On a combination of amplitude and frequency modulation used for processing speech signals in cochlear implants

Svetlin Antonov[1] and Snejana Pleshkova-Bekiarska[2]

Abstract – **A research, which combines the measurement of both amplitude and frequency modulation of speech signals and their processing in the processing unit of the cochlear implant, is being proposed. Numeric simulation is used as the basis for a comparison between the usage of the aforementioned combination of both modulations and the usage of only amplitude modulation. Using the proposed algorithm, a comparison between the original and processed signals is drawn.**

Keywords – **cochlear implants, amplitude and frequency modulation, speech processing.**

I. Introduction

Acoustic characteristics in speech signals allow listeners to derive not only the meaning of the speech but also the speaker's identity and emotion. Previous studies using either naturally produced whispered speech [1] or artificially synthesized speech [2], [3] have isolated and identified several important acoustic cues for speech recognition. For example, computers relying on primarily spectral cues and human cochlear-implant listeners relying on primarily temporal cues can achieve a high level of speech recognition in a quiet environment [4]- [6].

The goal of this study is to verify the relative contributions of spectral and temporal cues to speech recognition in realistic listening situations. A speech signal produced by a male talker is chosen for the purpose. We propose a combination of slowly varying amplitude modulation (AM) and frequency modulation (FM) from a number of frequency bands in speech signals and testing their relative contributions to speech recognition in acoustic and electric hearing. Different from previous studies using relatively ''fast'' FM to track formant changes in speech production [8], [11], or fine structure in speech acoustics [9], [10], the ''slow'' FM used here tracks gradual changes around a fixed frequency in the subband. We evaluate the AM-only, AM plus FM, and the original unprocessed speech signal to compare these 3 situations, and to extract the MSE and the distortion.

II. Methods

We conducted an experiment to test this hypothesis about the relative contribution of the added frequency modulation in the speech signal processing method in the cochlear implants.

In this experiment the processed stimuli contain either the AM cue alone or both the AM and FM cues. The main parameter is the number of frequency bands varying from 1 to 34.

We use a speech signal produced by a male talker (1,5s.). We conducted an experiment to test this hypothesis. The stimuli used are processed to contain either the AM cue alone or both the AM and FM cues. The main parameter is the number of frequency bands varying from 1 to 34. Different from previous studies, this experiment found that four AM bands were not enough to support good speech performance.

Thirty-four bands were used to match the number of auditory filters estimated psychophysically over the 80- to 8,800-Hz bandwidth [12].

Fig. 1 shows the block diagram for stimulus processing. To produce the AM-only and AM plus FM stimuli, a stimulus was first filtered into a number of frequency analysis bands ranging from 1 to 34. The distribution of the cutoff frequencies of the bandpass filters was approximately logarithmic according to the Greenwood map [13]. The band-limited signal was then decomposed by the Hilbert transform into a slowly varying temporal envelope and a relatively fast-varying fine structure [12], [14], [15]. The slowly varying FM component was derived by removing the center frequency from the instantaneous frequency of the Hilbert fine structure and additionally by limiting the FM rate to 400 Hz and the FM depth to 500 Hz, or the filter's bandwidth, whichever was less [16].

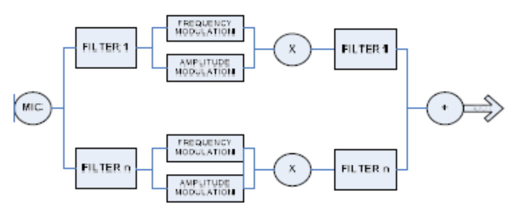

Fig. 1. Signal processing block diagram. The input signal is first filtered into a number of bands, and the band-limited AM and FM cues are then extracted. In the AM-only condition, the AM is modulated by either a noise or a sinusoid whose frequency is the bandpass filter's center frequency (not shown). In the AM_FM condition, the FM is smoothed in terms of both rate and depth and then modulated by the AM. In either condition, the same bandpass filter as in the analysis filter is applied before summation to control spectral overlap and resolution.

The AM-only stimuli were obtained by modulating the temporal envelope to the subband's center frequency and then

[1]Svetlin Antonov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: svantonov@yahoo.com.

[2]Snejana Pleshkova-Bekiarska is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: snegpl@tu-sofia.bg.

y_i – value of the i^{th} sample in y.

IV. Discussion

Because the FM cue is derived from phase, the present study argues strongly for the importance of phase information in realistic listening situations. We note that for at least two decades phase has been suggested to play a critical role in human perception [17], yet it has received little attention in the auditory field.

The most direct and immediate implication is to improve signal processing in auditory prostheses. Currently, cochlear implants typically have 12–22 physical electrodes, but a much smaller number of functional channels as measured by speech performance in a quiet environment [18]. The results of our research strongly suggest that frequency modulation in addition to amplitude modulation should be extracted and encoded to improve cochlear implant performance. Recent perceptual tests have shown that cochlear implant subjects are capable of detecting these slowly varying frequency modulations by electric stimulation [19].

Acknowledgement

Reserch in the subject has recently commenced through a NIS TU - Sofia funded project, № 121 ПД 0063-07.

References

[1] Tartter, V. C., "Percept. Psychophys" 49, 365–372, 1991

[2] Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K. & Rabinowitz, W. M.,Nature 352, 236–238, 1991

[3] Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. , Science 270, 303–304, 1995

[4] Remez, R. E., Rubin, P. E., Pisoni, D. B. & Carrell, T. D., Science 212, 947–949, 1981

[5] Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C. & Holden, T. A.,Ear Hear. 23, 207–223, 2002

[6] Rabiner, L. (2003) Science 301, 1494–1495.

[7] Dorman, M. F., Loizou, P. C. & Rainey, D., J. Acoust Soc. Am. 102, 2403–2411, 1997

[8] Alexandros, P. & Maragos, P., Speech Commun. 28, 195–209, 1999

[9] Smith, Z. M., Delgutte, B. & Oxenham, A. J., Nature 416, 87–90, 2002

[10] Sheft, S. & Yost, W. A., "Air Force Research Laboratory Progress Report No. 1", Contract SPO700-98-D-4002 (Loyola University, Chicago), 2001

[11] Doupe, A. J. & Kuhl, P. K., Annu. Rev. Neurosci. 22, 567–631, 1999

[12] Moore, B. C. & Glasberg, B. R., Hear. Res. 28, 209–225, 1987

[13] Greenwood, D. D., J. Acoust. Soc. Am. 87, 2592–2605, 1990

[14] Flanagan, J. L. & Golden, R. M., Bell Syst. Tech. J. 45, 1493–1509, 1966

[15] Flanagan, J. L., J. Acoust. Soc. Am. 68, 412–419, 1980

[16] Nie, K., Stickney, G. & Zeng, F. G., IEEE Trans. Biomed. Eng. 52, 64–73, 2005

[17] Oppenheim, A. V. & Lim, J. S., Proc. IEEE 69, 529–541, 1981

[18] Fishman, K. E., Shannon, R. V. & Slattery, W. H., J. Speech Lang. Hear. Res. 40, 1201–1215, 1997

[19] Chen, H. & Zeng, F. G., J. Acoust. Soc. Am. 116, 2269–2277, 2004

[20] G. Casella and E.L. Lehmann, Theory of Point Estimation. New York: Springer-Verlag, 1999