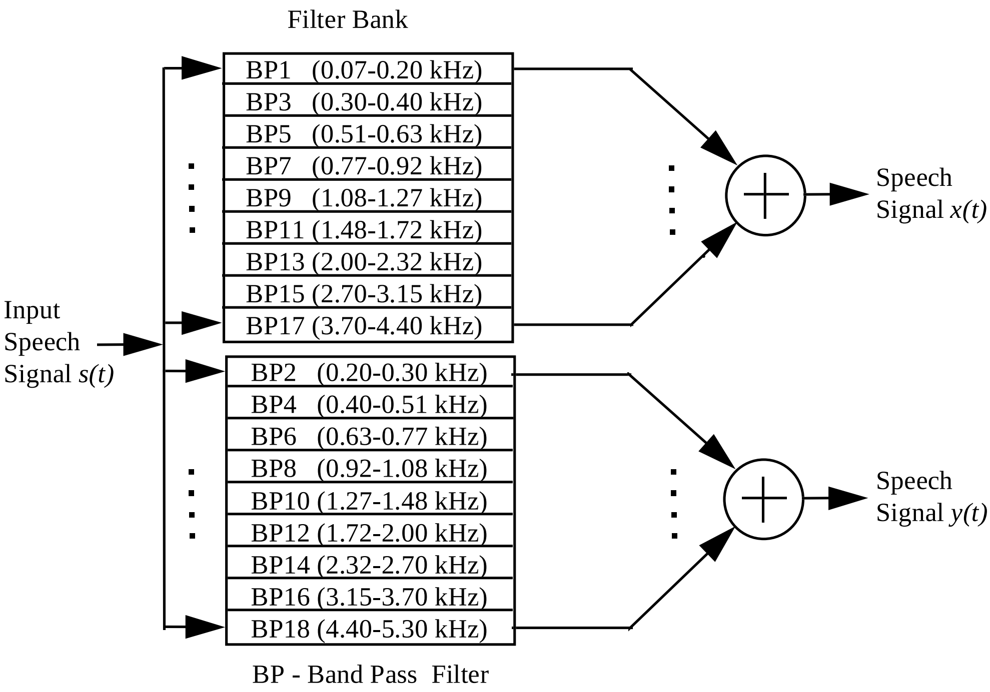

FIGURE 3. Splitting of speech based on multiband filtering.

band are in one part, components lying in next non-overlapping critical band are in second part, component of third non-overlapping critical band in first part and so on, as shown in Figure 3. The odd numbered filter outputs were fed to one ear and even numbered filter outputs were fed to the other ear.

The testing was done on normal subjects, with simulated sensorineural impairment. The sensorineural impairment was simulated by adding broadband noise to speech signal at 5 SNR conditions of ∞, 6, 3, 0, and -3 dB. The addition of the noise in each case has been done in such a way that the overall sound level remains unchanged. The masking noise used is a Gaussian white noise and was acquired using the set-up as shown in Figure 1. The processing of speech signal and noise addition was done off-line.

2. Listening tests

Listening tests were carried out for finding the confusions among the set of twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ in the VCV context with vowel /a/. These tests happen to be repetitive and time consuming, and hence conducted using an automated test administration system as shown in Figure 2 and described earlier, with the subject seated in the acoustically isolated chamber. The subject is briefed about the procedure. In a given test, each stimuli is presented five times, in a randomized order. For each presentation, the subject should respond by pressing a key. The location of response choices is also randomized. Each test uses a different order for presentations. In addition to the test response, the response time is also recorded. Before each test session, the subject can listen to the stimuli any number of time in any order in order to become fully familiar with them. The tests can be with feedback or without feedback. At the end of each session, the confusion matrix, and response time statistics are stored. Generally no feedback results were considered for analysis. The subject performance can vary with exposure to stimuli and fatigue, hence the stabilized scores with variation of 10 % were considered for analysis.

The speech signal was presented binaurally at the most comfortable listening level. Utmost care was taken to keep loudness level constant throughout the testing of an individual subject, since the clarity of voice also depends on loudness. For each subject tests were administered for (a) unprocessed speech presented to the left and to the right ears and (b) processed speech dichotically presented to the two ears. For each case the tests were carried out at five SNR conditions randomized across the test session, and these sessions were spread over two months. A typical test session consisted of two-to-four tests. Subjects were asked to also provide a qualitative assessment of the test stimuli.

II. RESULTS AND DISCUSSION

Listening tests were conducted with four subjects (three male and one female) with normal hearing. The recognition scores obtained from confusion matrices are given in Table 1, for the four subjects and also averaged across the subject. Paired t-test (Snedecor and Cochran, 1980) for testing the significance of differences of averaged scores for the unprocessed versus processed speech were carried out and these are also tabulated along with the scores. The scores for one subject SAK and scores averaged across the subject are plotted in Figure 4, and Figure 5 respectively.

Under no noise condition, all the subjects have nearly perfect scores with both unprocessed and the processed speech. However, all the subjects showed less response time (results not included here) for processed speech, indicating an improvement in listening condition with processing.

TABLE 1. Recognition scores for the 12-consonant listening tests and corresponding paired t-test significance level, for unprocessed versus processed speech for individual subjects, and averaged across the subjects.

Subject	SNR	Recognition Score: For Unprocessed Speech mean	s.d.	For Processed Speech mean	s.d.	Relative Score Improvement (%)	Test of Difference (two-tailed) t	p
SAK	No Noise	98.3	0.0	100.0	0.0	1.7	∞	< 0.001
	SNR = 6 dB	96.7	0.0	100.0	0.0	3.4	∞	< 0.001
	SNR = 3 dB	92.8	2.5	98.3	0.2	5.9	3.79	< 0.05
	SNR = 0 dB	85.0	5.0	95.6	2.5	12.5	3.28	< 0.05
	SNR =-3 dB	83.3	0.0	90.0	0.0	8.0	∞	< 0.001
MSC	No Noise	100.0	0.0	100.0	0.0	0.0	∞	< 0.001
	SNR = 6 dB	97.8	0.9	100.0	0.0	2.2	4.23	< 0.025
	SNR = 3 dB	96.7	0.0	100.0	0.0	3.4	∞	< 0.001
	SNR = 0 dB	96.1	1.9	98.3	0.0	2.2	2.01	< 0.2
	SNR =-3 dB	90.0	1.7	98.3	0.0	9.2	8.46	< 0.005
CKS	No Noise	98.9	1.0	100.0	0.0	1.1	1.85	< 0.2
	SNR = 6 dB	97.8	0.9	100.0	0.0	2.2	4.23	< 0.025
	SNR = 3 dB	96.1	1.9	98.9	1.0	2.9	2.26	< 0.2
	SNR = 0 dB	92.2	0.9	95.6	1.0	3.7	4.38	< 0.025
	SNR =-3 dB	90.0	1.7	95.0	0.0	5.6	5.09	< 0.025
HBN	No Noise	98.3	0.0	100.0	0.0	1.7	∞	< 0.001
	SNR = 6 dB	87.2	1.0	98.3	0.0	12.7	19.22	< 0.001
	SNR = 3 dB	81.1	1.7	95.0	0.0	17.1	14.16	< 0.001
	SNR = 0 dB	78.8	1.9	93.9	1.0	19.2	12.18	< 0.005
	SNR =-3 dB	74.4	1.0	84.4	1.0	13.4	12.25	< 0.005
Avg.	No Noise	98.9	0.8	100.0	0.0	1.1	2.74	< 0.1
	SNR = 6 dB	94.9	5.1	99.6	0.9	5.0	1.82	< 0.2
	SNR = 3 dB	91.7	7.3	98.1	2.2	7.0	1.68	< 0.2
	SNR = 0 dB	88.0	7.7	95.8	1.8	8.9	1.97	< 0.2
	SNR =-3 dB	84.4	7.4	91.9	6.1	8.9	1.75	< 0.2

For all the subjects, score generally decreases as the masking noise level increases. We further see that the scores for processed speech is higher than that for the unprocessed speech under the same condition of masking noise. It is to be noted that the improvements due to processing are more for higher levels of masking noise (i.e. higher level of sensori-neural loss).

For subjects SAK, and HBN the differences in recognition score are highly significant under all masking conditions. For subjects MSC and CKS, the differences are significant at all conditions except at 6 and 0 dB SNR, and 6 dB SNR conditions respectively.

We see that for a given masking level, the unprocessed scores do vary across the subjects. Relative improvements in recognition score (R.S.) were calculated as

$$\left((R.S.)_{\text{processed speech}} - (R.S.)_{\text{unprocessed speech}} \right) / (R.S.)_{\text{unprocessed speech}}$$

and these are also given in Table 1. For 6, 3, 0, and -3 dB SNR the relative improvement in recognition scores ranges from 2.2 to 12.7, 2.9 to 17.1, 2.2 to 19.2, and 5.6 to 13.4 respectively. Averaged across the subjects, the percentage improvements at these SNR levels are

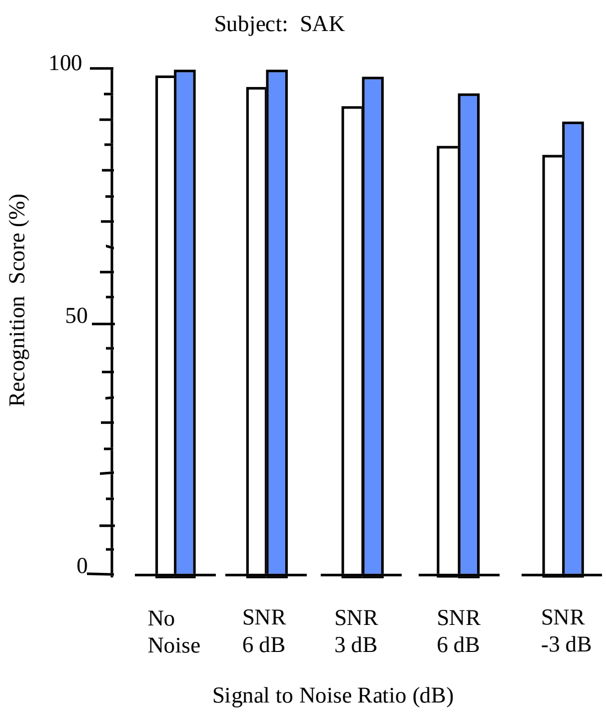

FIGURE 4. Recognition score at different SNRs by combining three confusion matrices.

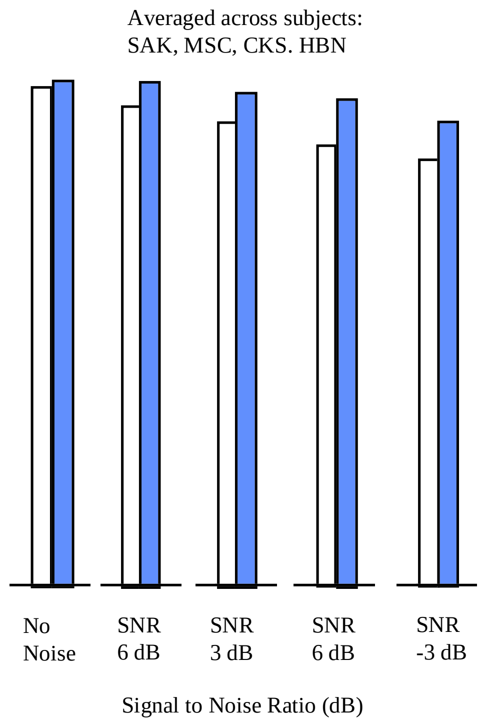

FIGURE 5. Recognition score at different SNRs averaged for four subjects.

1.1,

5.0, 7.0, 8.9, and 8.9 respectively indicating that processing of the speech and dichotic presentation improves recognition scores and improvements are higher under adverse listening condition. A compilation of qualitative assessments, by the subjects, about the set of test stimuli under various listening conditions indicated that the speech quality was better with processing for dichotic presentation.

III. CONCLUSIONS

It is clearly observed from the results for all the subjects that the recognition score generally decreases as the masking noise level increases. Further, we see that under equal condition of masking noise, the score for processed speech is higher than that for the unprocessed speech. The important finding is that the improvements due to processing are more for higher levels of masking noise, i.e. higher levels of simulated sensorineural loss. However, these improvements tend to level, at very high levels of loss.

REFERENCES

Baragi B. N. (1996). *A Speech Training Aid for the Deaf,* M.Tech. dissertation, Dept. of Electrical Engineering, IIT, Bombay.

DeGennaro S., Braida L. D., and Durlach N. I., "Study of multi-band syllabic compression with simulated sensorineural hearing loss," J. Acoust. Soc. Am., vol. 69, S16.
Dorman M. F., and Dougherty K. (1981). "Shifts in phonetic identification with changes in signal presentation level," J. Acoust. Soc. Am., vol. 69, pp. 1439-1440.
Fletcher H. (1952). "The perception of sound by deafened persons," J. Acoust. Soc. Am., vol. 24, pp. 490-497.
Instruction Manual, Precision Sound Level Meter (1985). (Bruel & Kjaer, Denmark).
Jesteadt W. (1997). "Modeling sensorineural hearing loss," Personal communication.
Leek M. R., and Summers V. (1996). "Reduced frequency selectivity and the preservation of spectral contrast in noise," J. Acoust. Soc. Am., vol. 100, pp. 1796-1806.
Lochner J. P. A., and Burger J. F. (1961). "Form of the loudness function in the presence of masking noise," J. Acoust. Soc. Am., vol. 33, pp. 1705-1707.
Moore B. C. J. (1982) *An Introduction to Psychology of Hearing,* New York: Academic.
Simon C. (1979). "On the use of comfortable listening levels in speech experiments," J. Acoust. Soc. Am., vol. 64, pp. 744-751.
Skinner M. W., Holden L. K., and Holden T. A. (1997). "Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants," J. Acoust. Soc. Am., vol. 101, pp. 3766-3782.
Snedecor G. W., and Cochran W. G. (1980). *Statistical Methods,* Ames, Iowa: The Iowa State University Press.
Thomas T. G., Pandey P. C., and Agashe S. D. (1994). "A PC-based multiresolution spectrograph," J. IETE, vol. 40, pp. 105-108.
User's Manual, PCL-DSP25 TMS 320C25 Digital Signal Porcessor Board (1993). (Dynalog MicroSystems, Bombay).
User's Manual, PCL-208 (PCL-718) Data Acquisition card (1989). (Dynalog MicroSystems, Bombay).
User's Guide, HP 33120A Function Generator / Arbitrary Waveform Generator (1994). (Hewlett Packard, USA).
Zwicker E. (1961). "Subdivision of audible frequency range into critical bands (Frequenzgruppen)," Letter to Editor, J. Acoust. Soc. Am., vol. 33, p. 248.