LANGUAGE IN INDIA

Strength for Today and Bright Hope for Tomorrow

Volume 10 : 6 June 2010

ISSN 1930-2940

Role of Encoding Temporal Fine Structure Cues in Time Compressed Word Recognition

Vidit Vidyarthi, Ritika Mittal, Imran Anwar Ali Dhamani, and S.G.R Prakash

---

Abstract

With traditional envelope cues and limited spectral and temporal information it's quite challenging for a Cochlear implant listener to understand rapid speech. A deficit in temporal processing may further contribute to difficulty understanding fast speech by a Cochlear implant listener. Hence the present study was aimed at investigating the benefit of encoding temporal fine structure information in understanding rapid speech.

32 normal hearing adults with an age range of 18 to 30 years participated in the study. Two experiments were conducted. In Experiment 1 - Words were compressed to 30% and in Experiment 2 – words were compressed to 50% of their original length. Vocoding of words was done in MATLAB 6.5. In one condition, only envelope cues were given while in the other, fine structure cues were extracted using phase orthogonal demodulation. Low pass filtering was done for getting 400 Hz modulation rate and FM bandwidth of 500 Hz with an envelope cutoff of 500 Hz. Word recognition was measured in both the conditions and subjected to further statistical analysis.

A paired t-test revealed statistically significant improvement in word recognition scores with temporal fine structure cues in both the conditions. (Exp-1: $t=6.984$, $p<0.000$ & Exp-2: $t=4.399$, $p<0.000$) with a mean difference in the scores in Exp-1 = [1.875] & Exp-2 = [1.0625].

Based on the mean difference in the scores obtained and our clinical observation in both the experiments, we can conclude that in spite of getting statistically significant improvement in the scores ($p<0.00$) with temporal fine structure information, the improvement is not clinically significant. This may be due to the adequate spectral information (22 channels) provided along with the envelope and fine structure information. Hence, with good spectral resolution there is not much effect of encoding temporal fine structure information on perception of rapid speech.

Introduction

Cochlear implants allow most patients with profound deafness to successfully communicate under optimal listening conditions, with good users being able to communicate on the telephone. Although much progress has been achieved in the design and performance of the cochlear implant systems, much remain to be done. A signal can be discomposed into its amplitude modulation and frequency modulation components. At present most of the cochlear implant signal processing devices are encoding temporal envelope (AM) cues into a restricted number of channels. These devices allows adequate speech perception in quiet environments for most of its users (Stickney et al 2005) but the utility of these temporal envelope cues is seriously limited to only optimal listening conditions (high context speech materials and quite listening environments) (Zeng et al., 2004).

Most of the recent studies showed that encoding fine structure (FM) information could improve CI listener's speech understanding in more realistic listening situations especially when background noise is one of the confounding variable. According to Zeng et. al., 2004, Frequency modulation serves as a salient cue that allows a listener to separate, and then assign appropriately, the amplitude modulation cues to form foreground and background auditory objects. Frequency modulation extraction and analysis can be used to serve as front-end processing to help solve, automatically, the segregating and binding problems in complex listening environments, such as at a cocktail party or noisy cockpit.

Nie et al., (2005) measured sentence recognition in presence of competing voice in 40 normal hearing adult subjects and showed that amplitude modulation from limited number of spectral bands is sufficient to support speech recognition in quiet listening situations but performance deteriorates in presence of competing voice and then it is important to encode frequency modulation along with traditional amplitude modulation (AM) cues to improve speech understanding.

According to Stickney et. al., (2005), FM is responsible for providing the formant transition and voice-pitch cues and is critically important when speech is severely impoverished as in Cochlear implants, In addition, it is plausible that fine-structure cues could-play a role in speech comprehension under conditions in which some of the other redundant cues are absent (e.g. Stickney, Nie & Zeng., 2002) such as might be the case of sensorineural hearing loss.

All these studies clearly indicated that fine structure encoding plays an important role in the improvement of speech perception in realistic listening situations especially in presence of noise, but realistic listening situations are not only confined to noisy backgrounds but can vary from noisy, reverberant environments to distorted speech signals in terms of either frequency or time.

When signal is distorted in temporal domain it could be either time expanded or time compressed. Time stretching can be used to compress the rate of presentation of speech without any changes in pitch, articulatory properties, or prosody of the original speech material. Time stretching preserves the pitch cues while only compressing the time domain. Artificial time compression of speech is mostly useful for the purpose of fast playback of long recordings, e-mails or voicemail messages and more commonly in making disclaimers.

Fu et. al., (2001), conducted a study to find out the effect of time compression and expansion on sentence recognition by normal hearing subjects and recipients of Nuclues-22 device. Results showed that majority of the CI listeners performed poorer in recognizing time compressed speech and also on simple temporal gap detection tasks. They concluded that the difficulties faced by the CI listener in recognizing time compressed speech are due to reduced spectral resolution and deficits in auditory temporal processing. Furthermore sensorineural hearing loss may be associated with a reduction in the ability to use fine temporal information that is coded by neural phase-locking (Buss et al., 2004).

Earlier lot of research work has been done and proved the effectiveness of encoding fine structure cues in noisy backgrounds, but still some areas are scantly explored i.e time compressed speech perception and need further investigation especially in Indian languages. Hence, present study took a step further and tried to explore the effect of encoding fine structure cues on time compressed word recognition.

Aim

The present study was aimed at investigating the effect of encoding temporal fine structure cues on time compressed word recognition.

Method

A. Subjects

A total of 32 adult (7 females and 25 males) normal hearing native Hindi speaker's with a age range of 18-30 years participated in the study. All the participants were high school pass with some pursuing their graduation while rest of them was doing their Masters degree in their respective discipline. Their pure tone audiometric threshold was within 20 dB with no indication of any middle ear infection.

B. Instrumentation

For the recruitment of subjects calibrated clinical audiometer (MAICO MA53) alongwith calibrated Immitance meter (Amplaid A756) were used to confirm normal hearing sensitivity in the test subjects and to rule out any middle ear pathology. Adobe audition 1.0 software installed on a Dell laptop (Inspiron 1420) was used for recording of words, generation of noise and further mixing of noise with the recorded words. Further processing and creation of acoustic simulation of CI were done in MATLAB-6.5. Presentation of test stimuli during the experimental tasks were done via good quality Logitech headphones connected to a Dell laptop (Inspiron 1420) with inbuilt sigmatel high definition audio card.

C. Test material

All the recording and generation of stimuli were done in Adobe Audition 1.0 software. Twenty standardized Hindi phonetically balanced words (taken from Indian speech, language and hearing tests - the ISHA battery, 1990) were recorded from male native Hindi speaker in a double wall sound attenuated booth. The words were uttered at a comfortable level and recorded with a microphone at a distance of about 12 cm from mouth. Recording was done at a sample frequency of 48 KHz with 16 bit. Words were time compressed to 30% and 50% of their original length in Adobe Audition 1.0 software.

D. Testing environment

All the testing for subject selection was performed in an air conditioned, acoustically treated room with taking care of the ambient noise level which was within the permissible limits (ANSI S3.1 1977). Experimental tasks were performed in a quiet environment.

E. Signal processing

MATLAB 6.5 software was used for the processing of the recorded stimuli. All stimuli were pre-emphasized with a first-order, high-pass Butterworth filter above 1.2 kHz. The analysis filters were fourth-order elliptic bandpass filters with 50-dB attenuation in the stop band and a 2-dB ripple in the passband. The speech signal processing bandwidth was 80 Hz - 8000Hz. Signals were processed through 22 band pass filters using 6^{th} order Butterworth filters. The length of the noise/masker signal was always kept greater than the speech signal. The filter bank parameters were kept according to the Greenwoods map (Greenwood., 1990) to mimic cochlear filters.

The AM was extracted by full-wave rectification of the output of the bandpass filter, followed by a low-pass filter in each band pass using 6^{th} order Butterworth filter. The low pass cutoff frequency for AM was kept at 500 Hz. Phase orthogonal demodulation technique was used to extract FM from each band by removing the centre frequency at each band.

The group delay between AM and FM pathways were compensated. Low-pass filters were used to limit the frequency modulation range. Instantaneous frequency was then calculated from the in-phase signal and the out-of-phase signal. The instantaneous frequency was further band-limited and low-passed to limit the frequency modulation rate at (400 Hz) and bandwidth at 500Hz, additionally, an amplitude threshold device was used to remove artificial frequency modulations exaggerated by a differential process in the algorithm.

The center frequency was re-introduced and the instantaneous frequency integrated for the recovery of the original subband phase value. In the last stage of the synthesis part, the recovered bandpassed signals were summed to form the synthesized speech containing the processed slowly varying amplitude and frequency modulations.

The AM-only stimuli were obtained by modulating the temporal envelope to the subband's center frequency and then summing the modulated subband signals. The AM+FM stimuli were obtained by additionally frequency modulating each band's center frequency before amplitude modulation and subband summation. Before the subband summation, both the AM and the AM+FM processed subbands were subjected to the same bandpass filter as the corresponding analysis bandpass filter to prevent crosstalk between bands and the introduction of additional spectral cues produced by frequency modulation.

All the frequency bands were then summed to produce the acoustic simulation of the cochlear implant and the processed stimuli was saved on a CD and played back during the experimental tasks in random order.

F. Procedure

All the subjects performed experimental tasks in a quiet environment. Each word was assigned to both of the strategies. Two experiments were conducted. In Experiment 1 - Words were compressed to 30% and in Experiment 2 – words were compressed to 50% of their original length. Over all set of stimuli comprised of 80 processed words [Experiment-1=40words {20(AM) & 20(AM+FM) and Experiment-2=40{20(AM) & 20(AM+FM)}].

In each trial, presentation of processed words was done through good quality (Logitech) headphones connected to a Dell laptop (Inspiron 1420) with an inbuilt sigma tel high definition audio card. A selection trial with unprocessed speech was offered prior to testing to screen out subjects who scored below 80%. Following this selection trial, a practice session was given to familiarize the subjects with the processed speech. No score was calculated for this practice session. During the main experiment subjects were asked to repeat the words after hearing them. The listeners were not familiar with the speaker whose speech samples were used.

Results

As shown in the table 1.0, in general with fine structure information an overall improvement of 9.37% percent was obtained in Experiment-1 (30% time compression) as compared to 5% improvement observed in Experiment 2 (50% time compression). A paired t-test was administered on the word recognition scores obtained in both the experiments to find out the significance of effect of encoding temporal fine structure cues on the word recognition scores. Results revealed statistically significant improvement in word recognition scores with temporal fine structure cues in both the experiments. While with traditional envelope encoding a mean score of 14.87 in Experiment-1 and 15.15 in Experiment-2 was obtained, temporal fine structure encoding yielded a mean score of 16.75 in Experiemnt-1 and 16.21 in Experiement-2 (Exp-1: $t=6.984$, $p<0.000$ & Exp-2: $t=4.399$, $p<0.000$) with a mean difference in the scores in Exp-1 = [1.875] & Exp-2 = [1.0625].

SUBJECTS	30% COMPRESSION SCORES			50% COMPRESSION SCORES		
	AM	AM+FM	DIFFERENCE(%)	AM	AM+FM	DIFFERENCE(%)
1	17(85%)	18(90%)	5%	16(80%)	17(85%)	5%
2	14(70%)	18(90%)	20%	18(90%)	18(90%)	0%
3	17(85%)	18(90%)	5%	15(75%)	17(85%)	10%
4	14(70%)	16(80%)	10%	17(85%)	17(85%)	0%
5	16(80%)	18(90%)	10%	16(80%)	16(80%)	0%
6	16(80%)	19(95%)	15%	20(100%)	20(100%)	0%
7	11(55%)	16(80%)	25%	13(65%)	15(75%)	10%
8	10(50%)	12(60%)	10%	12(60%)	14(70%)	10%
9	14(70%)	16(80%)	10%	17(85%)	17(85%)	0%
10	18(90%)	18(90%)	0%	18(90%)	18(90%)	0%
11	17(85%)	18(90%)	5%	17(85%)	17(85%)	0%
12	11(55%)	15(75%)	20%	14(70%)	15(70%)	0%
13	17(85%)	17(85%)	0%	17(85%)	19(95%)	5%
14	14(70%)	16(80%)	10%	11(55%)	17(85%)	30%
15	19(95%)	20(100%)	5%	18(90%)	19(95%)	5%
16	18(90%)	19(95%)	5%	18(90%)	18(90%)	0%
17	12(60%)	15(75%)	15%	9(45%)	11(55%)	10%
18	13(65%)	18(90%)	25%	14(70%)	16(80%)	10%
19	19(95%)	19(95%)	0%	16(80%)	16(80%)	0%
20	15(75%)	15(75%)	0%	12(60%)	14(70%)	10%
21	16(80%)	19(95%)	15%	17(85%)	18(90%)	5%
22	13(65%)	15(75%)	10%	15(75%)	15(75%)	0%
23	16(80%)	16(80%)	0%	14(70%)	14(70%)	0%
24	17(85%)	18(90%)	5%	18(90%)	18(90%)	0%
25	15(75%)	19(95%)	20%	17(85%)	19(95%)	10%
26	13(65%)	15(75%)	10%	12(60%)	12(60%)	0%
27	12(60%)	16(80%)	20%	16(80%)	16(80%)	0%
28	12(60%)	13(65%)	5%	14(70%)	14(70%)	0%
29	17(85%)	17(85%)	0%	13(65%)	15(75%)	10%
30	12(60%)	14(70%)	10%	13(65%)	16(80%)	15%
31	15(75%)	17(85%)	10%	14(70%)	17(85%)	15%
32	16(80%)	16(80%)	0%	14(70%)	14(70%)	0%

Table 1.0, showing the word recognition scores obtained in the two experiments.

SUBJECTS	AGE	SEX
SI	25	F
S2	23	F

S3	**25**	**M**
S4	**26**	**F**
S5	**27**	**M**
S6	**27**	**M**
S7	**23**	**M**
S8	**26**	**M**
S9	**25**	**M**
S10	**27**	**M**
S11	**22**	**M**
S12	**23**	**M**
S13	**25**	**M**
S14	**26**	**M**
S15	**24**	**M**
S16	**23**	**F**
S17	**19**	**F**
S18	**23**	**F**
S19	**19**	**M**
S20	**20**	**M**
S21	**23**	**F**
S22	**22**	**M**
S23	**21**	**M**
S24	**21**	**M**
S25	**21**	**M**

S26	**24**	**M**
S27	**31**	**M**
S28	**18**	**M**
S29	**24**	**M**
S30	**29**	**M**
S31	**27**	**M**
S32	**24**	**M**

Table 1.2 showing total number of subjects with their age and sex. Total 7 females and 25 males participated in the study.

Discussion

Recent studies have shown that temporal envelope cues with limited spectral information are sufficient for good speech perception for the current Cochlear implant listeners in quiet listening situations (Stickney et al 2005). But problem occurs, when a Cochlear implant listener has to listen in a more realistic listening situation. Earlier research work has proved that only envelope encoding is not sufficient to understand speech in noisy listening environment, and there is a need to encode additional temporal fine structure cues to improve the CI listener's performance in such kind of listening situations (Zeng et al., 2004, Stickney et al., 2004, Nie et al., 2005) but when we talk about realistic listening situation, it is not only confined to noisy backgrounds but can vary from noisy reverberant environments to distortion in the speech signal either in temporal or frequency domain.

There was a need to further assess the utility of encoding these fine structure cues on speech perception in other kinds of realistic listening situations. Hence, present study took a step forward by investigating the effect of encoding temporal fine structure cues in perception of temporally distorted speech. A technique called time stretch was used to compress the speech signals without altering the frequency component of the signals. Speech recognition was measured in two time compression ratios (30% and 50% time compression) with only temporal envelope encoding and with both temporal envelope plus fine structure encoding.

The results of the present study imply that the improvement in the word recognition scores in the two experiments with the encoding of fine structure information is statistically highly significant ($p<0.000$). But from our clinical experience, overall percentage improvement in the word recognition scores with the encoding of fine structure information (Experiment.1 = 9.37% and Experiment.2 = 5%) and the mean difference in the scores obtained in both experiments (Experiment1 = 1.875, Experiment2 = 1.0625), it is quite clear that although the improvement in the scores is statistically highly significant but not much significant from clinical point of view. As showed by Fu et. al., (2001) that understanding of time compressed speech is also dependent upon the spectral resolution and with improvement in spectral resolution, speech recognition also improves, we can contribute the clinically non significant improvement in the scores with the fine structure encoding in the present study to the good spectral resolution (22 frequency bands) provided in both the experimental conditions. Hence in future research work could be carried out to investigate the effect of encoding fine structure cues in time compressed speech recognition with varying number of frequency bands.

Conclusion

Results were quite conclusive and indicated that by providing good spectral resolution in the present day cochlear implant speech processors alongwith traditional envelope cues a good listening experience can be facilitated to its users especially when listening to rapid speech. Additional fine structure encoding may be beneficial in presence of restricted spectral resolution alongwith traditional envelope encoding, but this is still a question for future research and need further investigation to arrive at any final conclusion. Hence, in future research work could be carried out to find out the effectiveness of encoding fine structure information in perception of time compressed speech with varying number of channels.

References

Fitzgibbons, P. J and Wightman, F. L. (1982). Gap detection in normal hearing and hearing impaired listeners, *Journal of acoustical society of America, 72, 3.*

Fu Q. J., Galwin, J. J. & Wang, X. (2001). Recognition of time-distorted sentences by normal hearing and cochlear implant listeners, *Journal of acoustical society of America, 109, 1.*

Greenwood, D. D (1990). A Cochlear frequency-position function for several species-29 years later. *Journal of acoustical society of America,* 87, 6, 2592-2605.

ISHA phonetically balanced word list-3, (1990). In Kacker, S.K. & Basavraj .V, *Indian speech, language and hearing tests – the ISHA battery. The proposed ISHA battery of tests were discussed on the 8th and 9th February, 1990 in a seminar at AIIMS, New Delhi and approved, (1990),* Appendices.

Konkel, D., Beasly, D. & Bess, F. (1977). Intelligibility of time altered speech in relation to chronological ageing, *Journal of Speech Hearing Research*, 20, 108-115.

Letoweski, T. & Poch, N. E. (1995). Understanding of time-compressed speech by older adults: effect of discard interval, Journal *of American academy of Audiology*, 6, 433-439.

Nie, K. B., Stickney, G. & Zeng, F.G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise, *IEEE Transactions on Biomedical Engineering*, 52, (1), 64–73.

Stickney, G.S., Nie, K.B. & Zeng F.G., (2005). Realistic listening improved by adding fine structure. *Journal of the Acoustical society of America*, 112, 2355.

Stickney, G.S. & Zeng, F.G. (2004). Cochlear implant speech recognition with speech maskers, *Journal of acoustical society of America*, 116, 2.

Stickney, G.S., Nie, K.B. & Zeng F.G., (2005). Contribution of frequency modulation to speech recognition in noise, *Journal of acoustical society of America*, 118, 4.

Acknowledgement

We thank all the subjects for their active participation and cooperation in the study. We would also like to thank Mr. R. Rangasai (Director-AYJNIHH), for his motivation and encouragement, which led us to do this kind of research work. We are also thankful to Mr. Jitender Pathak (Lecturer Electronics, AYJNIHH-SRC) for his valuable guidance and support on technical issues, Mr. Umar Khan (founder and Director Helen Keller institute of speech and hearing, Secundrabad) and Mr. Jai Prakash (NIMH-Secunderabad) for their immense support during data collection.

Vidit Vidyarthi, M.Sc ASLP – II year
AYJNIHH – SRC
Manovikas Nagar

Old Bowenpally
Seunderabad – 500 009
Andhra Pradesh, India
vidit_vidyarthi@yahoo.co.in,
viditvidyarthi2885@gmail.com

Ritika Mittal, M.A. SLP - II Year,
Manipal College of Allied Health Sciences,
Manipal University
Manipal 576 104
Karnataka, India
riti_mittal2885@yahoo.co.in,
ritimittal2885@gmail.com

Imran Anwar Ali Dhamani
Department of Audiology
Manipal College of Allied Health Sciences
Manipal University
Manipal, 576 104
Karnataka, India.
imrandhamani@yahoo.co.in.

S. G. R. Prakash, Ph.D.
AYJNIHH - SRC
Secunderabad, 500 009
Andhra Pradesh, India.
prakash_nihh@rediffmail.com