Speech Intelligibility Enhancement in the context of Hearing -Impairment

Dr N H SHOBHA

Department of Electronics & Communication Engineering, Methodist college of Engineering & Technology, Hyderabad, India

ABSTRACT

The current paper focuses on the enhancement of an important attribute of clear speech, consonant-vowel Intensity Ratio on listeners suffering from Sensorineural hearing loss. This work is an extension of author's previous papers. A case for synthetic clear speech in the context of hearing impairment was tested on (i) Hearing Impaired subjects, and (ii) Normal hearing subjects in the presence of masking noise (+6dB, +12 dB). Consonant recognition in noise free and noisy situations using non-sense syllable test was investigated. The Stops and Fricative consonants of English language with cardinal vowels were processed for consonant-vowel intensity modifications ranging from 0dB to + 12dB at 3dB step. The speech perception in noise tests were quantified in terms of Consonant Recognition scores and Information transmission analysis measures. The results reported that the consonant-vowel intensity ratio modification of +9 to +12dB has positive effect on speech intelligibility improvement for both Hearing-impaired and Normal-hearing listeners. The maximum intelligibility benefit in Information transmitted scores were reported as 44, 45 percent points for HI listeners and 18, 31 percent points for masked normal-hearing listeners for /SV/ and /FV/ syllables respectively.

Keywords:- Consonant-vowel intensity ratio; Non-sense syllable test; Speech perception in noise; Speech intelligibility; Consonant Recognition

1.INTRODUCTION

Listeners almost always listen to speech, which is degraded by the addition of competing speech and non-speech signals. Unfortunately, hearing-impaired listeners often have the greatest difficulty in understanding speech in noisy environments as they have difficulty in isolating a specific speech signal from the background noise and understanding what is said. Sensorineural hearing losses (SNHL) are due to reduced sensitivity of the neural receptor that distorts the perception of sounds. The SNHL is characterized by compression in dynamic range, elevated hearing threshold, poor frequency and temporal resolutions. SNHL are not amenable to medical intervention and patients need to use the hearing-aids for speech perception. Owning to the complexity of the speech perception in general and consonant perception in particular, there is a need for more efficient algorithms in hearing aids. The present work is one such contribution to a broad research line whose aim is to develop improved hearing aids for people suffering from SNHL.

When listeners are confronted with difficult environments or when speaking to hearing-impaired person, humans instinctively change the way they speak and adopt a speaking style called *clear speech*. The acoustic analysis show that naturally produced clear speech typically involves a wide range of acoustic & articulatory adjustments or special attributes [1,2,3] such as more salient consonant contrasts (enhanced consonant-vowel intensity ratio-CVR, longer formant transitions, less vowel reduction[1,4,5],decrease in speaking rate (longer segments), wider dynamic range, greater sound-pressure levels, more salient stop releases, greater rms intensity of non-silent portions(release burst, frication, and/or aspiration) of obstruent consonants. Pre-processing speech for some of these parameters is expected to improve speech intelligibility for impaired listeners [1,4,5] and speech development in HI children[6] .

Two important temporal attributes of clear speech that are found to increase at phoneme level are the consonant-vowel Intensity ratio (CVR) and the consonant duration (CD). The process of strengthening CVR and CD are said to increases the salience of the consonant cues to weaken the masking effect or in other words results in reducing the vowel emphasis. Vowel perception is relatively simple for the hearing impaired while consonant perception is an area of continuing controversy. The present work focused on the role of CVR Modifications on speech perception in the context of hearing-impairment.

2. METHOD

2.1 Subjects

Two category of subjects such as Type 1 and Type 2 participated in listening/perception tests. Type 1 listeners were Hearing Impaired (HI) listeners, and Type 2 were Normal Hearing (NH) listeners. None of the subjects were experienced with perceptual experiments; subjects went through a speech token familiarization training session before the experiment started.

Figure 1 below illustrates the hearing losses and their associated thresholds in human beings, (i) Normal hearing :-10 to 25dB, (ii) Mild hearing loss: 26 to40 dB, (iii) Moderate hearing loss: 41 to 55dB, (iv)Severe hearing loss: 56 to 90dB, (v) Profound hearing loss: >91dB.

Type -1 Subjects: Five senior citizens in the age group of 58- 62 years, native listeners with hearing impairment ranging from moderate to severe hearing loss ranging from 45 dB to 85 dB in both ears. These listeners underwent audiological investigation with Pure Tone Audiometry (PTA) tests in speech and hearing clinic.

Type- 2 Subjects: Five subjects in the age group of 16 -45 years, native normal hearing individuals, participated in the listening experiments. They were tested for normal hearing as per PTA tests with less than 25dB hearing thresholds indicative of normal hearing ability.

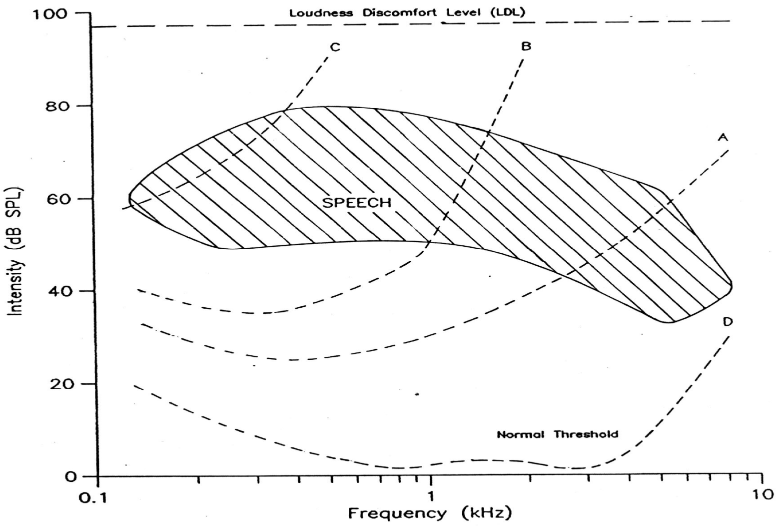

Figure 1. Sensorineural Hearing Thresholds: Hatched portion- Normal speech spectrum , Curve A-Mild to Moderate HL;B-Severe HL, C-Profound HL;D-Normal Hearing [6]

2.2 Target Stimuli System

The Stop consonants - /p, t, k, b, d, g/, and the Fricative consonants - /f, θ , s, v, ð, z/, with the accompanying cardinal vowels /a, ɛ, o/ and /a, i, u/ respectively, to form nonsense syllables were used as target stimuli system. The baseline set of 18 syllables are referred as Stop-Vowels (/SV/) and Fricative-Vowels (/FV/), further grouped into voiceless and voiced sub-sets. In total, the baseline stimuli in /SV/ context are: (i) voiceless sub-set: /pa, p ɛ , po, ta, t ɛ , to, ka, k ɛ , ko/ / (ii) voiced sub-set: /ba, b ɛ , bo, da, d ɛ, do, ga, gɛ , go/; similarly in /FV/ context are: (i) voiceless sub-set: /fa, fi, fu, θa, θi, θu, sa, si, su/ and (ii) voiced sub-set: /va, vi, vu, ða, ði, ðu, za, zi, zu/.

Stops and Fricative consonants are short speech sounds and people suffering from hearing loss are often said to have greatest difficulty in identifying them. 'Stops' are 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 voice bar. 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. 'Fricatives' are produced when the turbulent air-flow occurs at a point of narrow constriction in the vocal tract. Fricative consonants are characterized by a high frequency noise of sufficient duration. In general, stops and fricatives are both characterized by high frequency random noise, which occurs due to opening of oral cavity. The duration of the high frequency noise is longer and its intensity is greater for voiceless than for voiced for stops/fricatives. These consonants are common to all regional accents in Indian English.

2.3 Speech Signal Processing

The experiment constituted two phases, in the first phase or the signal processing phase, the speech stimuli were processed for CVR modifications comprising of selection and modifications of consonant segments. In the second phase the developed database was subjected to perception tests and the results were analyzed quantitatively to assess the benefit on speech intelligibility. A detailed explanation is as below

In the first stage of signal processing, the natural speech tokens were recorded and were subjected to resynthesis. The natural stimuli were recorded in a quiet room, sampled at 44.1 kHz, using a Praat monosound recorder. The best utterance out of 20 utterances of the author (middle aged, female) was selected based on the clarity and stress. The speech tokens were subjected to resynthesis using the procedure of LPC (linear prediction) analysis-synthesis as provided in Praat [7]. The idea behind the resynthesis was to get synthetic copy which renders efficient and independent manipulation of the spectral, temporal and intensity characteristics, and sounds as similar as possible to a human utterance. After the process of resynthesis, the synthesized tokens referred as baseline syllables were normalized to 70 dB Intensity level to avoid the signal clipping in subsequent processing stages.

In the second stage of signal processing, consonant-vowel intensity modifications (CVRM) were carried out on the baseline syllables set. CVR is defined as the difference in decibels between either the power/energy of the consonant and that of the adjoining vowel. CVRM can be achieved either by reducing the intensity of vowel or by increasing the intensity of the preceding consonant. The latter method has been reported to be more efficient over the former [8, 9]. The baseline syllables were manipulated under five CVRMs such as, 0 dB, +3 dB, +6 dB, +9 dB, and +12 dB, where 0 dB refers to the unmanipulated set (natural). In the process, the consonant and vowel segments were identified on simultaneous consultation with timing and spectrogram waveforms with repeated visual and auditory monitoring. The intensity of the vowel segment was fixed while that of the consonant segment was adjusted to the required CVR level. CVR modification was restricted to +12 dB so as to avoid the possibility of weak-vowel cue [10].

CVR modifications and their effects on syllable intensities are displayed for some syllables /ba/ and /va/ in Figures 2 and 3. These sample graphs obtained for /SV/ and /FV/ set, have established an enhancement in consonant intensity with respect to the proceeding vowel intensity with respect to five modification levels(0 dB, +3 dB, +6 dB, +9 dB, +12 dB) under study.

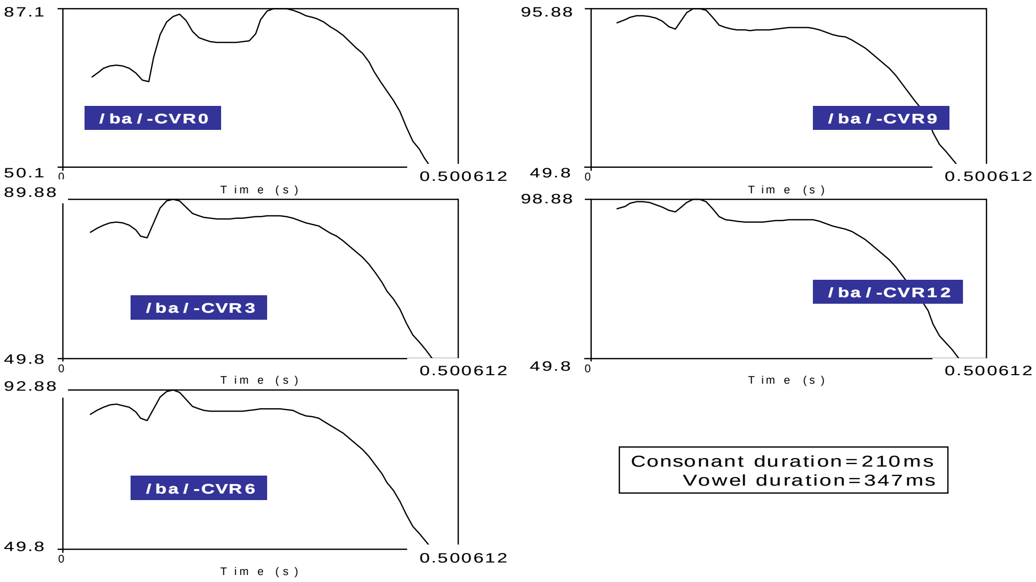

Figure 2. Intensity-time plots for stop-vowel syllable /ba/ for 0 to 12 dB CVR modifications

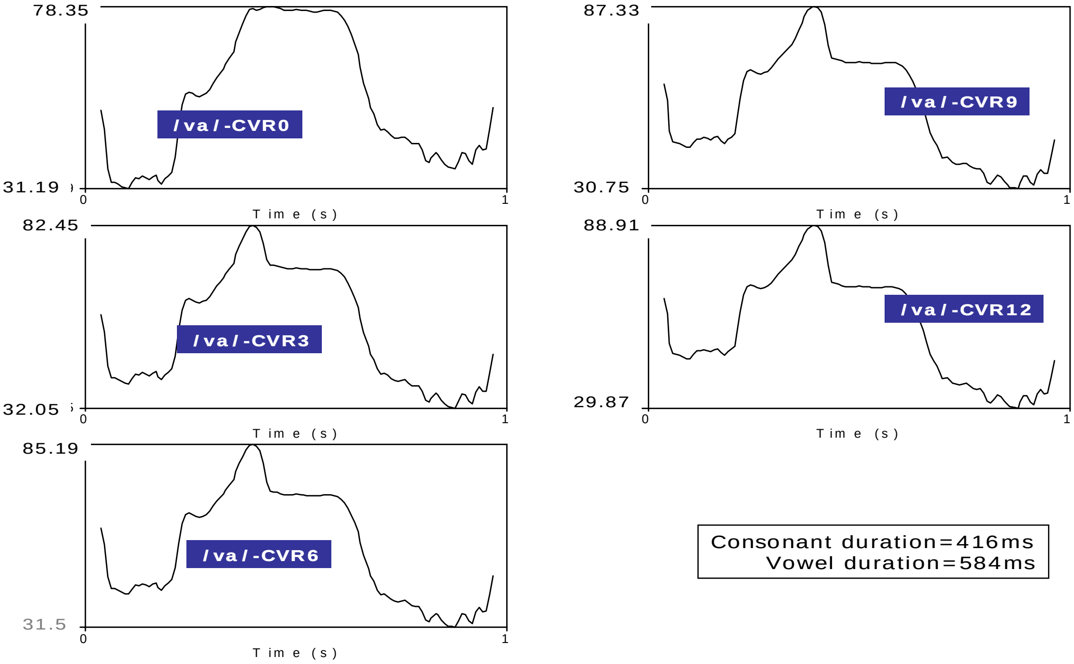

Figure 3. Intensity-time plots for fricative-vowel syllable /va/ for 0 to 12 dB CVR modifications

2.4 Speech in Noise Task

The third stage of processing holds relevance for Type 2 (NH) context stimuli but not for Type 1 (HI) context. In other words, the processed/CVR modified stimuli from the second stage served as test stimuli for Type 1 without the presence of noise masking . Thus Type 1 stimuli corpus held 90 (=18*5) tokens for /SV/ and /FV/ syllables respectively.

For Type 2 context, this stage of processing was designed to simulate hearing impairment by reducing the acoustic dynamic range. As reported in the literature, the reduction of the acoustic dynamic range (elevated hearing threshold) can be simulated by addition of white (Gaussian) noise [11, 12]. This masking noise responsible for the threshold elevation is believed to be predominantly of cochlear origin [11]. Some researchers have employed multi-talker babble instead of white noise [13, 14, 15]. However, due to its non-stationary nature, the effective masking it may provide during stimulus presentation is unpredictable. Simulation of hearing impairment is a complicated phenomenon. However, noise masking does to some extent; provide a partial model for some of the perceptual effects of hearing impairment. Hence, it was decided to use white noise masker, Praat script was run for the generation of white noise [16] . The CVR modified tokens from the second stage were now additively mixed along with the synthesized noise at three noise conditions, *ie*, no-masking noise, +12 dB and +6dB SNRs, where noise free (natural) tokens were considered to be those at no-masking noise.. The SNR refers to the ratio of the average power in CV token to the average power of the noise token in decibels.

$$\mathrm{SNR}_{\mathrm{dB}}=10\log_{10}\left(\mathrm{P}_{\mathrm{Avg\text{-}CV}} / \mathrm{P}_{\mathrm{Avg\text{-}Noise}}\right) \qquad \text{…… (1)}$$

The average power level of the speech token was fixed while that of the noise was adjusted for fixing the SNR to the required value. The process of mixing processed syllable with masking noise was accomplished by a PRAAT script [16]. In this algorithm, sounds are summed up by point-to-point values, preserving real time across different time domains and sampling rate. Finally, for Type 2 context, the stimuli corpus holds 270(=18*5*3) tokens spanning across 18 baseline stimuli with 5 CVR modifications and 3 SNR conditions.

3. EXPERIMENTAL SESSIONS

The perception tests were automated using MATLAB code with graphic user interface. Stimuli were presented using computerized testing procedure at the most comfortable listening level of 75 dB to 85 dB SPL for the listeners. The set of tokens at one CVR level (0/3dB/6dB/9dB/12dB) and one SNR level (noise free/+12dB/+6dB) were administered to listeners at one test run. 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. Each run lasted for 20 -25 min, spanning a period of 6 -8 hrs for the entire experimentation per listener. Results were cast into three groups of six by six confusion matrices (CM) per run; sub-matrices (3*3) were derived for analyzing the effect on the production-based categories [17].

3.1 Experimental Measures

Speech discrimination test results were summarized as the percentage of correct responses for many experimental runs. The diagonal cell entries in the stimulus-response confusion matrix correspond to the correct responses and the off-diagonal entries correspond to the confusion errors.

Recognition Score: The sum of these diagonal elements gives the empirical probability of correct responses, known as Recognition Score RS (or articulation score). Though computation of RS is simple, it obscures the detailed and important information on the distribution of errors among the off-diagonal cells [17] ,also it is sensitive to the subject's bias or chance scoring (an artificially high score).

$$Rs=\sum_{i=1}^{n}P(Xi;Yj) \qquad \text{…… (2)}$$

Relative information Transmission Score:

The Information Transmission analysis approach [10, 17, 18, 19] provides a measure of covariance between stimuli and responses. It 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[10,17,18, 19]. The information measures of the input stimulus X and output response Y are defined in terms of the mean logarithmic probability MLP,

$$I(X;Y)=-\sum_{i}\sum_{j}p(x_i,y_j)\,\log_2\left(\frac{p(x_i)\,p(y_j)}{p(x_i,y_j)}\right) bits \qquad \text{……. (3)}$$

The Relative information Transmission (Rtr) score from X to Y is given by,

$$Rtr\left(X;Y\right)=\frac{I\left(X;Y\right)}{I_s\left(X\right)} \qquad \ldots\ldots (4)$$

where, $I_s(x)$ is referred as the information measure of the input-stimulus in terms of MLP.

4. RESULTS

The speech perception in noise tests (SPIN) were quantified using two intelligibility scores (i) RS: Recognition scores, (ii) Rtr: Relative Information Transmitted scores. As per the previous discussion, even though relative information transmitted scores are better estimates of intelligibility measures, in this paper both the scores are analyzed. The perceptual ability of subjects with five versions of CVRM (0, +3, +6, +9, +12 dB) are reported in Tables 1, 2 and 3. The Intelligibility benefit which was defined as the percentage rise in Mean Score for CVR modification from 0 dB to the new CVR level under consideration (+3/+6/+9/+12dB).

4.1 Type- 1 Subjects

For Type -1 (HI) listeners, Recognition Scores and Relative Information transmitted scores are reported in Tables 1 and corresponding graphs in Figures 4. The mean scores for voiced and voiceless /SV/ and /FV/ syllables averaged across all listeners for individual vowel context are reported in the table. The scoring pattern reported the effect of CVRM on 'voiceless/voiced syllable' recognition and their vowel dependency. The stimuli included voiced and voiceless consonants under a single vowel context forming three subsets of 6 syllables for /SV/ and /FV/ syllables. For /SV/ set syllable subsets were, {/pa/, /ta/, /ka/, /ba/, /da/, /ga/}, {/pɛ/, /tɛ/, /kɛ/, /bɛ/, /dɛ/, /gɛ/}, {/po/, /to/, /ko/, /bo/, /do/, /go/}; similarly for /FV/ set subsets were, {/fa/, /θa/, /sa/, /va/, /ða/, /za/}, {/fi/, /θi/, /si/, /vi/, /ði/, /zi/}, {/fu/, /θu/, /su/, /vu/, /ðu/, /zu/}.

The results of the perceptual analysis for hearing impaired subjects were analyzed with two scoring patterns, Recognition scores and Relative information transmission scores .Table 1 and Figure 4 have summarized these scores for /SV/ and /FV/ syllables, where the identification ability of subjects were compared across five versions of CVRM's 0, +3, +6, +9, +12 dB. The perceptual analysis results have reported positive benefit on speech intelligence ranging from marginal to significant levels with CVR modifications. The maximum intelligibility benefit in Recognition Scores (RS) were, 56% (for Stops) and 47% (for fricatives) ; while maximum intelligibility improvement in Relative information transmitted scores (Rtr) were 44% (for stops) and 45% (for fricatives) .

4.2 Type- 2 Subjects

For Type-2 (NH) listeners, Recognition Scores are summarized in Tables 2 and the corresponding graphs in Figure 4; while the Relative Information transmitted scores are summarized in Tables 3 and the corresponding graphs in Figure 5. The mean scores for voiced and voiceless consonants averaged across all listeners for individual vowel context are recorded in these tables. The scoring pattern reported the effect of CVRM on 'voiceless/voiced syllable' recognition and their vowel dependency as a function of SNR (no-noise, +12dB, +6dB). The perceptual analysis results have reported positive benefit in speech intelligence albeit by different amounts.

According to Table 2, maximum intelligibility benefits in RS at 6 dB and 12 dB SNRs are summarized as follows, /SV/ syllables reported :15% (SNR=6dB) and 19% (SNR=12dB) for unvoiced stop syllables /o/context; while /FV/ syllables reported: 15% (SNR=6dB) and 16% (SNR=12dB) for voiced fricative consonant in /i/ context.

According to Table 3, maximum intelligibility improvement in Rtr at 6dB and 12 dB SNRs are summarized as follows, /SV/ syllables reported: 17% (SNR=6dB) and 18% (SNR=12dB) for unvoiced stop consonant in /o/context; while /FV/ syllables reported: 31% (SNR=6dB) and 27% (SNR=12dB) for unvoiced fricative consonant in /i/context.

4.3 Implication of Results

The results of our investigation based on the perceptual analysis tested on HI (Type- 1) and masked NH (Type- 2) listeners, involving two sets of stimuli (Stops and Fricatives), analyzed with two intelligibility scores (RS and Rtr) have altogether indicated a positive impact on speech intelligibility with CVR enhancement. The results suggest that there is an improvement in scores with respect to increasing CVR for voiceless and voiced consonant recognition, as well as the overall consonant recognition.

For stops and fricatives, the CVR modification of +9 and +12dB have contributed significant improvement of intelligibility in the presence of noise. It is also to be noted that the improvements due to processing are more for higher level of masking noise(+6dB), indicating an improvement in listening condition with processing. Though both the stops and fricatives have reported positive intelligibility improvement, the benefits were higher for fricatives compared to stops, in other words fricatives have reported better intelligibility improvement in the presence of noise. Further, in the presence of masking noise - the voiceless and voiced syllables exhibited nearly equal percent-benefits for stop-vowels; while voiceless syllables exhibited higher benefit than their voiced counterparts for fricative-vowels.

Table 1. Type-1 Subjects: Recognition scores and Relative information transmitted scores

Test Stimuli	Vowel context	Test CV9 Consonant Recognition Scores(%)					Test CV9 Rel. Infor. Transmitted(%)			
		CVR (dB)					CVR (dB)			
		0	3	6	9	12	0	3	9	12
Voiceless stop-vowels	/a/	43	60	73	67	60	14	24	54	58
	/ɛ/	50	53	67	53	50	44	19	20	22
	/o/	67	73	87	97	93	53	60	90	80
Voiced stop-vowels	/a/	73	60	70	77	100	59	44	54	100
	/ɛ/	27	83	73	83	73	27	59	61	60
	/o/	100	100	100	93	80	100	100	84	61
Voiceless Fricative-vowels	/a/	73	100	83	100	97	52	100	100	90
	/i/	30	60	70	77	63	15	59	54	58
	/u/	50	73	70	70	87	56	65	60	75
Voiced Fricative-vowels	/a/	100	100	97	97	100	100	100	100	100
	/i/	50	67	100	100	97	45	58	100	90
	/u/	97	100	100	100	100	90	100	100	100

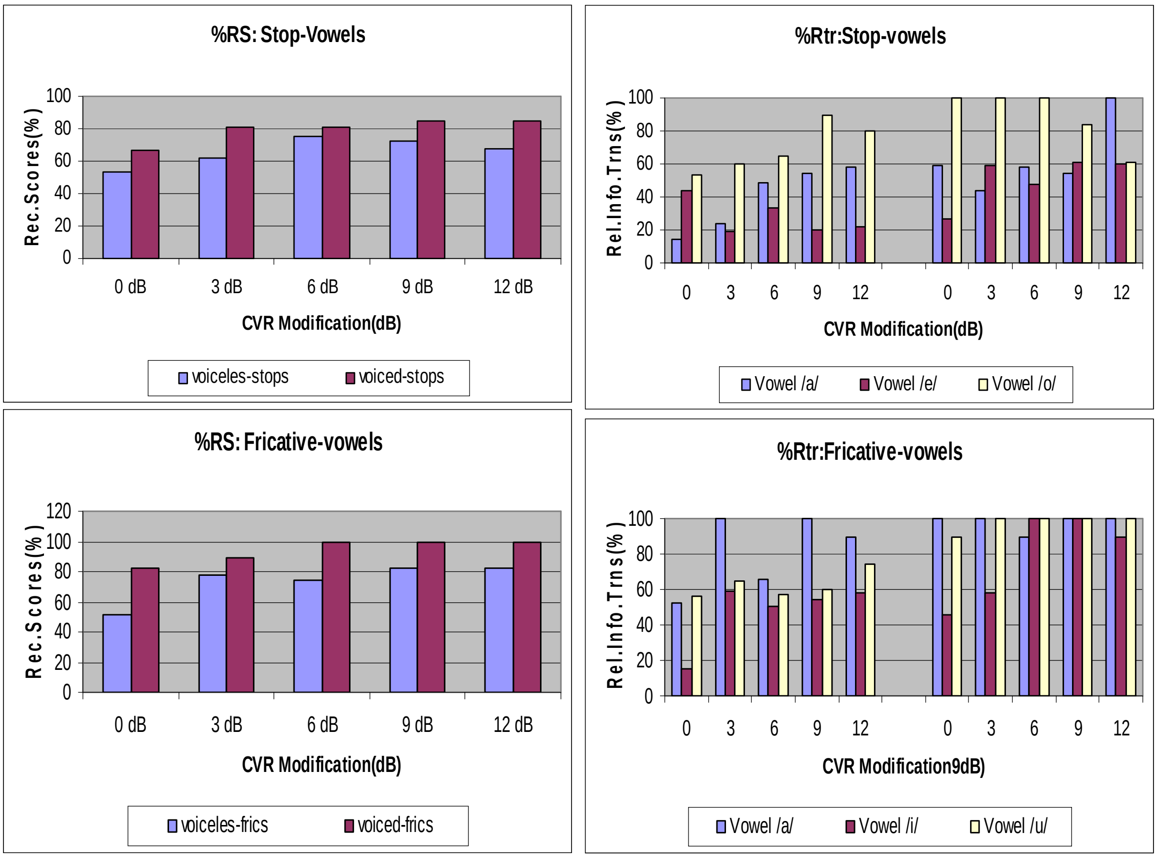

Figure 4. Recognition scores and Relative information transmitted pattern corresponding to Table 1
Top panel: - /SV/ syllables, Bottom panel: - /FV/ syllables

Table 2. Type-2 Subjects: Recognition score pattern for voiceless/voiced syllables

Consonant Recognition Scores :CV9																
Test Stimuli	Vowel context	No masking noise					SNR=12 dB					SNR=6 dB				
		CVR (dB)					CVR (dB)					CVR (dB)				
		0	3	6	9	12	0	3	6	9	12	0	3	6	9	12
Voiceless stop-vowels	/a/	88	98	98	97	99	81	83	86	87	86	78	82	83	85	85
	/ɛ/	95	97	98	99	99	92	94	95	96	97	87	90	93	94	93
	/o/	97	99	97	100	100	70	73	75	82	89	69	73	74	78	84
Voiced stop-vowels	/a/	94	96	98	98	98	71	75	78	77	78	68	70	73	75	77
	/ɛ/	97	99	97	99	99	89	91	91	94	99	87	89	89	92	93
	/o/	97	99	98	98	97	89	89	92	95	96	87	90	91	94	95
Voiceless Fricative-vowels	/a/	99	98	99	100	97	77	79	83	87	93	81	82	84	86	87
	/i/	97	99	96	97	95	77	78	81	87	89	77	86	86	87	87
	/u/	100	100	99	97	98	79	81	90	89	87	77	82	82	84	87
Voiced Fricative-vowels	/a/	94	97	98	98	96	81	83	85	87	91	80	81	81	85	89
	/i/	97	99	94	95	100	78	80	81	83	94	74	85	86	88	89
	/u/	98	97	99	99	100	86	86	89	90	95	89	88	82	87	92

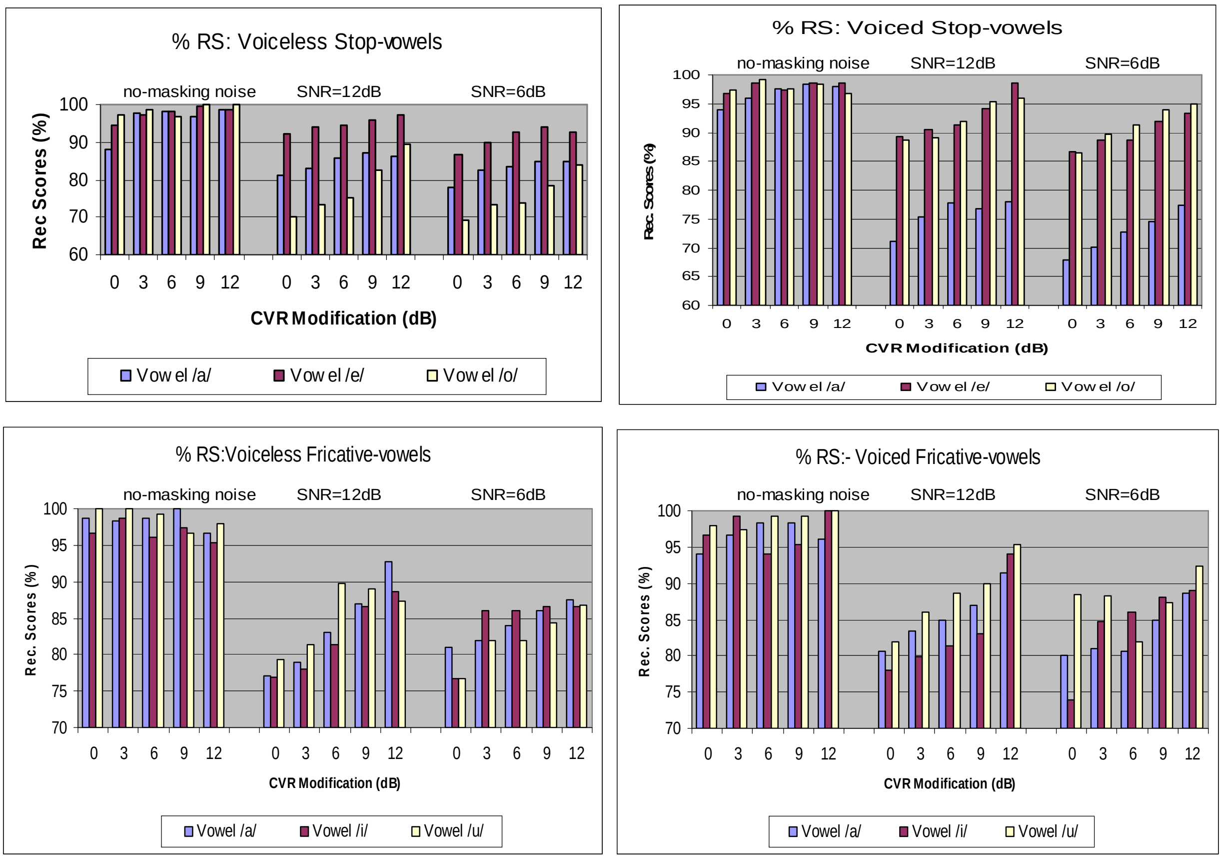

Figure 5. Recognition score Graphs with reference to Table 2 Top panel: - /SV/ syllables, Bottom panel: - /FV/ syllables

Table 3. Type-2 Subjects: Relative information transmitted scores pattern for voiceless/voiced syllables

Information transmission analysis: CV9																
Test Stimuli	**Vowel context**	**Relative information transmitted (%)**														
		No masking noise					**SNR=12 dB**					**SNR=6 dB**				
		CVR (dB)					**CVR (dB)**					**CVR (dB)**				
		0	**3**	**6**	**9**	**12**	**0**	**3**	**6**	**9**	**12**	**0**	**3**	**6**	**9**	**12**
Voiceless Stop-vowels	**/a/**	100	100	100	100	100	90	93	94	93	96	89	91	92	94	95
	/ε/	96	95	96	98	98	90	93	90	89	96	79	87	88	89	87
	/o/	100	98	98	100	100	65	69	67	71	83	58	61	62	71	75
Voiced Stop-vowels	**/a/**	100	100	98	100	98	94	94	96	98	98	82	85	87	93	96
	/ε/	100	100	100	96	100	91	92	92	94	96	83	87	87	88	88
	/o/	96	100	100	98	98	92	94	95	97	96	90	94	94	95	96
Voiceless Fricative-vowels	**/a/**	100	100	100	100	100	93	95	95	100	100	89	88	97	98	98
	/i/	97	95	96	98	98	73	96	98	98	100	67	85	87	91	98
	/u/	100	98	100	98	100	77	89	94	97	97	74	80	85	90	98
Voiced Fricative-vowels	**/a/**	100	100	100	100	98	92	92	93	92	98	92	88	96	94	100
	/i/	100	100	98	100	100	78	94	96	96	98	79	83	95	98	100
	/u/	99	100	100	100	100	90	91	92	96	99	88	91	92	94	99

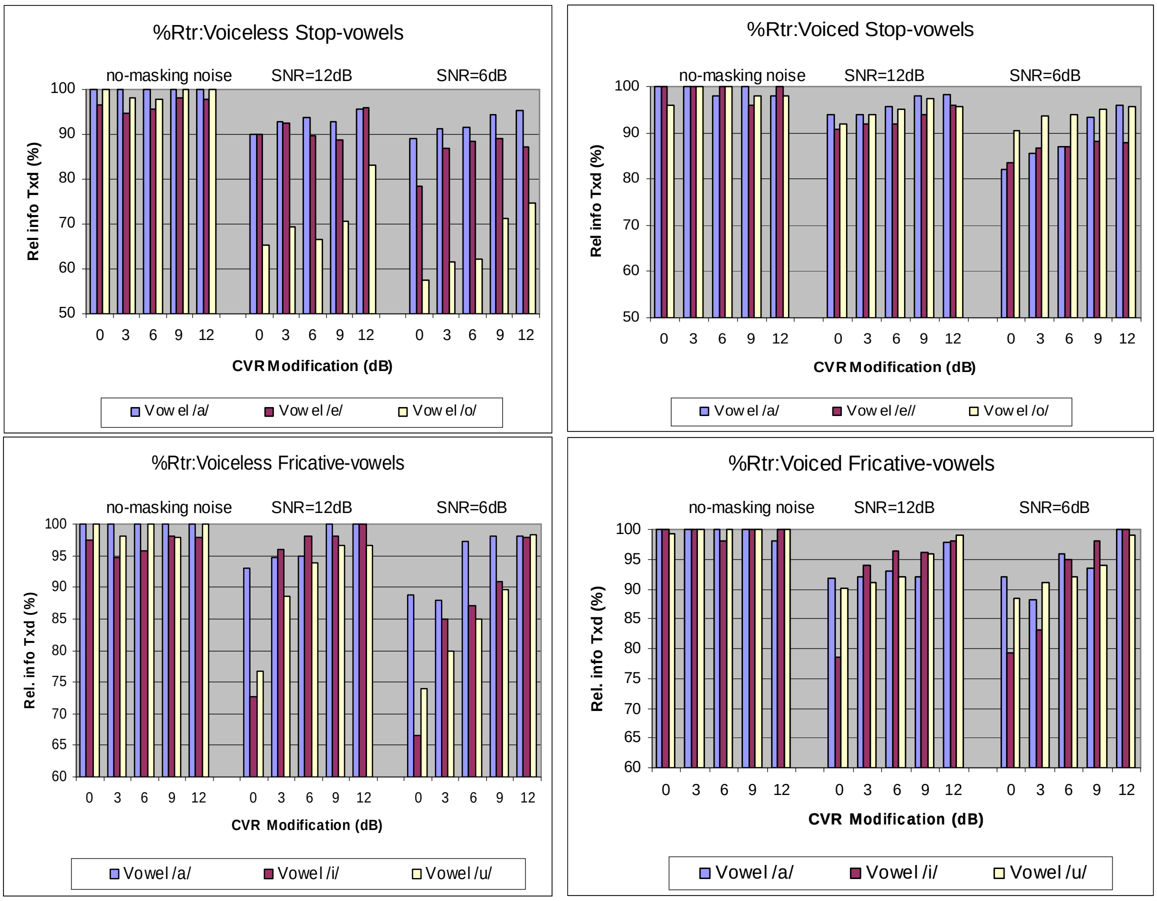

Figure 6. Relative information transmitted pattern with reference to Table 3
Top panel: - /SV/ syllables, Bottom panel: - /FV/ syllables

4.4 Discussion

A substantial body of previous research, has provided with important insights into the characteristics and benefits of clear speech. It is worthwhile to explore a few of recent studies in correlation with the present investigational results. However, it is crucial to understand that the variability in benefits across different studies depends on speaker- listener effects, signal-dependant effects, implementation of clear speech strategies etc. The focus of the reviewed studies shown below has exclusively used recorded speech material; on shifting the focus to synthesized clear speech material as in the present investigation, effectively revealed what acoustic-phonetic features are present/absent and how their manipulation affected intelligibility.

In a notable work by Ferguson and Kewley-port [20], conversational and clear vowels were recorded by one male talker in a CVC context and mixed with 12-talker babble. The stimuli were presented to young normal hearing adults and elderly HI listeners, they found that listeners with normal hearing enjoyed a 15 percent points clear speech advantage for vowels presented in noise, while elderly HI listeners showed no clear speech benefit.

In a detailed exploration of talker variability, Ferguson [21], collected conversational and clear speech from 41 talkers, thus created a large database that allows assessments of the relationship between acoustic-phonetic variations across talkers' variability in intelligibility. The study showed that for 41 talkers, clear speech vowel intelligibility for normal hearing listeners in noise varied widely from 12 to 33 percent points. Female talkers tended to produce more intelligible clear speech compared to male talkers despite having similar conversational intelligibility scores. Similar variability in clear speech intelligibility across talkers was found for words and sentences by Gagne et al. (1994), Schum, (1996), suggesting that intelligibility variability is characteristic of individual talker's production rater than vowel intelligibility only.

Bradlow et al. [22] explored clear speech perception in children with and without learning disabilities, in order to see whether they would be able to utilize clear speech cues to the same extent as the more experienced adults. Although, children with learning disabilities exhibited lower intelligibility scores compared to children with no deficits, the clear speech effect was substantial for both groups: 9% and 16-18%. This study demonstrated that speech perception for children with learning problems may be enhanced in everyday communication by employing a simple strategy of speaking clearly, which can be adopted in clinical settings and in educational settings.

Liu et al. [23] compared clear speech perceptual benefits for normal hearing adults and adults with cochlear implants varied in age of implant use. Using a slightly different approach in this study, conversational and clear speech intelligibility was measured as a function of S/N ratios. Both groups of listeners a significant clear speech advantage, although listeners with cochlear implants needed somewhat better S/N ratios in order to perform at the same level as normal-hearing adults.

Thomas et al. [24] studied the role of the two important attributes of clear speech namely, the consonant-to-vowel intensity ratio (CVR) and the consonant duration (CD) for dasiasynthetic clear speechpsila in the context of the persons with hearing impairment. The results suggested that the CVR and the formant transition duration play an important role in speech perception.

Maniwa et al. [25] conducted two experiments to determine whether clear speech enhances fricative intelligibility for normal hearing and for listeners with simulated impairment. The experiments measured babble signal-to-noise ratio thresholds for fricative minimal pair distinctions for 14 normal-hearing listeners and 14 listeners with simulated sloping, recruiting impairment. Results indicated that clear speech helped both groups overall. However, for impaired listeners, reliable clear speech intelligibility advantages were not found for non-sibilant pairs. Correlations between acoustic and perceptual data were less consistent for listeners with simulated impairment, and suggested that lower-frequency information may play a role.

Based on the present and past investigations of the author[19,26], CVR enhancement by +9 to +12dB levels have enjoyed significant intelligibility benefit in the context of hearing impairment, albeit by different amounts. The tests have reported that the maximum intelligibility benefit in relative information transmitted as 44, 45 percent points (for HI listeners) and 18, 31 percent points (for masked normal-hearing listeners) for /SV/ and /FV/ syllables respectively. The complex interactions between C-V modification, consonant feature, and vowel context indicate that the benefits of clear speech are not uniform for all features and contexts.

5. CONCLUSIONS

The work focused on speech perception on people with varying degree of hearing loss, using synthetic or digitally manipulated features of natural speech to the clear speech advantage. The avenue of research focused on an interesting approach to create more intelligible speech, based on clear speech attribute- consonant vowel intensity ratio. The results of the investigation involving two sets of stimuli analyzed with two intelligibility scores have altogether indicated a positive impact on speech intelligibility. For stops and fricatives, the CVR modification of +9 and +12dB have contributed significant improvement, hence an adjustment of the C-V intensity ratio can yield significant improvement in consonant recognition for hearing-impaired listeners in quiet, and for normal hearing listeners in noise [27-31]. The CVR level for hearing benefit for both hearing-impaired and masked normal-hearing listeners have been almost equivocal. In summary, the findings suggested that efforts to emphasize potentially weak consonant amplitude should be beneficial in surmounting some of the speech recognition difficulties of hearing impaired listeners with sensorineural hearing loss.

REFERENCES

[1] Liu, S., and Zeng, F., "Temporal properties in clear speech perception," Journal of Acoustic Society of America, 120(1), 424-432, 2006
[2] J C Krause & L D Braida,, "Investigating alternative forms of clear speech:the effects of speaking rate and speaking mode on intelligibility," Journal of Acoustic Society of America., vol 112, 2165-2172, 2002.
[3] J C Krause, L D Braida., "Accoustic properties of naturally produced clear speech at normal speaking rates," Journal of Acoustic Society of America , vol 115 (1), 362-378, 2004.
[4] M I Picheny, N I Dulach and L D Briada. "Speaking Clearly for the Hard of Hearing I : Intelligibility Differences between Clear and Conversational Speech." Journal of Speech Lang and Hear Res, vol 28, 1985.
[5] J C Krause and L D Braida. "Accoustic Properties of Naturally Produced Clear Speech at Normal Speaking Rates," Journal of Acoustic Society of America, vol 115, no 1, 2004.
[6] N Schiavetti, R L Whitehead and D E Metz. "The Effects of Simultaneous Communication on Production and Perception of Speech," Journal of Deaf Studies and Deaf Education, vol 3, 2004.
[7] P Boersma and D Weenik. "Praat : Doing Phonetics by Computer" vol 4.4.22, 2005.
[8] M I Picheny, N I Dulach and L D Briada. "Speaking Clearly for the Hard of Hearing I : Intelligibility Differences between Clear and Conversational Speech," Journal of Speech Lang and Hear Res, vol 28, 1985.
[9] M I Picheny, N I Dulach and L D Briada. "Speaking Clearly for the Hard of Hearing II : Acoustic Characteristics of Clear and Conversational Speech," Journal of Speech Lang and Hear Res, vol 29, 1986.
[10] T G Thomas. "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.
[11] S DeGennaro, L D Braida and N I Dulach. "Study of Multi-band Compression with Simulated Sensorineural Hearing Loss," Journal of Acoustic Society of America, vol 69, 1981, S16.
[12] A J Oxenham and B C J Moore. "Modelling the Effects of Peripheral Non-linearity in Listeners with Normal and Impaired Hearing. W Jesteadt (ed)." N J Mahwah. 'Modelling Sensorineural Hearing Loss.' Lawrence Earlbaum Associates, 1987, p 273.
[13] H Flecher. "The Perception of Sounds by Deafened Persons," J Acoust Soc Am, vol 24, 1952, p 490.
[14] S A Simpson and M Cooke. "Consonant Identification in N-talker Babble in a Non-monotonic Function of N," Journal of Acoustic Society of America ,vol 118, no 5, 2005, p 2775.
[15] H D Lewis, V A Benignus, K E Muller and C M Malott. "Babble and Random Noise Masking of Speech in High and Low Content Cue Identifications," Journal of Speech and Hearing Response, vol 31, 1988, p 108.
[16] http://www.lifesci.Sussex.ac.uk/home/Chris_Darwin/praatscripts/
[17] G A Miller and P E Nicely. "An Analysis of Perceptual Confusions among Some English Consonants," Journal of Acoustic Society of America , vol 27, no 2, 1955, p 338.
[18] R C Bilger and M D Wang. "Consonant Confusions in Patients with Sensorineural Hearing Loss," Journal Speech and Hearing Res, vol 19, 1976, p 718.
[19] N H Shobha, T G Thomas and K Subbarao. "Influence of Consonant-vowel Intensity Ratio on Speech Perception for Hearing Impaired Listeners," Proceedings of 2008 Congress in Image and Signal Processing, Hainan, CHINA, 2008, p 342.
[20] Ferguson,S.H., & Kewley-port, D. "Vowel intelligibility in clear &conversational speech for normal-hearing & hearing-impaired listeners," Journal of Acoustic Society of America , 112, 259-271.
[21] Ferguson. (2004). "Talker differences in clear & conversational speech; vowel intelligibility for normal-hearing listeners," Journal of Acoustic Society of America , 116(4), 2365-2373.
[22] Bradlow, A.R., Kraus. N., & Hayes. E. "Speaking clearly for children with learning disabilities: Sentence perception in noise," Journalof Speech Lang. Hear. Res. 46, 80-97.
[23] Liu, S., Del Rio, E., Bradlow, A.R., Zeng, E.G. "Clear speech perception in acoustic & electrical hearing," Journal of Acoustic Society of America, 116, 2374-2383.
[24] Thomas, T.G., Sekhar, S.C. " Perception studies on the attributes of synthetic clear speech for the hard of hearing," Proc. Signal processing & its Application, ISSPA, 1-4.
[25] Maniwa, K., & Jongman, A. " Perception of clear fricatives by normal-hearing & simulated hearing-impaired listeners," Journal of Acoustic Society of America , 123(2), 1114-1125,2008.
[26] N H Shobha, T G Thomas and K Subbarao ."Experimentation to Improve Consonant Recognition for the Hearing Impaired listeners", International Interdisciplinary Journal, Institute of Engineers-INDIA, May 2010.

http://www.ieindia.info/public.asp
[27] R L Freyman and G P Nerbonne. "The Importance of Consonant Vowel Intensity Ratio in the Intelligibility of VoicelessConsonants," Journal of Accost Speech Hear Res, vol 32, 1989.
[28] R L Freyman, G P Nerbonne and H A Cote. "Effect of C-V Ratio Modification on Amplitude Envelope Cues for Consonant Recognition," Journal of Accost Soc Am, vol 34, 1991.
[29] E Kennedy, H Levitt, Neumann and Weiss. "C-V Intensity Ratios for Maximizing Consonant Recognition by Hearing Impaired Listeners," Journal of Accost Soc Am, vol 103, no 2, 1998, p 1098.
[30] Uma Balakrishnan, Freyman, Chiang and Patrick Kelly. "Consonant Recognition for Spectrally Degraded Speech as a Function of C-V Intensity Ratio," Journal of Accostic Society of America, vol 99, no 6, 1996.
[31] AA Montgomery and R A Edge. "Evaluation of Two Speech Enhancement Techniques to Improve Intelligibility for Hearing-impaired Adults," Journal of Speech and Hearing Res, vol 31, 1988.

BIOGRAPHY OF AUTHOR

Dr N H SHOBHA obtained her Bachelor's in Electronics and Communication Engineering from Mysore University, Master's in Power Electronics from Bangalore University. She obtained her PhD from Osmania University, Hyderabad in Speech Perception/ Psychoacoustics domain. Her thesis titled 'Effects of Consonant-Vowel Intensity Ratio and Consonant-Duration Modifications on Speech Intelligibility for the Hearing-Impaired' addressed the social problem of hearing-impairment and focused on developing appropriate speech signal processing mechanisms to improve speech perception for the advantage of those individuals. She has rich experience in teaching, research and administration in Karnataka and Hyderabad. Presently, working as Professor and Head in ECE dept. at Methodist College of Engineering, Hyderabad. Her expertise includes Speech signal processing, Digital signal processing, Signals and systems, Microprocessors, Automatic control systems. She has published 14 research papers indexed in IEEE,IEI, AES .