DOI: 10.5958/j.2319-5886.3.2.077

International Journal of Medical Research & Health Sciences

www.ijmrhs.com Volume 3 Issue 2 (April - Jun) Coden: IJMRHS Copyright @2014 ISSN: 2319-5886

Received: 10th Feb 2014 Revised: 8th Mar 2014 Accepted: 12th Mar 2014

Research Article

CONTRIBUTION OF FREQUENCY MODULATIONS TO THE PERCEPTION OF SPECTRALLY SHIFTED SPEECH IN QUIET AND NOISE

Somashekara HS[1], Nisha Krishnani[2], Jayashree S Bhat[3], *Arivudai P Nambi[4]

[1]Assistant Professor, [3]Professor, [4]Assistant Professor, Department of Audiology and Speech Language Pathology, Kasturba Medical College, Manipal University, Mangalore, Karnataka, India.
[2]Clinical Audiologist and Speech language Pathologist, Nisha Speech and Hearing centre, Ahmednagar, Maharashtra, India

*Corresponding Author email: arivudai.nambi@manipal.edu

ABSTRACT

Cochlear implant is electronic device which is surgically implanted into the cochlea to provide the sense of hearing for hearing impaired individual who may not benefit from hearing aids. The current day's cochlear implant codes only the temporal envelope cues. Speech perception simulation studies have shown that, adding frequency modulation cues to the amplitude modulation improves the speech recognition under adverse listening condition. A similar processing cannot be incorporated in the cochlear implant as it would result in spectral mismatch. The current study evaluated the effect of such mismatch on speech recognition scores. **Method:** The study involved the subjects with normal hearing who listened to the spectrally shifted HINT sentences having only amplitude modulation (AM) cues and amplitude modulation with frequency modulations (FM). Sine wave vocoders were used to synthesize the signals with only AM and AM with FM cues. For simulating the spectral mismatch, carrier frequencies were decided based on Greenwood's map. Sine wave vocoded speech was presented with and without background noise. **Results:** Paired't' test showed significant main effect of FM on spectrally shifted speech in quiet as well as noise. Mean scores significantly improved when the speech was processed with AM+FM spectral shift than AM spectral shift alone in both the conditions. **Conclusion:** The results of the current study indicated the importance of frequency modulation cues even in the spectrally mismatched conditions.

Keywords: Cochlear implant, Spectral shift, Amplitude modulations (AM), Frequency modulations (FM)

INTRODUCTION

The cues for speech recognition can be broadly classified as spectral and temporal cues. The contribution of this spectral, temporal envelope and temporal fine structure cues for speech recognition has been studied extensively in recent decades. The envelope cues from as few as four bands are sufficient for good speech recognition in quiet[1]. Increasing the number of spectral channels improves the speech perception. Adding fine structure cues along with envelope significantly improves the perception in adverse listening conditions[2]. Envelope is a slow fluctuation in the amplitude that rates below 50Hz and the fine structure is a fast frequency fluctuation that rates above 500Hz. Figure 1 depicts the envelop and fine structure of a speech signal.

Contemporary cochlear implants code only the spectral and temporal (envelope) information. The spectral information is coded by number of stimulating electrodes and electrode position in the cochlea and temporal information is coded by presenting the band specific envelope to each corresponding electrodes[3]. But in the cochlear implants the spectral information is presented to wrong place of the auditory nerve array, due to the fact that electrodes can only be inserted part of the way into the cochlea. This partial insertion causes spectral shift in the speech signal carried by the auditory nerves. The effect of this insertion depth is difficult to study in cochlear implantees due to the interaction of following variables such as duration of deafness, age of implantation, insertion depth across individuals, cognitive and linguistic performance and the amount of neuronal survival. So, efforts have been made to simulate the effect of insertion depth by shifting the spectrum towards high frequencies. There is a significant decrement in the performance when the spectrum is shifted towards the high frequencies to mimic the basal position of the electrodes in cochlear implant[4,5]

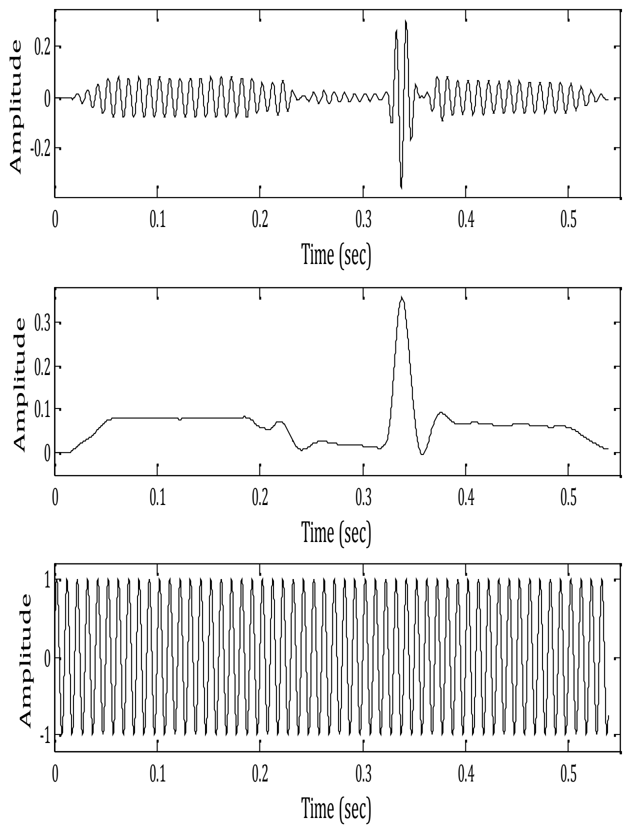

Fig 1: Representing the envelope and fine structure of the speech signal

The effect of spectral shift on speech recognition is explored in quiet and as well as in noisy situation using the speech stimuli containing only envelope cues. But, it is shown that temporal envelope cues alone are not sufficient for speech recognition in noise [1,6,7]. It has been reported that adding frequency modulation along with envelope improves the perception under adverse listening condition[8]. Adding frequency modulations will help the listeners to utilize the fine structures which is essential for speech perception in noise especially, ecologically relevant conditions such as speech perception in fluctuating noise.

TFS cues are used for speech perception in fluctuating noise is through the process of dip listening [9]. Normally hearing listeners identify target speech better in fluctuating noise than steady background noise [10,11]. This occurs because of masking release due to dip listening [12]. Dip listening refers to the ability to take a snapshot of target signal when fluctuating background noise levels momentarily decrease relatively preserving the signal. Dip listening seems to depend at least partly on TFS information. [9,11,13,14]. TFS of the target signal in the dips of the fluctuating noise is important to determine the target signal.[9] It is found that for individuals with poor TFS sensitivity, speech perception is similar in steady and fluctuating noise. This could be because they are not able to use information in dips to enhance speech perception.

Nie, Zeng & Stickney [2]proposed a method to encode the frequency modulations in the cochlear implants. The proposed method was to frequency modulate the fixed pulse width carrier with the slow varying frequency modulations extracted from the speech signal. In earlier approach only envelop cues were extracted from one carrier band and presented to wrong place of the cochlea. However this method will present both AM as well as FM cues to the wrong place of the cochlea, thus resulting in a perception of spectrally shifted speech. Inline with this, the present study compared the sentence recognition scores in AM spectrally shifted speech with AM+FM spectrally shifted speech (Spectrally shifted conditions were produced to simulate 25mm insertion depth and the active length of the electrode array was considered to be 15mm).

METHOD

Subjects: The current study is in accordance with the ethical standards of the Helsinki declaration of 1975 (revised in 1983) [15]. It followed an experimental study design with non-random convenient sampling.

A group of 39 (19 males, 20 females) normal hearing individuals within the age range of 18 to 25 years participated in the present study. They further divided into group1 and group 2. The group1 involved 24 individuals (12 males, 12 females) who were subjected to experiment 1 which assessed speech recognition abilities for the speech stimulus having amplitude modulations alone and amplitude modulations with frequency modulations in quiet, whereas in group 2 the other 15 (7 males, 8 females) individuals were assessed the same in presence of noise. All the subjects had hearing thresholds better than 15dBHL at audiometric test frequencies from 250 to 8000Hz and were exposed to English language at least for 5 years.

Stimuli: Two lists each containing 10 English sentences taken from HINT sentences[16] were used. The familiarity of the sentence lists was ascertained by individuals with exposure to English language at least for ten years. In each list three key words were identified, with a total of 30 key words. Female speaker with Indian English accent spoke the speech stimuli. The stimuli were recorded digitally on a data acquisition system at 44.1 kHz sampling frequency and using a 16-bit A/D converter in a sound treated room. Responses were scored using "loose method" in which, only the key words were considered for scoring. The correct identification of the key word received a score of one. [17]

Signal processing: Eight channel sine wave vocoders were used to simulate cochlear implant speech processing. Speech signals were band pass filtered into 8 frequency bands with a slope of 24dB/octave. From each band pass signal AM and FM was extracted in two different pathways. For AM extraction the sub-band signals were subjected to full wave rectification and low-pass filtering at 160Hz. In another pathway the sub-band signals were sent through phase orthogonal demodulation filter to extract slowly varying frequency modulation at 400Hz or at filters bandwidth. The extracted FM was used to frequency modulate the center frequencies of the shifted band. The output of the two pathways were multiplied to produce the speech signal with AM and FM. For spectral shift conditions the envelope was extracted from 80 - 8000 Hz at 24dB/octave and modulated on a frequency band of 490Hz to 5060Hz. The center frequencies were calculated based on greenwood's map[18]. Both the sentence list were processed for both the conditions i. e., spectral shift with AM alone and spectral shift with AM+FM. Signal processing for experiment1 and experiment2 was same except that, for experiment2 "Four talker babble" was added to input speech at '0dBSNR' prior to frequency amplitude modulation encoding (FAME) processing.

Procedure: The experiments were performed on a PC equipped with a Creative Labs SoundBlaster 16 soundcard. The subjects listened to the sentences via Senheiser stereo headphones at a comfortable level set by the subjects themselves. The processed stimuli lists were presented in a randomized fashion, where half of the individuals within the group listen to the AM spectrally shifted sentences alone in the list 1 and other half of the subjects listened to the list 2, which was AM+FM spectrally shifted. Subjects were instructed to listen to the sentences and write down the responses. Scoring was according to the correct key word identification through written mode. Subjects were also encouraged to write down the perceived words of the sentences, if not a complete correct sentence.

RESULTS:

The number of correctly identified key words were counted for both the conditions (AM spectrally shifted and AM+FM spectrally shifted) in quiet and noise. The maximum possible raw score for each condition was 30. Then the scores were converted into rationalized arcsine units. The rationalized arcsine transformation was considered as the inferential statistics assumes that given dependent variable on interval scale. The rationalized arcsine transformation arranges the raw scores on an interval scale and also accounts for ceiling & floor effects that are inherent in the conventional scoring method[19]. Scores obtained in both experiments were converted into rationalized arcsine units (RAU) score using following equations.

$$t = asin\left(\sqrt{\frac{x}{n+1}}\right) + asin\left(\sqrt{\frac{x+1}{n+1}}\right)$$

$$RAU = \left(\frac{146}{\pi}\right) \times t - 23$$

'x' is the raw speech recognition scores; 'n' is maximum possible score. In the current n=30

RAU scores were subjected to further statistical analysis. Shapiro Wilk's test of normality was administered to investigate whether data were normally distributed. Statistical analysis revealed that, speech recognition scores in quiet (W_{24}= 0.92, p>0.05) speech recognition scores in noise (W_{15} = 0.97, p>0.05).Paired t test was used to investigate the main effect of addition of FM cues on speech recognition in quiet as well as noise. Results revealed that, there was a significant effect of addition of FM cues on speech recognition ability in quiet (t_{14} = 5.35, p=0.00) as well as noise (t_{24} = 6.86, p=0.00). Addition of FM cues significantly improved speech recognition ability in both quiet and noise. Mean and standard deviation of speech recognition scores in quiet and in noise is represented in figure 2 & 3 respectively. However, scores represented the figure raw scores not RAU scores. Maximum possible raw score was 30.

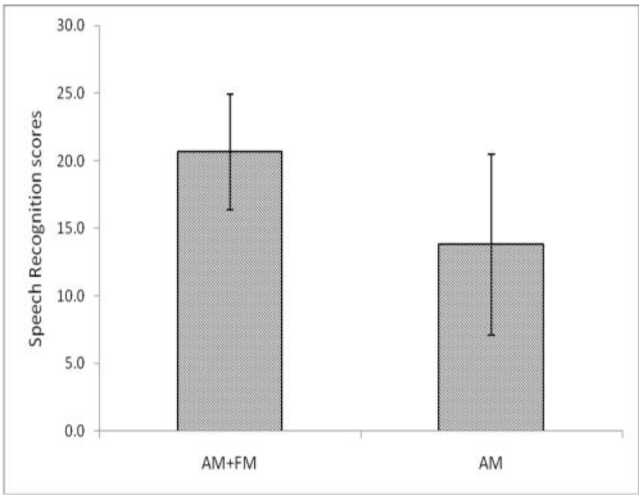

Fig 2: The mean and standard deviations of speech recognition abilities of Group1 for spectrally shifted AM (amplitude modulations) + FM (frequency modulations) and spectrally shifted AM alone in quiet.

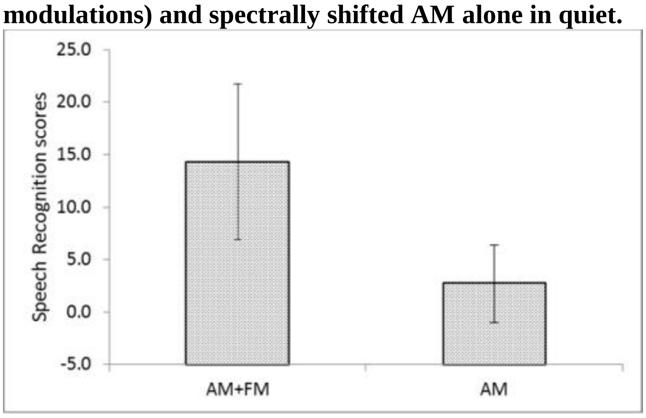

Fig 3: The mean and standard deviations of speech recognition abilities of group2 for spectrally shifted with AM (amplitude modulations) + FM (frequency modulations) and AM alone in noise.

Error bars show improved scores when the speech was processed with AM+FM spectral shift than AM spectral shift alone in both quiet and in noise. These results indicate that the AM cues alone are not sufficient for speech recognition as it was earlier reported[2,8]. FM added speech showed enhanced speech recognition ability. Thus it is observed that speech is perceived better when the FM is added to the spectrally shifted speech. However, even with FM added none of the subjects achieved 100% scores. Rosen Faulkner & Wilkinson [17]reported that speech material such as sentences and vowels requires effective transmission of spectral information for good performance when compared to consonants.

DISCUSSION

The results of the study reveal that spectrally shifted speech with both AM and FM was significantly better than spectrally shifted speech with only AM. The AM only condition has limited cues, which codes the fundamental frequency and without coding other critical cues such as, transition and formants etc. But amplitude with frequency modulation can account for this phenomenon. Adding frequency modulation to the amplitude modulation codes the transition, formant bandwidth[2,8]. Since the spectrally shifted speech with FM contained rich acoustic cues the speech perception is better with spectral shifted speech of frequency plus amplitude modulation. Rosen, Faulkner & Wilkinson, [17] reported that speech material such as sentences and vowels requires effective transmission of spectral information for good performance when compared to consonants. Even though additional cues were provided by FM, they were transmitted through different spectral channels. This could be the reason for not obtaining 100% scores even with the addition of FM.

TFS helps in speech understanding in noise is through stream segregation mechanism. When target and masker sentences occur to form the mixed signal, some distinctive envelope peaks of each are separately preserved while others combine. If only amplitude modulation is used, it impairs the subject's ability to segregate speech and noise. However, when TFS cues are used it allows the segregation of target envelope into one stream and masker envelope into another stream [2]. Poor stream segregation for harmonic complex tone by cochlear hearing loss has been earlier demonstrated by [20]. Inability to use TFS could be a possible reason for poor stream segregation ability observed in cochlear impaired individual. Poor TFS based stream segregation could contribute to the difficulties experienced by individuals with cochlear loss in understanding speech when the competing signal is also speech or music [20,21].

Over the decades the major concern was to enhance the speech intelligibility through improved signal processing in auditory prosthesis. The superior performance of the normal hearing subjects observed in earlier studies[2,8] on speech recognition when AM were supplemented with the FM suggested the importance of the FM cues in addition to AM cues. However, application of these results to the actual cochlear implant is not straight forward as they have not focused on the spectral shift and overall bandwidth compression that are inherent in actual cochlear implant. The current study accounted for these inherent properties of the cochlear implant, thereby designating the advantage of frequency modulations even under spectral shifted condition.

The addition of frequency modulations mainly indicated to be imperative in speech perception under the background noise. The FM cue lets the individuals to segregate the speech and noise to a separate perceptual stream[22]. Current study focused on the effect of spectral shift only in the quiet conditions. However it is crucial to evaluate the advantages of the frequency modulation cues in spectrally shifted carrier in the presence of background noise. Adding Frequency modulation cues codes rich temporal fine structure (TFS) which helps in speech perception in noise[9].One possible mechanism by which TFS helps in speech understanding in noise is through stream segregation mechanism. When target and masker sentences occur to form the mixed signal, some distinctive envelope peaks of each are separately preserved while others combine. If only amplitude modulation is used, it impairs the subject’s ability to segregate speech and noise. However, when TFS cues are used it allows the segregation of target envelope into one stream and masker envelope into another stream [2]. Individuals exhibit poor stream segregation if they have poor sensitivity to temporal fine structure[20]. Poor TFS sensitivity has been attributed to poor stream segregation by cochlear hearing loss individuals [20,21].

CONCLUSION

Current study compared the sentence recognition scores in normal individuals under spectrally shifted speech having only AM cues and AM as well as FM cues in quiet condition. The results indicated that, significant superior performance of the subjects when the frequency modulations were augmented with the amplitude modulations in spectrally shifted speech. Current study derives an important clinical implication that, addition of FM cues along with AM cues might improve the perception in cochlear implants even the spectral shift is present.

REFERENCES

1. Shannon R V, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995;270:303–04
2. Nie KNK, Stickney G, Zeng F-GZF-G. Encoding frequency modulation to improve cochlear implant performance in noise. IEEE Trans Biomed Eng. 2005;52:64–73.
3. Nie K, Barco A, Zeng F-G. Spectral and temporal cues in cochlear implant speech perception. Ear Hear 2006;27:208–17.
4. Dorman MF, Loizou PC, Rainey D. Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. J Acoust Soc Am. 1997;102:2993–6.
5. Shannon R V, Zeng FG, Wygonski J. Speech recognition with altered spectral distribution of envelope cues. J Acoust Soc Am. 1998;104:2467–76.
6. Faure PA, Fremouw T, Casseday JH, Covey E. Temporal masking reveals properties of sound-evoked inhibition in duration-tuned neurons of the inferior colliculus. J Neurosci. 2003;23:3052–65.
7. Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P. Temporal envelope processing in the human left and right auditory cortices. Cerebral cortex (New York, N.Y. : 1991). 2004 p. 731–40.
8. Zeng F-G, Nie K, Stickney GS, Kong Y-Y, Vongphoe M, Bhargave A, et al. Speech recognition with amplitude and frequency modulations. Proc Natl Acad Sci U S A. 2005;102:2293–8.

9. Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BCJ. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci U S A.. 2006 Dec 5 [cited 2013 Aug 30];103(49):18866–9.
10. Baer T, Moore BC. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. The Journal of the Acoustical Society of America. 1994. p. 2277–80.
11. Gilbert G, Bergeras I, Voillery D, Lorenzi C. Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues. J Acoust Soc Am. 2007;122(3):1336.
12. Nelson PB, Jin S-H, Carney AE, Nelson DA. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. J Acoust Soc Am. 2003;113:961–8.
13. Moore BBCJ. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. J Assoc Res Otolaryngol. 2008;9(4):399–406.
14. Hopkins K, Moore BCJ, Stone MA. Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. J Acoust Soc Am. 2008;123(2):1140–53.
15. Carlson R V, Boyd KM, Webb DJ. The revision of the Declaration of Helsinki: past, present and future. Br J Clin Pharmacol. 2004;57:695–713.
16. Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. J Acoust Soc Am. 1994;95(2):1085–99.
17. Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. J Acoust Soc Am. 1999;106:3629–36.
18. Greenwood DD. A cochlear frequency-position function for several species--29 years later. J Acoust Soc Am. 1990;87:2592–605.
19. Studebaker GA. A “rationalized” arcsine transform. J Speech Hear Res. 1985 Sep;28(3):455–62.
20. Grimault N, Micheyl C, Carlyon RP, Arthaud P, Collet L. Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. Br J Audiol. 2001;35:173–82.
21. Rose MM, Moore BC. Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. J Acoust Soc Am. 1997;102:1768–78.
22. Bregman AS. Auditory Scene Analysis: The Perceptual Organization of Sound. Auditory Scene Analysis The Perceptual Organization of Sound. 1990.