Proc. 4th World multi conference on systemics, cybernetics and informaics (SCI'2000), Orlando, Fl. USA 2000, pp.346-353

INTER-AURAL SWITCHING WITH DIFFERENT FADING FUNCTIONS FOR DICHOTIC PRESENTATION TO REDUCE THE EFFECT OF TEMPORAL MASKING IN SENSORINEURAL HEARING LOSS

D. S. Jangamashetti and P. C. Pandey
Department of Electrical Engineering
Indian Institute of Technology, Bombay, India

ABSTRACT

Sensorineural hearing loss is associated with increased temporal masking of low energy sub-segments of speech by adjacent high energy segments, resulting in increased consonant confusion. Splitting the speech into segments and presenting the alternate segments to the two ears, may help in reducing the effect of increased temporal masking in cases of moderate bilateral loss. A method of inter-aural switching with a switching frequency of 50 Hz, with step and trapezoidal "fading functions" was investigated. It has been established in our earlier work that inter-aural switching with step transition with certain overlap of segments (duty cycle of 70%) helps in improving the perceived quality, by reducing the perception of temporal gaps. Use of trapezoidal fading function during switching may help in reducing the occurrence of high frequency components and thereby further increasing the speech intelligibility. Effect of different slopes (during trapezoidal variation) in improving the speech quality was studied. Stimuli consisted of 12 English non-sense syllables in VCV context. Five normal hearing subjects with simulated hearing loss participated in the tests. Results showed that inter-aural switching with trapezoidal fading function improved the scores further, and the relative improvements were larger at higher noise levels. Thus the inter-aural switching with trapezoidal transition helps in improving the perception by reducing the effect of temporal masking.

Keywords: Sensorineural hearing loss, Temporal masking, Binaural dichotic presentation.

1. INTRODUCTION

Sensorineural hearing loss, which occurs due to damage to hair cells in the cochlea and degeneration of auditory nerve fibers, is characterized by elevated thresholds, loudness recruitment, reduced frequency and temporal resolution, and increased spectral and temporal masking. Masking is a phenomenon in which presence of one signal component elevates the threshold of neighboring signal component. The masking takes place primarily, at the level of peripheral auditory system. Sensorineural loss is associated with broad auditory filters, resulting in poor frequency selectivity and thereby spectral masking. To reduce the effect of spectral masking, for patients with bilateral sensorineural hearing loss, speech processing schemes making use of binaural dichotic presentation have been investigated [1, 2, 3]. Increased temporal masking due to reduced temporal resolution results in forward and backward masking of weak energy components by the adjacent strong ones. Important cues namely voice-onset-time, formant transition, and burst duration that are required for consonant identification get masked by the preceding and following vowel segments thereby resulting in degraded perception. To reduce the effect of temporal masking, speech processing schemes based on the use of properties of clear speech have been investigated [4, 5]. These schemes have been tested using synthesized speech. Real time implementation requires reliable identification of acoustic cues and no implementations have been reported.

Provision of relaxation period between the sub-segments of speech may help in overcoming the problem associated with increased temporal masking. Lunner et al [1] studied the combined effect of spectral and temporal splitting for binaural dichotic presentation. For spectral splitting they used band pass filters of 700 Hz constant bandwidth. Odd and even bands were presented to the two ears. For temporal splitting, they used a symmetrical inter-aural switching with a frequency of 50 Hz to switch the odd and even bands alternately between the two ears. They have reported that temporal splitting did not contribute to improvement in scores. The sound quality deteriorated, and this was ascribed to small gap at inter-aural switching transitions. Provision of a certain overlap during inter-aural switching (symmetrical inter-aural switching with duty cycle > 50%) may help in reducing the perception of temporal gaps due to switching, and may thereby increase the speech intelligibility for persons with sensorineural hearing loss and for normal hearing persons under adverse listening conditions. Objective of our research was to investigate a scheme of temporal splitting of speech with inter-aural switching, for binaural dichotic presentation with alternate segments presented to the two ears. Study has been carried out to investigate the effect of "fading function" for switching. The experimental evaluation was done by conducting listening tests on normal hearing subjects with simulated sensorineural hearing loss.

2. PROCESSING

In our previous work [6], it has been established that fading function with step transition along with an overlap period helps in reducing the effect of increased temporal masking. Overlap of segments of speech during splitting helped in reducing the perception of temporal gaps during inter-aural switching. For temporal splitting, symmetrical inter-aural switching frequency of 50 Hz was used. Figure 1 shows the scheme of temporal processing, the outputs $s_1(n)$ and $s_2(n)$ of temporal processing were obtained by multiplying input $s(n)$ with fading functions $w_1(n)$ and $w_2(n)$ respectively. The two time windows used for switching with step transition are shown in Fig. 2. For an inter-aural switching period of N samples, "on" period of L samples corresponds to duty cycle $d = L/N$. For an inter-aural switching period of N samples with 50% duty cycle, alternate segments of $N/2$ samples are presented to the ears so that adjacent segments get presented to the different ears, i.e. while first $N/2$ samples are presented to the left ear, right ear is relaxed and during next $N/2$ samples of signal presentation to the right ear, left ear is relaxed. Implementation of this method and evaluation through listening tests indicated that duty cycle of 70% was optimum.

While providing relaxation interval and improving the speech perception degraded due to increased temporal masking, use of step transition during switching may result in undesired high frequency components which deteriorate the speech quality. Switching with trapezoidal fading function may help in improving the speech perception by avoiding the occurrence of high frequency components. The two switching time windows are shown in Fig. 3. For 20 ms switching interval and 70% duty cycle, transition durations of 0, 1, 2, and 3 ms were used. Transition duration of 0 ms refers to step transition and 3 ms refers to maximum transition duration. A program was written using C for offline processing. The program provides facility for selecting desired inter-aural switching period, "on" period, and transition duration. The processed signals were added with noise at different SNRs to simulate the sensorineural hearing loss.

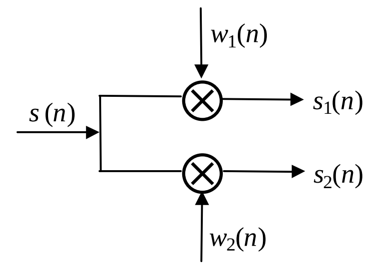

FIG. 1. Temporal splitting of the signal.

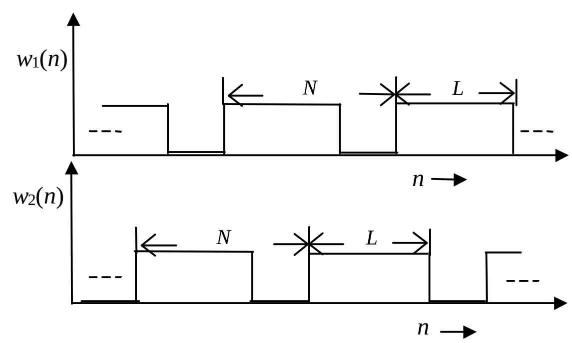

FIG. 2. Overlapping inter-aural switching with step fading

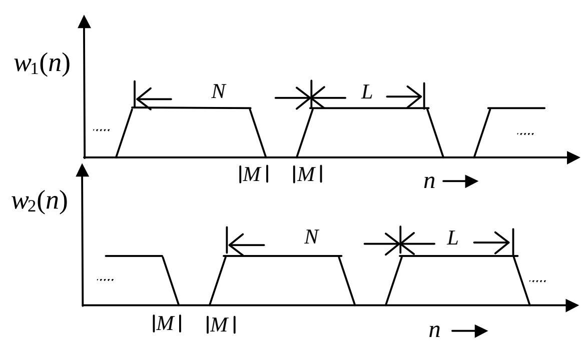

FIG. 3. Overlapping inter-aural switching with trapezoidal fading

3. METHODS AND MATERIALS

Five (3 male and 2 female) normal hearing subjects with simulated sensorineural hearing loss participated in the test. Their ages ranged between 20-35 years. All the subjects had pure tone hearing thresholds less than 20 dB HL in the frequency range of 125-6000 Hz. Test materials consisted of twelve English nonsense syllables / *apa, aba, ata, ada, aka, aga, ama, ana, asa, aza, afa, ava* /. The test stimuli were digitally recorded in an acoustically isolated room, using a microphone, amplifier, and lowpass filter with cutoff frequency of 4.8 kHz, at a sampling rate of 10 k Sa/s with 16-bit resolution. Simulation of hearing loss was done by adding broad band noise having Gaussian distribution to each of the test stimulus [7]. Signal-to-noise ratios (SNR) of ∞, 6, 3, 0, -3, and -6 dB were used to simulate hearing loss of varying degrees.

The scheme was experimentally evaluated by conducting listening tests using a PC based automated test set-up. The unprocessed and processed signals were outputted from the two D/A ports of a data acquisition card. The signals after passing through low pass filters with cut off frequency of 4.8 kHz, were fed to a pair of calibrated headphones (Telephonics TDH-39). Test materials were presented at most comfortable listening level which was between 70-75 dB SPL [8, 9, 10]. Experiments were carried out in an acoustically isolated room. There were 30 test conditions (5 processing conditions x 6 SNR). Each test condition was run a number of times in random order. In each test, there were a total of 60 presentations, i.e. 5 presentations for each stimulus. For each condition, results of five tests with stabilized scores were considered for analysis. Subject's responses were stored in the form of stimulus-response confusion matrix and response time statistics. The confusion matrix under a test condition was obtained with 25 presentation of each stimulus. Confusion matrices were used for finding recognition score and relative information transmission [11]. Information transmission analysis was also carried out for the feature grouping of phonemes for the features of voicing, place, manner, nasality, frication, and duration. Paired t-test was used for finding the significance of difference between the scores of unprocessed and processed signals [12].

4. RESULTS AND DISCUSSION

The test results were analyzed by comparing recognition scores of unprocessed and processed speech. Table 1 shows the percentage recognition scores of five subjects, for unprocessed signal and relative improvement for processed speech for different transition durations. For unprocessed speech, recognition scores for all the subjects decrease as the SNR degrades. For subject S1, there is a relatively small variation, while for subject S5, there is a very large variation. The averaged score decreases from 100% for no noise to 80% for -6 dB SNR. Processing with dichotic presentation has resulted in improvement in the recognition scores, and the extent of improvement appears to be related to the subject' s susceptibility to poor SNR. The improvements are there for all the transition durations, these are generally highest for 2 ms transition. Average relative improvement (%) in recognition score for 6, 3, 0, -3, -6 dB SNR conditions were 1.9, 3.4, 2.4, 5.5, 11.0 for step variation and 5.6, 5.6, 8.9, 12.8, 15.6 for trapezoidal variation with transition duration of 2 ms respectively. Improvements in recognition scores were statistically significant for higher levels of noise (-3 and -6 dB SNR).

Table 1. Recognition scores (%) for unprocessed speech (Su) and relative improvement (%) for processed speech (Sp), with 4 transition durations : A–0 ms, B–1 ms, C–2 ms, and D–3 ms. S: Subject, Avg: averaged across subjects. p : significance level for paired (unprocessed vs processed) t-test across subjects.

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR				
	Su	Sp				Su	Sp				Su	Sp				Su	Sp				Su	Sp				Su	Sp			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
S1	100	-0.6	-0.3	0	0	99	0.4	0.4	0.7	0.7	97	2.7	2.7	2.7	2.7	97	2.0	3.1	3.1	2.8	96	2.1	4.2	4.2	4.2	94	3.5	5.3	5.0	4.2
S2	100	-0.6	-0.3	0	0	98	1.3	0.7	2.0	1.7	97	1.3	1.7	1.3	0.7	96	0.7	3.8	2.1	3.8	95	1.5	4.0	4.2	3.5	93	5.4	5.7	6.5	7.5
S3	100	0	-0.3	0	0	92	2.2	4.3	3.3	5.4	91	1.0	3.6	3.3	1.8	80	4.0	8.2	3.3	5.7	71	3.3	10	20.2	13.1	78	8.5	10.6	7.3	12
S4	100	-0.3	0	0	0	93	2.1	4.3	6.4	4.3	93	4.6	4.3	4.3	3.2	91	2.5	0	2.6	4.0	84	5.1	11	14	13.0	79	11.4	19	20	20.6
S5	100	0	0	0	0	75	3.6	10.3	15.6	11	74	7.5	6.7	16.5	19	69	2.6	28	33.5	34	67	15.6	18	21.5	19.6	59	27.6	37	39	36.5
Avg.	100	-0.3	-0.2	0	0	91	1.9	4.0	5.6	4.6	90	3.4	3.8	5.6	5.5	86	2.4	8.6	8.9	10	82	5.5	9.4	13	10.7	80	11	15.5	15.6	16
P		0.07	0.07	-	-		.015	.07	.08	.04		.04	.004	.08	.15		.009	.12	.18	.12		.09	.015	.018	.018		.028	.029	.041	.024

Table 2. Overall information transmitted (%) for unprocessed (Su) and for processed (Sp) speech signal.

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR				
	Su	Sp				Su	Sp				Su	Sp				Su	Sp				Su	Sp				Su	Sp			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
S1	100	99	99	100	100	99	99	99	100	100	96	100	100	100	100	96	99	100	100	99	96	97	100	100	100	93	96	98	98	97
S2	100	99	99	100	100	97	99	98	100	99	98	100	99	98	98	96	96	98	98	100	94	98	98	98	97	93	97	97	98	100
S3	100	99	99	100	100	93	94	96	95	96	93	93	94	95	95	85	86	89	87	88	75	77	78	86	81	82	85	88	85	90
S4	100	100	100	100	100	83	84	89	90	90	82	86	87	90	91	81	82	89	91	91	83	85	82	84	84	80	81	83	83	83
S5	100	99	100	100	100	93	94	97	98	97	93	98	96	97	96	92	98	96	97	96	87	90	93	96	95	83	92	93	95	95
Avg.	100	99.2	99.4	100	100	93	94	95.8	96.6	96.4	92.4	95.4	95.2	96	96	90.6	92	94.4	94.6	94.8	87	89.4	90.2	92.8	91.4	86	90	92	92	93

Table 3. Grouping of consonants by features.

Feature	Groups
voicing (2)	unvoiced : / *p t k s f* /, voiced : / *b d g m n z v* /
place (3)	front : / *p b m f v* /, middle : / *t d n s z* /, back : / *k g* /
manner (3)	oral stop : / *p b t d k g* /, fricative : / *s z f v* /, nasal : / *m n* /
nasality (2)	oral : / *p b t d k g s z f v* /, nasal : / *m n* /
frication (2)	stop : / *p b t d k g m n* /, fricative : / *s z f v* /
duration (2)	short : / *p b t d k g m n f v* /, long : / *s z* /

Table 2 gives the overall information transmitted for all five subjects under all SNR conditions. It can be seen that there is a degradation in overall information transmission with increase in noise level for unprocessed speech. Average percentage relative improvement with processed speech for 6, 3, 0, -3, -6 dB SNR conditions were respectively 1.1, 3.3, 2.4, 2.8, 4.7 and 4.0, 4.1, 5.2, 7.0, 6.5 for step and trapezoidal variation (transition duration of 2 ms). An interesting observation here is that for unprocessed speech, subject S5 has a very low recognition score with low SNR. However, the relative information transmitted for this subject is not much lower than for other subjects. This indicates that errors in reception by this subject are not randomly distributed. It is also seen that dichotic presentation brings the relative information transmitted for this subject to almost the same level as for other subjects.

For a detailed look into the contributions by various consonantal features in the perception, information transmission analysis was carried out for grouping of phonemes by features of voicing, place, manner, nasality, frication, and duration as given in Table 3. It was seen that processing improved the relative information transmission for all the features, particularly at lower SNRs. However, the improvements vary over a large range for different features, being maximum for place and duration.

Information transmitted for the features of voicing, place, and duration are shown in Table 4. For unprocessed speech, it was observed that with decrease in SNR, reception of voicing was not affected much. However, reception of place and duration was severely degraded. Averaged across the subjects, the relative information changes from 100% to 54% for place and 100% to 49% for duration. Subjects S4 and S5 had great difficulty in the reception of these two features under poor listening conditions. With processing, the reception of place and duration improved for all the subjects, and particularly for subjects S4 and S5. The improvements for trapezoidal fading are higher than those for step fading. Averaged across the subjects, relative improvement (%) in information transmitted for place feature were 3.3, 9.7, 10.7, 20.2, 54.3 and 13.5, 15.4, 29, 39.6, 66.5 for step and trapezoidal variation (transition duration of 2 ms) respectively for 6, 3, 0, -3, -6 dB SNR conditions. Averaged across the subjects, the relative improvement (%) in information transmitted for duration feature for 6, 3, 0, -3, -6 dB SNR conditions were -0.6, 0.4, 17.4, 64, 49 and 9.7, 4.74, 40.1, 70.4, 244 for step and trapezoidal variations (transition duration of 2 ms) respectively. It is further seen that transition duration does not have much effect on the reception of duration feature, while longer transition duration generally results in higher improvement in the reception of place feature. It is to be noted that trapezoidal fading results in lesser amount of spectral distortion, and therefore it might be more helpful in place perception.

Table 4. Relative information transmitted (%) for unprocessed (Su) and for processed (Sp) speech signal for feature groupings of : (a) voicing, (b) place, (c) duration.

(a) Feature: voicing

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR				
	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D
S1	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
S2	100	100	100	100	100	100	100	97	100	100	100	100	100	100	100	97	100	100	100	100	100	100	100	100	100	100	100	100	100	100
S3	100	100	100	100	100	100	97	100	100	100	100	100	100	100	100	90	100	100	100	97	88	94	95	95	95	97	93	100	93	100
S4	100	100	100	100	100	97	100	100	100	100	100	97	97	100	100	97	97	100	100	100	100	100	100	100	100	97	100	97	100	100
S5	100	100	100	100	100	97	97	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	97	100	100	100	100
Avg.	100	100	100	100	100	98.8	98.8	99.4	100	100	100	99.4	99.4	100	100	96.8	99.4	100	100	99.4	97.6	98.8	99	99	99	98	98.6	99.4	98.6	100

(b) Feature: place

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR				
	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D
S1	100	97	98	100	100	97	98	98	100	100	90	100	100	100	100	92	97	100	100	98	88	91	100	100	100	76	90	95	93	91
S2	100	98	98	100	100	92	97	97	100	98	97	100	98	95	97	88	93	97	94	100	85	95	93	95	91	76	90	93	95	100
S3	100	98	98	100	100	81	81	85	82	88	76	75	83	81	81	59	61	70	60	68	37	45	51	65	57	50	60	60	57	65
S4	100	100	100	100	100	53	57	70	72	74	53	64	63	74	74	37	41	73	80	84	34	52	54	55	58	19	51	57	56	57
S5	100	97	100	100	100	79	81	88	94	89	80	90	88	89	87	78	100	88	89	87	62	69	79	85	84	47	69	79	83	82
Avg.	100	98	98.8	100	100	80	82.8	87.6	89.6	89.8	79	85.8	86.4	87.8	87.8	71	78.4	85.6	84.6	87.4	61	70.4	75.4	80	78	54	72	76.8	76.8	79

(c) Feature: duration

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR				
	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D	Su	Sp A	Sp B	Sp C	Sp D
S1	100	94	100	100	100	100	100	100	100	100	96	96	100	100	100	89	95	100	100	100	88	100	100	100	100	84	100	95	95	89
S2	100	95	95	100	100	100	100	100	100	100	100	100	100	100	100	100	95	100	100	100	91	100	96	100	94	77	91	92	94	100
S3	100	95	100	100	100	73	72	73	73	85	85	74	83	85	88	52	73	76	54	71	41	56	55	64	50	38	57	78	61	74
S4	100	100	100	100	100	63	55	60	64	65	60	59	59	65	65	32	34	67	80	94	20	64	75	62	80	07	70	80	77	72
S5	100	94	100	100	100	64	71	94	94	91	80	90	88	89	87	72	100	96	96	96	59	83	83	96	81	42	96	89	94	92
Avg.	100	95.6	99	100	100	80	79.6	85.4	86.2	88.2	84	83.8	86	87.8	88	69	79.4	87.8	86	92	60	80.6	96.8	84.4	81	49	82.8	86.8	84.2	85.4

5. CONCLUSIONS

To reduce the effect of temporal masking, scheme of temporal splitting using symmetrical inter-aural switching at the rate of 50 Hz with 70% overlap and trapezoidal fading for switching was implemented and evaluated. There was maximum improvement in information transmission for duration feature. Hence this scheme helps in reducing the effect of temporal masking. Slower transitions resulted in better reception of place feature. Studies should be carried out for establishing the optimal combination of inter-aural switching period, overlap duration, and transition slope. Also processing schemes with appropriate combination of temporal and spectral splitting need to be investigated.

6. REFERENCES

1. T. Lunner, S. Arlinger, and J. Hellgren, "8-channel digital filter bank for hearing aid use: preliminary results in monaural, diotic and dichotic modes," *Scand. Audiol*. Suppl. Vol. 38, pp. 75-81, 1993.
2. D. S. Chaudhari and P. C. Pandey, "Dichotic presentation of signal using critical filter bank for bilateral sensorineural hearing impairment," *Proc. of 16th Int. Congress on Acoust. (ICA)*, Seattle, Wash., U.S.A., 1998, pp. 213-214.
3. D. S. Chaudhari and P. C. Pandey, "Dichotic presentation of speech signal with critical band filtering for improving speech perception," *Proc. ICASSP'98*, Seattle, Wash., U.S.A., 1998, Paper AE 3.1.
4. M. A. Picheny, N. I. Durlach, and L. D. Braida, "Speaking clearly for the hard of hearing II: Acoustic characteristic of clear and conversational speech," *J. Speech Hear. Res*., vol. 29, pp. 434–446, 1986.
5. T. G. Thomas, P.C Pandey, and S. D. Agashe, "Effect of consonantal intensity and duration modification on speech perception by listeners with simulated hearing impairment," *J. Acoust. Soc. India*, XXIV, VI-4.1-4.5, 1995.
6. D. S. Jangamashetti and P. C. Pandey, "Dichotic presentation with inter-aural switching for reducing the effect of temporal masking due to sensorineural hearing loss," *Proc. National Conference on Biomedical Engg.*, Roorkee, India., 2000, pp. 346-353.
7. L. E. Humes, D. D. Dirks, T. S. Bell, and G. E. Kincaid, "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," *J. Acoust. Soc. Amer*., vol. 81, no. 3, pp. 765–773, 1987.
8. S. Gordon-Salant, "Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects," *J. Acoust. Soc. Amer*., vol. 81, pp. 1199–1202, 1987.
9. C. Simon, "On the use of comfortable listening levels in speech experiments," *J. Acoust. Soc. Amer*., vol. 64, no. 3, pp. 744–750, 1978.
10. M. F. Dorman and K. Dougherty, "Shifts in phonetic identification with changes in signal presentation level," *J. Acoust. Soc. Amer*, vol. 69, no. 5, pp. 1439–1440, 1981.
11. G. A. Miller and P. E. Nicely, "Analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Amer.*, vol. 27, no. 2, pp. 338-352, 1955.
12. G. W. Snedecor and W. G. Cochran, *Statistical Methods*, Ames, Iowa: The Iowa State University Press, 1980.