Original Article

Effect of Spectrally Modified Speech on Speech Perception of Native Speakers of Marathi

S. B. Rathna Kumar, Prajakta Prafulla Patil[1], Udit Saxena[2], Mendem Bapuji[2], Gish Chacko[2]

Department of Audiology, Ali Yavar Jung National Institute of Speech and Hearing Disabilities, [1]Sanwad Speech and Hearing Clinic, Mumbai, Maharashtra, [2]Department of Audiology, MAA Institute of Speech and Hearing, Hyderabad, Telangana, India

Abstract

Objective: We investigated the effect of spectrally modified speech on speech perception of native speakers of Marathi. **Methods:** A total of eight word-lists with each list consisting of 25 words in Marathi were used as speech stimuli. The speech stimuli in List 1, List 2, List 3, List 4, List 5, List 6, and List 7 were spectrally modified by filtering them at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies, respectively. The speech stimuli in List 8 were unfiltered. The speech identification score (SIS) testing was performed as a measure of speech perception using eight word-lists on a total of 125 normal hearing native speakers of Marathi in the age range between 18 and 25 years. **Results:** The results revealed that there was a statistically significant effect of spectral modification on speech perception. An increase in cutoff frequency of speech stimuli was analogous to increase in speech perception performance. The normal range of SIS (90%–100%) was achieved for speech stimuli filtered at 2500 Hz cutoff frequency. **Conclusion:** Although speech is a broadband signal, the findings of this study suggest that the spectral information till 2500 Hz is sufficient for optimum perception of speech of Marathi.

Keywords: Cutoff frequency, filtered speech stimuli, Marathi, spectrally modified speech, speech-identification score

Introduction

Speech is one of the most important media of human communication system. Speech sounds are more meaningful and reflect the critical activities of life and the comprehension of social communication. Speech perception is the process by which the sounds of a language are heard, interpreted, and understood. "Speech perception" is defined as the receptive language process in which the input signal is speech. "Hearing" is one of the sensory processes that may be used in speech perception. Thus, "speech perception is a particular form of receptive language processing and hearing is a particular sensory modality contributing to speech perception".[1] Speech perception is most commonly assessed in terms of auditory capabilities of an individual to recognize the regular components of speech such as phonemes, words, and sentences.[2] In addition to regular components of speech, speech perception is dependent on how the psychophysical aspects of speech such as spectral and temporal characteristics are recognized and discriminated which indeed dependent on the auditory capabilities of an individual.[2]

There are spectral variations such as differences in formant frequencies and temporal variations such as changes in speaking rate in the way they contribute in speech perception across languages.[3] Hence, a systematic study is required to assess the contribution of spectral or temporal aspects of speech in speech perception. It might be difficult to study the contribution of spectral and temporal characteristics in speech perception simultaneously. The contribution of spectral or temporal characteristics in speech perception can be studied by altering one parameter and keeping another parameter unaltered.[3] Thus, by altering the spectral properties of speech like use of filtered speech, one can study the contribution of different spectral components in speech perception.[4] Bornstein *et al*.[4] investigated the effect of filtered speech on speech-identification performance in English. They observed that individuals with normal hearing obtained 70% correct

Submitted: 30-Dec-2019 **Revised:** 20-Mar-2020 **Accepted:** 18-May-2020
Published: 07-Dec-2021

Access this article online

Quick Response Code:

Website: www.indianjotol.org

DOI: 10.4103/indianjotol.INDIANJOTOL_142_19

Address for correspondence: Dr. S. B. Rathna Kumar, Ali Yavar Jung National Institute of Speech and Hearing Disabilities, Mumbai, Maharashtra, India.
E-mail: sarathna@yahoo.co.in

How to cite this article: Rathna Kumar SB, Patil PP, Saxena U, Bapuji M, Chacko G. Effect of spectrally modified speech on speech perception of native speakers of Marathi. Indian J Otol 2021;27:63-7.

speech-identification score (SIS) for speech stimuli filtered at 1500 Hz cutoff frequency. Similar results were reported by Avilala *et al*.[3] who observed that individuals with normal hearing obtained 70% correct SIS for speech stimuli filtered at 1200 Hz in Kannada, the only Indian language in which the effect of spectral modification on speech perception is studied till date.

The findings of the studies on one language does not hold for the other, for example, findings of investigation in the English language cannot be adapted to Indian languages and one Indian language to other Indian language because of the variations in psychophysical characteristics including the spectral energy across the languages.[3] There is a tremendous need to understand the contribution of spectral energy in different Indian languages, with the growing implication of speech perception in developing better assessment tools and management strategies, especially for individuals with auditory disorders, especially auditory processing disorders. As a first step, we investigated the effect of spectrally modified speech on speech perception among the native speakers of Marathi in the present study. The findings of this study help to understand the contribution of spectral energy in the perception of speech and determine the effective cutoff frequency of speech required for optimum perception of speech in Marathi as well.

Materials and Methods

The research design of the present study was analytical, and the sampling technique was convenient sampling. The present study was carried out in the following two phases: Phase I: development of speech stimuli (i.e., speech stimuli filtered at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies and unfiltered speech stimuli) and Phase II: A formal study for assessing SIS as a measure of speech perception using developed speech stimuli.

Phase I: Development of speech stimuli

The word lists available in the conventional speech-identification test in Marathi developed by Kumar *et al*.[5] were used to develop the speech stimuli for this study. The conventional speech-identification test in Marathi consisted of four word lists with 25 words in each list. These word lists were reported to be most familiar, commonly used and equally difficult words by native speakers of Marathi. Each word list was randomized once to form a total of eight word lists, i.e., List 1, List 2, List 3, List 4, List 5 (formed by randomization of List 1), List 6 (formed by randomization of List 2), List 7 (formed by randomization of List 3), and List 8 (formed by randomization of List 4). This was done to avoid the order and practice effect.

The recording was done in a sound-treated room, and the noise levels were maintained as per the permissible levels. Each word list was spoken by an adult female native speaker of Marathi who is a professional voice user (for recording). The recording was carried out using 16 KHz sampling rate and 16-bit quantization, and the signal was digitized at a sampling rate of 16 KHz using a 12-bit analog-to-digital converter housed within the computer. Each word list was saved as a separate file. The recorded material was then edited to carry out noise and hiss reduction. Amplitude normalization of the signals was done using the adobe audition (version 3.0) software (Adobe, San Jose, California, USA) to maintain the constant amplitude across the words. The interstimulus interval between the two words was set to 5 s. A calibration tone of 1 kHz was inserted before beginning of the word list to adjust the volume unit meter at zero. The eight word lists were then copied into an audio compact disc using a compact disc writer.

List 1, List 2, List 3, List 4, List 5, List 6, and List 7 were filtered at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies, respectively, and List 8 was unfiltered. Filtering was carried out using the adobe audition software (version 3.0) at the attenuation rate of 115 dB/octave using Butterworth filters. The filtering was done on the Intel CORE i5 laptop. A 32-bit analog-to-digital converter was selected with a sampling rate of 44.1 kHz. The recorded signal was normalized at 75% for each word list so that all the words had the same intensity.

Phase II: A formal study for assessing speech-identification score

A total of 125 normal-hearing young adults in the age range between 18 and 25 years and mean age of 23.6 years served as the participants. All the participants were the native speakers of Marathi belonging to different regions of Maharashtra (i.e., Vidarbha, Marathwada, Khandesh and Northern Maharashtra, Konkan, and Pune). It was ensured that an equal number of participants were included from each region. All the tests were conducted in an air-conditioned, double-room suite where the ambient noise levels were within the permissible limits. The audiological evaluations, including otoscopic examination, pure-tone audiometry, and tympanometry, were conducted to ensure that the participants were suitable with normal hearing for the study.

The SIS testing was carried out on each subject in a sound-treated audiometric room, and the noise levels were maintained as per the permissible levels. The stimuli were played through a CD player, which was routed through digital diagnostic clinical audiometer and delivered through the Telephonics Dynamic Headphone (TDH) 39 headphones. The SIS was calculated on each subject at most comfortable loudness level using all the eight word lists (seven filtered and one unfiltered). A short break of 15–30 min was given to them whenever they required break for the rest. All the participants were tested monaurally with eight word lists, and ear selection was done randomly. An open-set response in the form of an oral response was obtained. The SIS (%) was then calculated for each subject for each word list separately for further assessment. The formula for calculating SIS (%) is the number of words correctly repeated by the subject divided by the total number of words presented and then multiple this calculation with 100. In the present study, we considered the lowest cutoff frequency at which the participants achieved normal range of SIS as the effective cutoff frequency of speech for the optimum perception of speech in Marathi. Individuals with normal hearing are generally expected to obtain approximately 90%–100% SIS.[6-8]

Ethical clearance

This is a behavioural study in which data is collected using non-invasive methods. Further, all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and it's later amendments or comparable ethical standards.

Results

The mean and standard deviation of SIS (%) obtained by 125 participants for eight word lists, List 1, List 2, List 3, List 4, List 5, List 6, and List 7 which were filtered at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies, respectively, and List 8 which was unfiltered is depicted in Figure 1. One-way repeated measures of ANOVA showed a statistically significant difference in mean SIS (%) between and within eight word lists (F [7, 992] = 1665.968; $P < 0.001$]). Least significant difference (LSD) *post hoc* multiple comparisons were then done to check the significant difference in mean SIS (%) between each word list and are depicted in Tables 1 and 2. The results were discussed under the following headings: comparison of mean SIS between different filtered speech stimuli and comparison of mean SIS between unfiltered speech stimuli and each filtered speech stimuli.

Comparison of mean speech-identification score between different filtered speech stimuli

LSD *post hoc* multiple comparison showed a significant difference in SIS between speech stimuli filtered at 500 Hz

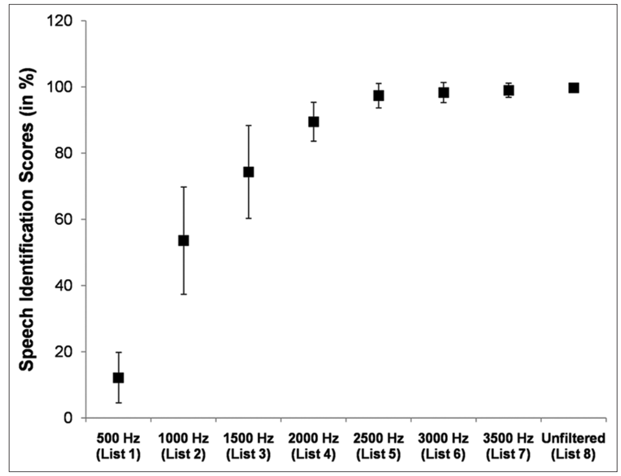

Figure 1: Mean and standard deviation of speech-identification score (%) obtained by 125 participants for eight word lists, List 1, List 2, List 3, List 4, List 5, List 6, and List 7 which were filtered at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies, respectively, and List 8 which was unfiltered

Table 1: Least significant difference *post hoc* multiple comparison of mean speech-identification score (%) with each word list (List 1, List 2, List 3, List 4, List 5, List 6, and List 7 which were filtered at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies, respectively

Speech stimuli		Mean difference (I–J)	SE	Significant	95% CI	
I	J				Lower bond	Upper bond
500 Hz	1000 Hz	−41.40800*	1.07851	0.00	−45.4623	−37.3537
	1500 Hz	−62.19200*	1.07851	0.00	−66.2463	−58.1377
	2000 Hz	−77.36000*	1.07851	0.00	−81.4143	−73.3057
	2500 Hz	−85.20000*	1.07851	0.00	−89.2543	−81.1457
	3000 Hz	−86.16000*	1.07851	0.00	−90.2143	−82.1057
	3500 Hz	−86.86400*	1.07851	0.00	−90.9183	−82.8097
1000 Hz	1500 Hz	−20.78400*	1.07851	0.00	−24.8383	−16.7297
	2000 Hz	−35.95200*	1.07851	0.00	−40.0063	−31.8977
	2500 Hz	−43.79200*	1.07851	0.00	−47.8463	−39.7377
	3000 Hz	−44.75200*	1.07851	0.00	−48.8063	−40.6977
	3500 Hz	−45.45600*	1.07851	0.00	−49.5103	−41.4017
1500 Hz	2000 Hz	−15.16800*	1.07851	0.00	−19.2223	−11.1137
	2500 Hz	−23.00800*	1.07851	0.00	−27.0623	−18.9537
	3000 Hz	−23.96800*	1.07851	0.00	−28.0223	−19.9137
	3500 Hz	−24.67200*	1.07851	0.00	−28.7263	−20.6177
2000 Hz	2500 Hz	−07.84000*	1.07851	0.00	−11.8943	−03.7857
	3000 Hz	−08.80000*	1.07851	0.00	−12.8543	−04.7457
	3500 Hz	−09.50400*	1.07851	0.00	−13.5583	−05.4497
2500 Hz	3000 Hz	−00.96000	1.07851	0.99	−05.0143	03.0943
	3500 Hz	−01.66400	1.07851	0.94	−00.7183	02.3903
3000 Hz	3500 Hz	−00.70400	1.07851	1.00	−04.7583	03.3503

*The mean difference is statistically significant at the 0.05 level. SE: Standard error, CI: Confidence interval

Table 2: Least significant difference *post hoc* multiple comparison of speech‑identification score (%) of word List 8 (unfiltered speech stimuli) with List 1, List 2, List 3, List 4, List 5, List 6, and List 7 which were filtered at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies, respectively

Speech stimuli		Mean difference (I−J)	SE	Significant	95% CI	
I	J				Lower bond	Upper bond
500 Hz	Unfiltered	−87.63200*	1.07851	0.00	−91.6863	−83.5777
1000 Hz	Unfiltered	−46.22400*	1.07851	0.00	−50.2783	−42.1697
1500 Hz	Unfiltered	−25.44000*	1.07851	0.00	−29.4943	−21.3857
2000 Hz	Unfiltered	−10.27200*	1.07851	0.00	−14.3263	−06.2177
2500 Hz	Unfiltered	−02.43200	1.07851	0.65	−06.4863	01.6223
3000 Hz	Unfiltered	−01.47200	1.07851	0.99	−05.5263	02.5823
3500 Hz	Unfiltered	−00.76800	1.07851	0.99	−04.8223	03.2863

*The mean difference is statistically significant at the 0.05 level. CI: Confidence interval, SE: Standard error

cutoff frequency and speech stimuli filtered at 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz; speech stimuli filtered at 1000 Hz cutoff frequency and speech stimuli filtered at 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies; speech stimuli filtered at 1500 Hz and speech stimuli filtered at cutoff frequencies 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz; speech stimuli filtered at 2000 Hz cutoff frequency and speech stimuli filtered at 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies. However, no significant difference SIS was found between speech stimuli filtered at 2500 Hz cutoff frequency and speech stimuli filtered at 3000 Hz cutoff frequency; speech stimuli filtered at 2500 Hz cutoff frequency and speech stimuli filtered at 3500 Hz cutoff frequency; speech stimuli filtered at 3000 Hz cutoff frequency and speech stimuli filtered at 3500 Hz cutoff frequency.

Comparison of mean speech‑identification score between unfiltered speech stimuli and each filtered speech stimuli

LSD *post hoc* multiple comparisons showed a significant difference in SIS between unfiltered speech stimuli and speech stimuli filtered at 500 Hz, 1500 Hz, and 2000 Hz cutoff frequencies. However, there was no significant difference in SIS between unfiltered speech stimuli and speech stimuli filtered at 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies. In addition, it was found that with an increase in cutoff frequency of speech stimuli, there was a corresponding increase in speech perception performance. In addition, increase in cut‑off frequency of speech stimuli lead to lower standard deviation in SIS. On the contrary, higher standard deviation in SIS was noticed when there was a decrease in cut‑off frequency of speech stimuli. The normal range of SIS (90%–100%) was found for speech stimuli filtered at 2500 Hz cutoff frequency and remained the same for speech stimuli filtered at 3000 Hz and 3500 Hz cutoff frequencies, and unfiltered speech stimuli as well.

Discussion

Speech is a broadband signal with both spectral and temporal information, which are important for the perception of speech. There are variations across languages in the way these spectral and temporal cues contribute to speech perception. The present study assessed the effect of spectrally modified speech on speech perception among the native speakers of Marathi. Spectral modification of speech stimuli was accomplished by filtering the speech stimuli at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies. The SIS testing was carried out on the native speakers of Marathi as a measure of speech perception using speech stimuli filtered at above‑mentioned cutoff frequencies and unfiltered speech stimuli as well. The lowest cutoff frequency at which the participants achieved normal range of SIS (90%–100%) were considered as the effective cutoff frequency of speech for the optimum perception of speech in Marathi.

The results revealed that there was a significant effect of spectral modification on speech perception performance. It was found that with an increase in cutoff frequency of speech stimuli, there was a corresponding increase in speech perception performance. In addition, the results also revealed a narrow standard deviation with an increase in the cutoff frequency of speech stimuli. This indicates that speech perception was less variant with an increase in cutoff frequency of speech stimuli and more variant with the decrease in cutoff frequency of speech stimuli. This is expected because, with a decrease in cutoff frequency, the bandwidth of speech signals becomes narrow, resulting in limited access or audibility to relevant phonetic cues. On the other hand, with an increase in cutoff frequency, the bandwidth of speech signals becomes broader, resulting in widespread access or audibility to relevant phonetic cues.[9-11]

The optimum perception of speech, i.e., normal range of SIS (90%–100%) was achieved for speech stimuli filtered at 2500 Hz cutoff frequency and remained the same for speech stimuli filtered at 3000 Hz, 3500 Hz cutoff frequencies and unfiltered speech stimuli as well. The results suggest that spectral information till 2500 Hz is sufficient for obtaining optimum speech perception in Marathi. Speech signals contain redundant cues that can aid in the recognition of basic components of speech such as phonemes, words, and sentences.[12] Natural speech carries abundant acoustic cues in both spectral and temporal aspects.[13-17] Limited spectral information is sufficient to understand the speech in a quiet listening environment.[18,19] Thus, reduction in spectral information of speech did not cause significant reduction in speech perception in the present study. Several investigators who have extensively evaluated the speech pattern recognition under a wide variety of conditions which distorted or reduced spectral information of speech have also reported that dramatic reduction in spectral information of speech does not cause significant reduction in speech recognition and discrimination performance.[14-16,18,20,21]

Among Indian languages, the effect of spectrally modified speech on speech-identification performance was investigated in Kannada.[3] They reported that individuals with normal hearing obtained 70% correct SIS for speech stimuli filtered at 1200 Hz cutoff frequency in Kannada. Since we have taken a criterion of 90%–100% correct SIS, a retrospective analysis was carried out to make the comparison with the study done in Kannada. The participants obtained 70% correct SIS for speech stimuli filtered at 1500 Hz cutoff frequency in Marathi as compared to a cutoff frequency of 1200 Hz in Kannada. The discrepancy in the cutoff frequency for Kannada (1200 Hz) in comparison with Marathi (1500 Hz) could be due to the predominance of low-frequency information in Kannada. The spectral energy of speech sounds in Kannada is more concentrated at the low-frequency region.[22] The findings in Marathi or Kannada cannot be generalized and adapted directly to other Indian languages as variations were evident between Marathi and Kannada. Hence, there is a necessity for similar studies in different Indian languages.

Conclusion

The present study assessed the effect of spectrally modified speech on speech-identification performance of native speakers of Marathi. Spectral modification of speech stimuli was carried out by filtering the speech stimuli at 500 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 2500 Hz, 3000 Hz, and 3500 Hz cutoff frequencies. The results revealed that there was a significant effect of spectral modification on speech-identification performance. The speech-identification performance improved as the cutoff frequency of the filtered speech stimuli increased. Although speech is a broadband signal, the findings of the present study suggest that spectral information till 2500 Hz is sufficient for obtaining a normal range of speech-identification performance in Marathi.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Marschark M, Spencer PE. The Oxford Handbook of Deaf Studies, Language and Education. New York: Oxford University Press; 2011.
2. Winn MB, Won JH, Moon IJ. Assessment of spectral and temporal resolution in cochlear implant users using psychoacoustic discrimination and speech cue categorization. Ear Hear 2016;37:e377-90.
3. Avilala VK, Prabhu P, Barman A. The effect of filtered speech on speech identification scores of young normal hearing adults. J All India Inst Speech Hear 2010;29:115-9.
4. Bornstein SP, Wilson RH, Cambron NK. Low- and high-pass filtered Northwestern University Auditory Test No. 6 for monaural and binaural evaluation. J Am Acad Audiol 1994;5:259-64.
5. Kumar SB, Mohanty P, Ujawane PA, Huzurbazar YR. Conventional speech identification test in marathi for adults. Int J Otorhinolaryngol Head Neck Surg 2016;2:205-15.
6. Gelfand SA. Essentials of Audiology. 2nd ed. New York: Thieme Medical Publishers; 2008.
7. Lawson GD, Peterson ME. Speech Audiometry. San Diego: Plural Publishing, Inc.; 2011.
8. Kramier S, Brown DK. Audiology: Science to Practice. 3rd ed. San Diego: Plural Publishing, Inc.; 2019.
9. Pittman AL, Stelmachowicz PG. Perception of voiceless fricatives by normal-hearing and hearing-impaired children and adults. J Speech Lang Hear Res 2000;43:1389-401.
10. Stelmachowicz PG, Pittman AL, Hoover BM, Lewis DE. Effect of stimulus bandwidth on the perception of/s/in normal- and hearing-impaired children and adults. J Acoust Soc Am 2001;110:2183-90.
11. Stelmachowicz PG, Pittman AL, Hoover BM, Lewis DE, Moeller MP. The importance of high-frequency audibility in the speech and language development of children with hearing loss. Arch Otolaryngol Head Neck Surg 2004;130:556-62.
12. Xu L, Pfingst BE. Spectral and temporal cues for speech recognition: Implications for auditory prostheses. Hear Res 2008;242:132-40.
13. Stevens KN. Acoustic correlates of some phonetic categories. J Acoust Soc Am 1980;68:836-42.
14. Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. Science 1981;212:947-9.
15. Rosen S. Temporal information in speech: Acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 1992;336:367-73.
16. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303-4.
17. Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. Nature 2002;416:87-90.
18. Turner CW, Souza PE, Forget LN. Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. J Acoust Soc Am 1995;97:2568-76.
19. Fu QJ, Shannon RV, Wang X. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. J Acoust Soc Am 1998;104:3586-96.
20. Van Tasell DJ, Greenfield DG, Logemann JJ, Nelson DA. Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants. J Acoust Soc Am 1992;92:1247-57.
21. Drullman R. Temporal envelope and fine structure cues for speech intelligibility. J Acoust Soc Am 1995;97:585-92.
22. Sairam VV. Long Term Average Speech Spectrum in Kannada. Unpublished Independent Project. Mysore: University of Mysore; 2002.