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].

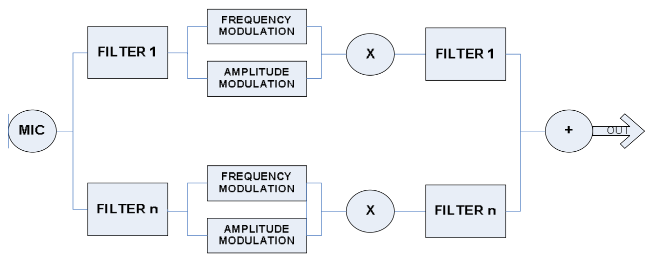

Fig. 1. Signal processing block diagram. The input signal is first filtered into a number of bands, and the band-limited AM and FM cues are then extracted. In the AM-only condition, the AM is modulated by either a noise or a sinusoid whose frequency is the bandpass filter's center frequency (not shown). 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.

summing the modulated subband signals [2], [7]. The AM plus 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 plus 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 stimuli were presented at an average root-mean-square level of 65 dB (A weighted) with the exception of the SRT measure in Exp. 3, in which the noise was presented at 55 dBA and the signal level was varied adaptively.

A signal, $s(t)$, can be approximated by a sum of N band-limited components, $x(t)$, containing both amplitude and frequency modulations

$$s(t) \approx \sum_{k=1}^{N} x_k(t) = \sum_{k=1}^{N} A_k(t)\cos\left[2\pi f_{ck} t + 2\pi \int_0^t g_k(\tau)\, d\tau + \theta_k\right] \tag{1}$$

Where $A_k(t)$ and $g_k(t)$ are the k-th band's amplitude and frequency modulations, whereas f_{ck} and θ_k are the k-th band's center frequency and initial phase, respectively.

Fig. 2 shows the block diagram for extraction of AM in the k-th subband. The AM is extracted by full-wave rectification of the output of the bandpass filter, followed by a low-pass filter LPF 1. The cutoff frequency of LPF 1 controls the maximal AM rate preserved in the AM signal. Additionally,

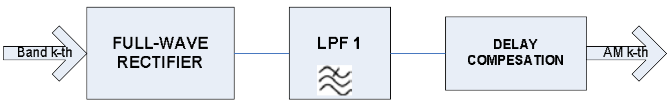

Fig. 2. Amplitude modulation block diagram.

the delay compensation box synchronizes signals between the AM and FM pathways.

Fig. 3 shows the block diagram for FM extraction in the k-th subband. First, the output of the k-th subband, $x_k(t)$, is subjected to a quadrature oscillator with the center frequency. This manipulation is equivalent to shifting the spectrum of $x(k)$ from f_{ck} to zero and $2f_{ck}$ in the frequency domain. The following low-pass filters (LPF 2 and LPF 2') then extract the slowly varying frequency components (a and b) by removing the high frequency component $2f_{ck}$. In signal processing nomenclature, the slowing-varying components and are termed in-phase and out-of-phase signals of the original subband signal $x_k(t)$, respectively.

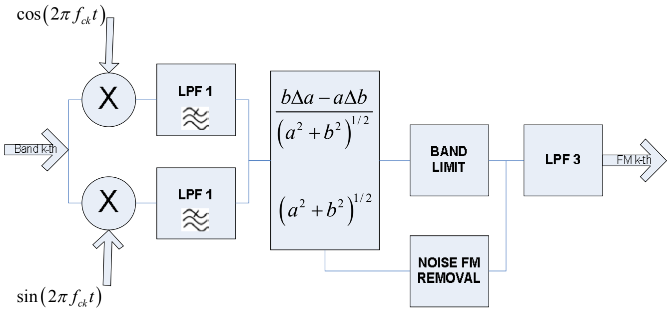

Fig. 3.Frequency modulation block diagram.

Mathematically, if $x_k(t)$ can be described as $x_k(t) = m(t)\cos\left[2\pi f_{ck} t + \varphi(t)\right]$, where $m(t)$ is the amplitude, is the center frequency and is the phase, then the in-phase signal can be derived

$$\begin{aligned}
& x_k(t) \times \cos(2\pi f_{ck} t) = \\
& m(t)\cos\left[2\pi f_{ck} t + \varphi(t)\right]\cos(2\pi f_{ck} t) = \\
& \frac{1}{2} m(t)\cos\left[2\pi f_{ck} t + 2\pi f_{ck} t + \varphi(t)\right] + \\
& + \frac{1}{2} m(t)\cos\left[2\pi f_{ck} t + \varphi(t) - 2\pi f_{ck} t\right] = \\
& \frac{1}{2} m(t)\cos\left[2(2\pi f_{ck})t + \varphi(t)\right] + \frac{1}{2} m(t)\cos\varphi(t)
\end{aligned} \tag{2}$$

Again, the first term in the above equation can be filtered out

$$b = -\frac{1}{2} m(t)\sin\varphi(t) = \frac{1}{2} m(t)\cos\left[\varphi(t) + \frac{\pi}{2}\right] \tag{3}$$

Dividing b by a will produce

$$\begin{aligned}
& \frac{b}{a} = -\tan\varphi(t) \\
& \varphi(t) = \tan^{-1}\left(-\frac{a}{b}\right)
\end{aligned} \tag{4}$$

Finally, the instantaneous frequency can be obtained

$$\begin{aligned}
& FM = \frac{1}{2\pi}\frac{d\varphi(t)}{dt} = \\
& \frac{d\tan^{-1}\left(-\frac{b}{a}\right)}{2\pi dt} = \\
& \frac{b\left(\frac{da}{dt}\right) - a\left(\frac{db}{dt}\right)}{2\pi\left(a^2+b^2\right)}
\end{aligned} \tag{5}$$

In discrete implementation, differentiation in Eq. (5) can be substituted by calculating the difference in time (Δ) to obtain the slowly varying frequency modulation

$$FM = \frac{b\Delta a - a\Delta b}{2\pi\left(a^2 + b^2\right) \times T_s} \qquad (6)$$

where T_s represents sampling period.

III. RESULTS

Fig. 4 shows the spectrograms of the original and processed speech sound: on the top - original test speech signal, on the left - the 1-, 8-, and 32-band amplitude modulation only, whereas on the right- the 1-, 8-, and 32-band amplitude and frequency modulation conditions

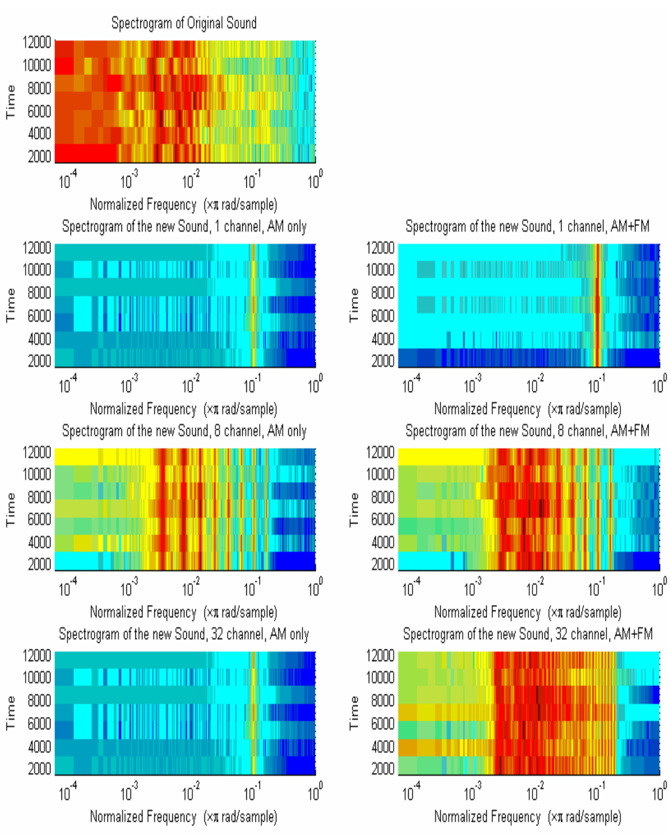

Fig. 4. Spectrograms of the original and processed speech sound. On the top- original speech signal. On left- the 1-, 8-, and 32-band amplitude modulation only. On right- the 1-, 8-, and 32-band amplitude and frequency modulation conditions.

First, we note that the original formant transition is not represented in the AM-only speech with few spectral bands (1 and 8 spectral bands left side), and only crudely represented with 32 bands (left side). In contrast, with as few as 8 bands, the AM plus FM speech (8 spectral bands right side) preserves the original formant transition. Second, we note that the decreasing fundamental frequency in the original speech is represented with even the 1-band AM plus FM speech (1 spectral band right side) but not in any AM-processed speech. The acoustic analysis result indicates that the present slowly varying FM signal preserves dynamic information regarding formant and fundamental frequency movements.

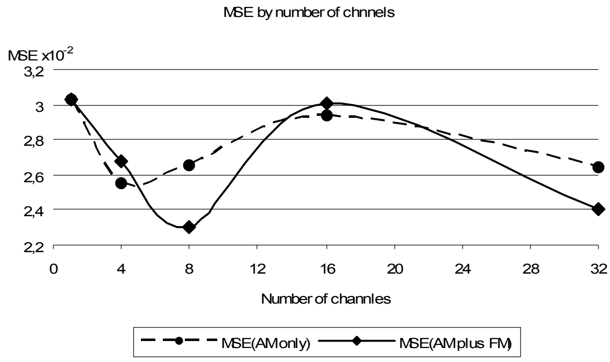

Fig. 5. Diagram showing the results of computing the mean squared error (MSE) depending of the number of the bandpass channels while AM only or AM plus FM condition.

Fig. 5 shows the diagram of the extracted mean squared error (MSE) depending on the number of the channels and the modulation conditions. We can see the lower value of MSE in 8 channels and AM plus FM conditions. We can see again that the MSE is lower in AM plus FM in 32 channels than AM only condition.

TABLE I
MSE DEPENDING ON PROCESSING CONDITIONS

Number of bandpass channels	MSE (AM only condition)	MSE (AM + FM condition)
1	0,03033781	0,03033854
4	0,02550918	0,02673373
8	0,02660029	0,02304768
16	0,02942653	0,03009370
32	0,02648830	0,02400135

Table I shows the exactly the values of MSE which are the base of Fig.5.

MSE is essentially a signal fidelity measure [20]. The goal of a signal fidelity measure is to compare two signals by providing a quantitative score that describes the degree of similarity/fidelity or, conversely, the level of error/distortion between them. Usually, it is assumed that one of the signals is a pristine original, while the other is distorted or contaminated by errors.

Suppose that $x = \left\{x_i \mid i = 1, 2, ..., N\right\}$ and $y = \left\{y_i \mid i = 1, 2, ..., N\right\}$ are two finite-length, discrete signals, original and processed. The MSE between the signals is given by the following Eq. (7).

$$MSE(x, y) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - y_i\right)^2 \qquad (7)$$

Where,
N – number of signal samples,
x_i – value of the i^{th} sample in x,

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