International Journal of Electronics Engineering, 1(2), 2009, pp. 179-184

Effect of Consonant Duration Modifications on Speech Perception in Noise

N. H. Shobha[1], T. G. Thomas[2] & K. Subbarao[3]

[1]Dept. of ECE, Osmania University, Hyderabad, India-500007
[2]Dept. of Electrical Engineering. BITS-Pilani, Dubai, UAE
[3]Dept. of ECE, Osmania University, Hyderabad, India-500007

Abstract: *Our investigations addressed the efficacy of temporal modifications to the clear speech advantage. A case for synthetic clear speech in the context of hearing impairment was developed. Plosive–vowels of English language were used as test stimuli. Stimuli were subjected to consonant-duration expansion, wherein burst duration, voice onset time, and formant transition segments were time-expanded independently by 50-100% of their original duration. The speech perception in noise (SPIN) tests were quantified in terms of information transmission analysis measures, in the presence of white noise-masker at three noise levels, 0 dB, +12 dB, and +6 dB. Lengthening burst duration by 50% and formant transition by 100% was found to improve speech intelligibility in simulated low level sensorineural hearing loss.*

Keywords: *Consonant-duration Modification, Information Transmission Analysis, White-noise Masker, Speech Intelligibility.*

INTRODUCTION

The speech that we hear is often degraded by the addition of competing speech and non-speech signals. People suffering from hearing loss often have the greatest difficulty understanding speech in noisy environments. When talkers slow their speech, intelligibility is improved for those hearing- impaired listeners. The speech produced when a talker intentionally tries to improve intelligibility by speaking slowly and clearly, but without exaggeration, is called 'clear speech'. Even though slowing is a consistent feature of clear speech, specific acoustic changes occur in addition to insertion of pauses and lengthening of durations of individual speech sounds. Previous studies on speech perception have demonstrated a significant intelligibility advantage for clear speech over conversational speech in both normal-hearing and hearing-impaired listeners across a wide range of listening conditions including quiet, noisy, and reverberant backgrounds [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11], additional insight comes from the work on the characteristics of a 'clear' speaking style [12, and 13]. The clear speech context included an increase in consonant duration (CD), an enhanced consonant-vowel intensity ratio (CVR), and some of the systematic changes in encoding phonetic contrasts like slower speaking rates, longer formant transitions, less vowel reduction [5,12,13, and 14]. Accordingly, a clear speech context, bursts are reported to be intense, while stop gaps and formant transitions are longer corresponding to plosive consonants.

The present study focuses on lengthening consonant duration of plosives to provide slow speech rate or in other words to provide extra processing time for the hearing impaired subjects. Nonsense syllables involving plosive consonants / p t k b d g / in CV context with vowels /a, ɛ, o/ were used as test stimuli. Plosives are those consonants produced by first forming a complete closure in the vocal tract via a constriction at the place of constriction, during which there is a either silence or a low-frequency hum called 'voicebar'. The vocal tract is then opened suddenly releasing the pressure built up behind the constriction; this is characterized acoustically by a transient and /or a short duration noise 'burst' [15]. The period between the release of the stop and the beginning of voicing in the vowel is called the' voice onset time' (VOT). During this period there is a silence and/or aspiration noise. The time interval between the onset of the following vowel and the instance when a formant frequency reaches its steady-state value is called the 'formant transition' (FT).

In the current investigation, natural syllables were recorded, subjected to resynthesis, processed for consonant duration lengthening in three independent schemes, (i) Burst Duration Modification (BDM), (ii) Voice Onset Time Modification (VOTM), and (iii) Formant Transition Duration Modification (FTDM). The consonant lengthening protocol employed was PSOLA (pitch-synchronous overlap and add) [16]. In PSOLA, the original pitch is being preserved during the processing, [17] and hence degradation of transitional portions in plosive consonants may be avoided [17].

SPEECH IN NOISE TASK

The various stages involved in the speech perception in noise tests are discussed below.

*Corresponding Author: shobhanh7@gmail.com

1.1. Speech Material

Nonsense syllables with consonant-vowel /CV/ structure were chosen for investigation. The idea behind this non sense syllable test (NST) was to maximize the contribution of acoustic factors, and minimize the impact of adjacent vowels. The test material consisted of plosive consonants, / p t k b d g / in the context of cardinal vowels /a, ɛ, o/ forming voiceless /CV/ subset, /pa, pɛ, po, ta, tɛ, to, ka, kɛ, ko/, voiced /CV/ subset , /ba, bɛ, bo, da, dɛ, do, ga, gɛ, go/. The vowels were chosen to have formant frequencies close to each other, with the goal of making them more confusable.

1.2. Speech Signal Processing

The signal processing was accomplished in four different stages as explained below. In the first stage, we recorded the natural speech tokens and subjected them to resynthesis. The natural stimuli were recorded in a quiet room, sampled at 44.1KHz, using a PRAAT monosound recorder. The best utterance out of 20 utterances of the first author (middle aged, female) was selected based on the proper phonetic clarity.

The speech tokens were subjected to resynthesis using the procedure of LPC (linear prediction) analysis-synthesis as provided in PRAAT [18]. The idea behind the resynthesis was two-fold; firstly, the synthetic copy renders efficient and independent manipulation of the spectral, temporal and intensity characteristics; secondly, synthetic speech is as similar as possible to a human utterance. The resynthesis tradition assumes five formants in the range between 0 to 5500 Hz for a female voice, 0 to 5000 Hz for a male voice and 0 to 10000 Hz for child voice. For implementing linear prediction with Praat, we have to implement this band-limiting by resampling the original signal to 11 KHz for female, 10 KHz for male, or 20 KHz for a young child. In the current investigation, we performed resampling at 11 KHz. We extracted the filter and the source from the resampled sound using linear-prediction analysis. The analysis procedure adopted 10 linear-prediction coefficients (yields at most 5 formant-bandwidth pairs) in each time frame of 5 or 10 ms, which is suited for capturing changes in the speech signal. Next, using the extracted source and filter, the speech sound was regenerated based on LPC synthesis. This procedure gave back the resynthesized version with the original quality except that the windowing caused few ms at the beginning and the end of the signal to be set to zero Finally, these tokens were normalized to 70 dB IL (referred as baseline syllables) to avoid the signal clipping in subsequent processing stages.

In the second stage of processing, consonant segment durations such as burst release/burst duration, VOT, and FTD measurements were measured by visual inspection of the time waveforms and wideband spectrograms using the PRAAT software. The release burst was identified as the short segment characterized as a 'spike' in the time domain and a sudden, sharp vertical line in the spectrogram [19]. The segmentation of a burst was performed visually by examining both the waveform and the spectrogram. VOT was identified as duration from the end of burst to the beginning of the vowel (the beginning of first waveform period) [19]. It is to be noted that the silence or closure interval of plosives cannot be defined for isolated CV syllables [20]. The current stage of investigation reported that, the release burst was longer for voiceless than voiced plosives; VOT durations were longer for velars than alveolar, which in turn were longer than for labials; formant transitions were longer for voiced than voiceless plosives.

The Formant transition durations were measured by simultaneous consultation of time domain waveform, spectrogram, linear-predictive coding (LPC) spectra, and short-time fast-fourier transform (ST-FFT) spectra [19] The LPC spectrum was constituted for a prediction order of 10 (at least twice as the number of spectral peaks that we want to detect), analysis window of 12.5 ms and 5 ms step, +6dB/octave filtering above 50 Hz. The three formants were originally located by examining the LPC spectra, FFT spectra, and spectrogram. The steady-state point of the vowel was centered at 100 ms after the onset. Formant analysis was performed for the detection of formant transition duration. After proper settings, formant contour was extracted and the formant values were written to a text file. Utilizing this data, the duration of the transitions and their onset and offset points were determined, and we then applied a time warp to all formants over the determined duration of the transition. The acoustic segmentations and measurements were done using PRAAT software.

In the third stage of processing, the extracted acoustic segments were subjected to duration modification or time-stretching. This stage of processing employed a time-stretching algorithm referred as Pitch-Synchronous Overlap and Add (PSOLA). Based on the modification strategies, consonant duration modifications took three different schemes; (i) Burst duration modification–BDM [20, and 21] (ii) Voice Onset time modification–VOTM [14, 19, and 20], and (iii) Formant Transition Duration modification–FTDM [14, 20, 22, and 23]. The PSOLA analysis-modification-synthesis method belongs to the general class of STFT (short-time Fourier Transform) analysis-synthesis method. The analysis phase performs the segmentation of the input speech, and the synthesis phase generates a time stretched version by overlapping and adding time segments extracted by the analysis phase. In the PRAAT object window, PSOLA can be found as sound > Convert > Lengthen (PSOLA). Here, the term 'factor' decides the factor for lengthening or shortening; by choosing factor value > 1 or < 1, the resulting sound could be longer or shorter than the original segment, but a factor value larger than 3 will not work. We selected a minimum pitch of 75Hz and a maximum pitch of 600Hz, while a 'factor' of 1.5 for 50% lengthening (compared to original duration) and a 'factor' of 2 for 100% lengthening (compared to original duration). Finally, the lengthened segment was blended back to its original location to result in time stretched version. We thus obtained three modifications for each stimulus, one without modification (0%) and other two with modifications (50% and 100%) under all three schemes (BDM/VOTM/FTDM).

The fourth stage of processing was designed to simulate hearing impairment, by reducing the acoustic dynamic range. The masking noise responsible for the threshold elevation is believed to be predominantly of cochlear origin [24]. As reported in literature, the reduction in the hearing threshold can be approximately simulated by addition of white noise [25, 26]. Some researchers have employed multi-talker babble instead of white noise [26, 27, and 28]. However, due to its non-stationary nature, the effective masking it may provide during stimulus presentation is unpredictable. Hence, we decided to use white noise masker to model the hearing loss to a good approximation.

The processed tokens from the previous stage were additively mixed with the synthesized noise at three noise conditions, i.e., no-masking noise, +12 dB and +6dB SNRs. The noise free (natural) tokens were considered as no-masking noise tokens. The SNR refers to the ratio of the average power in CV token to the average power of the noise token in decibels. For deriving +12dB and +6dB SNR-tokens, the average power level of the speech token was fixed while that of the noise was adjusted. PRAAT scripts were run for synthesizing the white noise and for the process of mixing [29]. The Chris-Darwin [29] algorithm which performed additive mixing summed up the sounds by point-to-point values, preserving real time across the time domains. Finally, after four stages of processing stimuli corpus holds 486 test tokens spanning across 18 syllables, 3 duration lengthening schemes, 3 versions of lengthening per scheme and 3 SNRs per version.

1.3. Speech Perception in Noise

Two female and two males in the age group of 16-45 years with normal hearing, participated in the listening experiments. None of the subjects were experienced with perceptual experiments; subjects went through a stimuli familiarization sessions before the experiment started.

The perception tests were automated using a MATLAB code with graphic user interface. Stimuli were presented using a computerized testing procedure at the most comfortable listening level of 75 to 85 dB SPL for the listeners. The test procedure used a similar protocol for all three experiments. The Experiment under each individual scheme (BDM/ VOTM/ FTDM*)* worked on a total of 162 tokens categorized under 9 or (3*3) listening conditions. These included the original and processed stimuli with 3 levels of CD lengthening (0%, 50% and 100%) and 3 noise levels (no-noise, +12dB, +6dB). Under every listening condition, subjects were played tokens with ten randomized replications of each token; they were prompted to choose from the set of choices displayed on the computer screen. Results were cast into three groups of six by six confusion matrices (CM) per run.

1.4. Speech Intelligibility Measures

Speech discrimination test results were summarized as the percentage of correct responses for many experimental runs. The sum of the diagonal elements gives the empirical probability of correct responses, known as Recognition Score-RS (or articulation score). The computation of RS is simple, but it obscures the detailed and important information on the distribution of errors among the off-diagonal cells [31]; also it is sensitive to the subject's bias or chance scoring (an artificially high score). We adopted the Information Transmission analysis approach [22, 31, and 32], which provides a measure of covariance between stimuli and responses, and takes into account the pattern of errors and the score in a probabilistic manner. The covariance measure of intelligibility can be applied to the sub matrices derived from the original matrix by grouping the stimuli in accordance with certain desired features [31, and 33]. The information measures of the input stimulus X and output response Y are defined in terms of the Mean Logarithmic Probability - MLP, given by,

$$I(X;Y) = -\sum_i \sum_j p(x_i, y_j) \log 2\left(\frac{p(x_i)\,p(y_i)}{p(x_i, y_j)}\right) bits \quad (1)$$

The Relative Information Transmission (RIT) from X to Y is given by ,

$$I_{tr}(X;Y) = \frac{I(X;Y)}{I_s(X)} \quad (2$$

Where, *Is*(*x*) is the information measure of the input-stimulus in terms of MLP.

2. EFFECTS OF CONSONANT DURATION MODIFICATIONS ON SPEECH INTELLIGIBILITY

The confusion matrices obtained were analyzed and quantified with perceptual (information transmission analysis) and statistical (two-tailed t-test) measures. The perceptual scores were obtained by averaging the scores for individual subject across three vowel contexts /a, i , u/; the last rows in the table indicates their means and standard deviations. The statistical tables reported the mean percent-correct recognition data, standard deviations (SD), probability value (p) and the corresponding statistical significance value corresponding to the perception test. The processing factor examined the intelligibility benefit between the unprocessed speech and the processed speech, a benefit was treated significant at 0.05 levels; $p <= 0.01$ was accepted as indicative of high significance and $0.01 < p < 0.05$ as moderate significance.

2.1. BDM Paradigm

Tables 1(a) and 1(b) represents the perceptual analysis and statistical analysis scores respectively. For no-masking noise presentations, voiceless and voiced plosives recorded a moderate benefit up to +7% corresponding to 50% and 100% BDM. For +12dB SNR, voiceless and voiced plosives recorded a minimal benefit up to +5% corresponding to 50% and 100% BDM. For +6dB SNR, voiceless and voiced plosives recorded a marginal benefit upto +9% corresponding to 50% and 100% BDM.

The statistical test (Table 1(b)) presents the significance status of the perceptual measures (Table 1(a)). For no-noise masking, voiceless plosives reported highly significant benefit ($p < 0.01$) corresponding to 100% BDM. For +12 dB SNR, voiceless plosives reported moderately significant benefit ($0.05 < p < 0.01$) corresponding to 50% BDM; while voiced plosives reported highly significant benefit corresponding to both 50% and 100% BDMs.

Table 1(a): BDM Scheme-Perceptual Analysis Results

TEST STIMULI	Listener	Relative Information Transmitted (%)								
		No-noise			SNR=12 dB			SNR=6 dB		
		BDM (%)			BDM (%)			BDM (%)		
		0	50	100	0	50	100	0	50	100
Voiceless stop-vowels	L1	89	100	98	86	95	83	70	84	84
	L2	100	100	100	85	86	84	75	84	73
	L3	91	100	100	83	86	88	89	88	90
	L4	89	81	100	85	91	83	70	83	84
	MEAN	92	95	99	84	89	85	76	85	83
	SD	5	10	1	1	4	2	9	2	7
Voiced stop-vowels	L1	81	97	94	93	97	96	86	87	96
	L2	100	100	100	93	97	97	85	85	90
	L3	97	97	97	89	100	97	99	93	95
	L4	81	88	97	93	93	100	86	100	88
	MEAN	90	95	97	92	97	97	89	91	92
	SD	10	5	2	2	3	2	7	7	4

Table 1(b): BDM Scheme–Statistical Analysis Results

Test Stimuli	SNR (dB)	BDM (%)	MEAN	SD	Two Tailed t Test of Difference		
					t	p	Result
Voiceless Stop-vowels	No-noise	0	92	5			
		50	95	10	0.537	0.6108	NS
		100	99	1	2.748	0.0355	S
	12	0	84	1			
		50	89	4	2.429	0.0919	S
		100	85	2	0.894	0.4055	NS
	6	0	76	9			
		50	85	2	1.852	0.0987	NS
		100	83	7	1.228	0.2655	NS
Voiced Stop-vowels	No-noise	0	90	10			
		50	95	5	0.894	0.4055	NS
		100	97	2	1.373	0.2189	NS
	12	0	92	2			
		50	97	3	2.774	0.0323	S
		100	97	2	3.536	0.0123	S
	6	0	89	7			
		50	91	7	0.404	0.7002	NS
		100	92	4	0.744	0.4849	NS

2.2. VOTM Paradigm

Tables 2(a) and 2(b), represents the perceptual analysis and statistical analysis scores respectively. For no-noise presentations, voiceless and voiced plosives reported a minimal benefit up to +4% for a few conditions, but a majority of the conditions have reported negative benefit. In the presence of noise, the effect was reported to be detrimental because of the reported negative benefits.

The statistical analysis (Table 2(b)) presents the significance status of the perceptual measures (Table 2(a)). The analysis has reported no significant benefit under for all three SNR presentations, and three VOT modifications.

Table 2(a): VOTM Scheme–Perceptual Analysis Results

TEST STIMULI	Listener	Relative Information Transmitted (%)								
		No-noise			SNR=12 dB			SNR=6 dB		
		VOTM (%)			VOTM (%)			VOTM (%)		
		0	50	100	0	50	100	0	50	100
Voiceless Stop-vowels	L1	89	89	62	86	86	78	70	76	62
	L2	100	91	83	85	74	72	75	78	82
	L3	91	97	97	83	95	90	89	82	76
	L4	89	87	89	85	82	76	70	76	84
	MEAN	92	91	83	84	84	79	76	78	76
	SD	5	4	15	1	9	8	9	3	10
Voiced Stop-vowels	L1	81	91	84	93	91	97	86	84	83
	L2	100	100	100	93	91	88	85	78	82
	L3	97	89	100	89	100	97	99	95	100
	L4	81	97	93	93	95	97	86	97	84
	MEAN	90	94	94	92	94	95	89	88	87
	SD	10	5	8	2	4	4	7	9	8

Table 2(b): VOTM Scheme–Statistical Analysis Results

Test Stimuli	SNR (dB)	VOTM (%)	MEAN	SD	Two Tailed t Test of Difference		
					t	p	Result
Voiceless Stop-vowels	No-noise	0	92	5			
		50	91	4	-0.312	0.7653	NS
		100	83	15	-1.138	0.2984	NS
	12	0	84	1			
		50	84	9	0	1	NS
		100	79	8	-1.24	0.2611	NS
	6	0	76	9			
		50	78	3	0.422	0.688	NS
		100	76	10	0	1	NS
Voiced Stop-vowels	No-noise	0	90	10			
		50	94	5	0.716	0.5012	NS
		100	94	8	0.625	0.5552	NS
	12	0	92	2			
		50	94	4	0.894	0.4055	NS
		100	95	4	1.342	0.2283	NS
	6	0	89	7			
		50	88	9	-0.175	0.8665	NS
		100	87	8	-0.376	0.7196	NS

2.3. FTDM Paradigm

Tables 3(a) and 3(b), represent the perceptual analysis and statistical analysis score pattern for FTDM scheme. In the no-noise case, voiceless plosives reported moderate benefit upto 9% ;while in the presence of +12 dB, the benefit was up to 11% for voiceless plosives, but +6 dB SNR presentations have reported negative benefit.

The statistical test (Table 3(b)) presents the significance status of the perceptual measures (Table 3(a)). For no-noise masking, voiceless plosives reported highly significant benefit ($p < 0.01$) corresponding to 50% and 100% FTDM. For +12 dB SNR, voiceless plosives reported highly significant benefit corresponding to 50% and 100% FTDM; while voiced plosives reported highly significant benefit corresponding to 100% FTDM; but +6 dB SNR presentations did not report any significant benefit either for 50% or 100% FTDM's.

Table 3(a): FTDMScheme–Perceptual Analysis Results

TEST STIMILI	Listener	Consonant Recognition Scores (%)								
		No-noise			SNR=12 dB			SNR=6 dB		
		FTDM (%)			FTDM (%)			FTDM (%)		
		0	50	100	0	50	100	0	50	100
Voiceless Stop-vowels	L1	89	86	98	89	100	93	80	78	88
	L2	84	99	99	84	94	98	87	90	89
	L3	82	98	91	82	99	99	94	74	74
	L4	89	91	88	89	90	97	80	81	79
	MEAN	86	93	94	86	96	97	85	81	83
	SD	3	6	5	3	5	2	7	7	7
Voiced Stop-vowels	L1	89	89	92	94	92	89	84	86	89
	L2	100	100	99	98	100	91	83	98	92
	L3	98	100	93	94	88	80	97	80	81
	L4	89	99	90	94	91	89	84	79	89
	MEAN	94	97	94	95	93	87	87	86	88
	SD	6	5	4	2	5	5	6	9	5

Table 3(b): FTDMScheme–Statistical Analysis Results

Test Stimuli	SNR (dB)	FTDM (%)	Mean	SD	Two Tailed t Test of Difference		
					t	p	Result
Voiceless Stop-vowels	No-noise	0	86	3			
		50	93	6	2.087	0.0815	S
		100	94	5	2.744	0.0336	S
	12	0	86	3			
		50	96	5	3.43	0.014	S
		100	97	2	6.122	0.0808	S
	6	0	85	7			
		50	81	7	-0.806	0.4408	NS
		100	83	7	-0.404	0.7002	NS
Voiced Stop-vowels	No-noise	0	94	6			
		50	97	5	0.788	0.4715	NS
		100	94	4	0	1	NS
	12	0	95	2			
		50	93	5	-0.743	0.4857	NS
		100	87	5	-2.871	0.0246	S
	6	0	87	6			
		50	86	9	-0.185	0.8594	NS
		100	88	5	0.256	0.8064	NS

3. SUMMARY AND CONCLUSIONS

In conclusion, the above results suggested that of the three acoustic segment modifications considered here, increase in release- burst duration and increase in formant transition duration have yielded positive results. Based upon the consistent pattern, it can be concluded that release-burst duration modification by 50% at +12 dB SNR and FTD modification by 100% at +12 dB SNR are beneficial for intelligibility improvement, for plosive-vowel syllables; hence at low-level masking noise the two can be treated as dominant cues for lengthening consonant duration. However, VOT did not appear to be suitable for consonant duration modification.

REFERENCES

[1] Liu, C. and Kewley-port, D., 'Vowel Formant Discrimination for High-fidelity Speech', *J. Acoust. Soc. Am.*, **116,** (2004), 1224-1233.

[2] Bradlow, A. R. Kraus, N. and Hayes, E., 'Speaking Clearly for Children with Learning Disabilities: Sentence Perception in Noise', *J. Speech Lang. Hear. Res.*, **46,** (2003), 80-97.

[3] Bradlow, A. R. and Bent, T., 'The Clear Speech Effect for Non-native Listeners', *J. Acoust. Soc. Am.*, 112, (2002), 272-284.

[4] Ferguson, S. H. and Kewley-port. D., 'Vowel Intelligibility in Clear and Conversational Speech for Normal-hearing and Hearing-impaired Listeners', *J. Acoust. Soc. Am.*, **112,** 259-271.

[5] Krause, J. C. Braida, L. D, 'Investigating Alternative Forms of Clear Speech: The Effects of Speaking Rate and Speaking Mode on Intelligibility', *J. Acoust. Soc. Am.*, **112,** (2002), 2165-2172.

[6] Helfer, K. S., 'Auditory and Auditory-visual Perception of Clear and Conversational Speech', *J. Speech Lang. Hear. Res.*, **40,** (1997), 432-443.

[7] Schum, D., 'Intelligibility of Conversational and Clear Speech of Young and Elderly Talkers', *Journal of American Academy Audiology*, **7**, (1996), 212-218.

[8] Uchanski, R. M. Choi, S., Braida, L. D. Reed, C. M. Durlach, N. I., "Speaking Clearly for the Hard of Hearing IV: Further Studies of the Role of Speaking Rates", *J. Speech Lang. Hear. Res.*, **39,** (1996), 494-509.

[9] Gagne, J. Qarengeer, C. Folkeard, P. Munhall, K. and Mastern, V., 'Auditory, Visual and Audiovisual Speech Intelligibility for Sentence-length Stimuli: An Investigation of Conversational and Clear Speech', *The Volta Review*, **97,** (1995), 33-51.

[10] Payton, Karen, L. Uchanski, Rosalie, M. and Braida, Louis, D., 'Intelligibility of Conversational and Clear Speech in Noise and Reverberation for Listeners with Normal and Impaired Hearing', *J. Acoust. Soc. Am.*, **95,** (1995), 1581-1592.

[11] Chen, F., 'Acoustic Characteristics of Clear and Conversational Speech at Segmental Level', MIT, Cambridge, MA, (1980).

[12] Picheny, M. I. Dulach, N. I. and Briada, L. D., 'Speaking Clearly for the Hard of Hearing I: Intelligibility Differences between Clear and Conversational Speech', *J. Speech and Hearing Res.*, **28,** (1985), 96-103.

[13] Picheny, M. I. Dulach, N. I. and Briada, L. D., 'Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech,' *J. Speech and Hearing Res.*, **29,** (1986), 434-446.

[14] Krause, J. C. Braida, L. D., 'Acoustic Properties of Naturally Produced Clear Speech at Normal Speaking Rates', *J. Acoust. Soc. Am.*, **115,** (1), (2004), 362-378.

[15] Stevens, K. N., 'Acoustic Phonetics' MIT Press, Cambridge, MA, (1998).

[16] Moulines, E. and Charpentier, F., 'Pitch-synchronous Waveform Processing Technique for Text-to-speech Synthesis using Diphones', Speech Communication, **9,** (5), (1990), 453-467.

[17] Liu, S. and Fan-Gang Zeng, 'Temporal Properties in Clear Speech', *J. Acoust. Soc. Am.*, **120,** (1), (2006), 424-432.

[18] Boersma, P. and Weenik, D., 'Praat: Doing Phonetics by Computer', Version 4.4.22, (2005).

[19] Jiang, J. Chen, M. and Alwan, A., 'On the Perception of Voicing in Syllable-initial Plosives in Noise', *J. Acoust. Soc. Am.*, **119,** (2), (2005) 1092-1105.

[20] Thomas, T. G., 'Experimental Evaluation of Improvement in Speech Perception with Consonantal Intensity and Duration Modification', Ph.D. Thesis, Department of Electrical Engineering, Indian Institute of Technology, Bombay, (1996).

[21] Revoile, S. Holden-Pitt. Pickett, J. and Brandt, F., 'Speech Cue Enhancement for the Hearing Impaired: II. Amplification of Burst/murmur Cues for Improved Perception of Final Stop Voicing', *J. Rehab. Res. Dev.,* **24,** (1987), 207-216.

[22] Nittrouer, S., 'Do Temporal Processing Deficits Cause Phonological Processing Problems ?', *J. Speech Lang. Hear. Res.*, **42,** (1999), 925-942.

[23] De Gennaro, S. Braida, L. D. and Dulach, N. I., 'Study of Multi-band Compression with Simulated Sensorineural Hearing Loss', *J. Acoust. Soc. Am.*, **69,** S16, (1981).

[24] Oxenham, A. J. and Moore, B. C. J., 'Modeling the Effects of Peripheral Nonlinearity in Listeners with Normal and Impaired Hearing', In W. Jesteadt (Ed.), 'Modeling Sensorineural Hearing Loss' (Mahwah, NJ: Lawrence Earlbaum Associates) (1987), 273-288.

[25] Sussman, J. E., 'Perception of Formant Transition Cues to Place of Articulation in Children with Language Impairments', *J. Speech Lang. Hear. Res.,* **30,** (1993) 1286-1299.

[26] Flecher, H., 'The Perception of Sounds by Deafened Persons', *J. Acoust. Soc. Am.*, **24,** (1952), 490-497.

[27] S. A. Simpson, & M. Cooke, 'Consonant Identification in N-talker Babble in a Non-monotonic Function of N', *J. Acoust. Soc. Am.*, **118,** (5), (2005), 2775-2778.

[28] H. D. Lewis, V. A. Benignus, K. E. Muller, C. M. Malott, 'Babble and Random Noise Masking of Speech in High and Low Content Cue Identifications, *J. Speech and Hearing Res.*, **31**, (1988), 108-114.

[29] *http://www.lifesci.Sussex.ac.uk/home/Chris_Darwin/praatscripts/.*

[30] Montgomery, A. A. and Edge, R. A., 'Evaluation of Two Speech Enhancement Techniques to Improve Intelligibility for Hearing-impaired Adults', *J. Speech and Hearing Res.*, **31,** (1988), 386-393.

[31] Bilger, R. C. and Wang, M. D., 'Consonant Confusions in Patients with Sensorineural Hearing Loss', *J. Speech and Hearing Res.*, **19,** (1976), 718-748.

[32] Miller, G. A. and Nicely, P. E., 'An Analysis of Perceptual Confusions Among Some English Consonants', *J. Acoust. Soc. Am.*, **27,** (2), (1955), 338-352.

[33] Ladefoged, P., 'A Course in Phonetics', Orlando: Harcourt Brace, 4th ed. 2001, Boston, (1975).