FLEXIBLE COCHLEAR SYSTEM BASED ON DIGITAL MODEL OF COCHLEA: STRUCTURE, ALGORITHMS AND TESTING*

Jaroslaw Baszun, Alexander Petrovsky
Dept. of Computer Science, Bialystok University of Technology,
Wiejska St. 45A, 15-351 Bialystok, POLAND
Tel: +48-85 7428206; fax: +48-85 7422393
e-mail: jb@ii.pb.bialystok.pl

ABSTRACT

This paper presents complete speech processing system for cochlear prosthesis. The main elements of the system are speech enhancement system, speech processor and system for speech reconstruction. Speech enhancement system exploits properties of modulation spectrum of the human speech. Speech signal is split into 64 equal bands using maximally decimated polyphase DFT filter bank. Power spectrum of the signal for each channel is computed and filtered with tunable bandpass filter. Characteristic of the bandpass filter is computed independently for each channel base on the properties of power spectrum of the signal in the channel. Output signal is reconstructed from filtered power spectrum and original phase of the input signal. Tests proved that the system is suitable for cochlear implants and as a preprocessing procedure in Automatic Speech Recognition (ASR) systems. Speech processor dedicated for CIS stimulation was design base on digital, two dimensional, nonlinear model of the human cochlea model. The analysis filter bank was built with use tunable bandpass filters.

1 INTRODUCTION

Cochlear implants have become one of the most successful prosthesis for profoundly deaf people. The implant consists of two parts: external and internal which is implanted for a patient. The most important element of the external part is a speech processor. The main task of one is to select some acoustic features from microphone signal important for understanding speech. Existing speech processors are usually built base on monolithic filterbank chips [1] or digital signal processors (DSP) [2] and uses digital bandpass filters or Fast Fourier Transform (FFT) algorithm for frequency analyze of speech signal. In some cases specialized digital chips are used dedicated for specific algorithms. Because of the limited number of electrodes and relatively to human, simple speech processing method used in cochlear implants people using that kind of electric hearing devices have problems to understand speech in adverse conditions such as reverberation and/or additive noise. The degradation of speech intelligibility is much bigger than in case of normal binaural hearing people. Therefore, speech enhancement techniques should be applied [3].

In our approach sound processor for cochlear implants was design with use previously developed nonlinear model of human cochlea [4]. Base on it system for design banks of tunable bandpass filters was developed [5]. The cochlear system consists of the following elements: speech enhancement system, speech processor for cochlear implants, system for speech reconstruction for test and tuning processes and control module for float point DSP processor and fixed point RISC stack processor. One of the advantages of the presented solution is one channel, environment independent speech enhancement system.

2 SYSTEM STRUCTURE

The main blocks of the system are shown in Fig. 1. Signal from microphone is lowpass filtered and converted into digital form. Next speech enhancement procedure is applied as an option. After enhancement signal is split into eight channels by the speech processor. Signals from the speech processor are used for speech reconstruction or/and after dynamic compression are used for control of electrodes in the implant.

2.1 Speech Enhancement

The speech enhancement system applied here exploits modulation spectrum properties of speech. It is known that reverberation has a low-pass effect on the temporal subband energy contours [6]. The idea of RelAtive SpecTrA (RASTA) [7] processing is to inverse high-pass filtering of subband energy signals. The block diagram of the speech enhancement system is shown in Fig. 2. Unlike classic RASTA method [8] tunable bandpass filters were apply to modify modulation spectrum of speech. In effect system can adapt to different kind of acoustical environments and cause less distortion in compare to solution with fixed filters [9].

*This work was supported by Bialystok University of Technology (Poland) under grant W/II/2/00

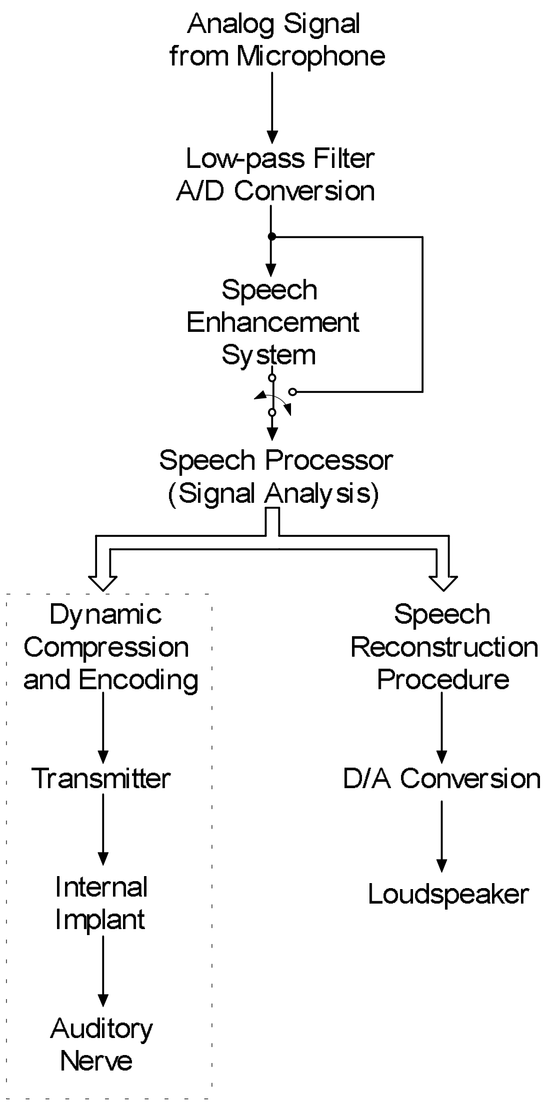

Figure 1: Signal processing steps

The transfer function of a tunable bandpass second-order digital filter [5] is defined by

$$H(z) = a_0 \frac{1 - z^{-2}}{1 + (a_0 - 1)gz^{-1} + (1 - 2a_0)z^{-2}}, \quad (1)$$

where the coefficients a_0 and g depend only on the bandwidth $\Delta\omega$ and central frequency ω_0 respectively:

$$\begin{aligned} a_0 &= \frac{\Delta\omega\Delta t}{2 + \Delta\omega\Delta t}, \\ g &= 2\cos\omega_0\Delta t. \end{aligned} \quad (2)$$

Speech signal is split into 64 equal bands (see Fig. 2) using maximally decimated polyphase DFT filter bank [10]. Polyphase network of modulated bandpass filters have been derived from the 640-th order FIR prototype lowpass filter denoted as $H(e^{j\omega}), and$ Discrete Fourier Transform (DFT). The frequency response of this uniform bank is given by equation:

$$H_m\left(e^{j\omega}\right) = H\left(e^{j(\omega - 2\pi m/M)}\right), \quad m = 0, ..., M-1, \quad (3)$$

where $H_m\left(e^{j\omega}\right)$ is m-th complex bandpass filter with centre frequencies $\omega_m = 2\pi m/M$.

Power spectrums of signal for each channel is computed and nonlinear transform T is applied:

$$y = \ln(1 + 1000 * x) \quad (4)$$

where x is the magnitude of the signal in the channel. Next the signal (power) is split into two paths. In upper path signal with modulation frequency higher than 16Hz is detected and compared to the power of the signal in the channel. The result of the comparison is transformed into a_0 coefficient of tunable bandpass filter in the bottom path. When more high frequencies appears in modulation spectrum of input signal then pass band of the tunable filter narrows and influence on signal becomes stronger (see Fig. 3). The idea of this procedure comes from that in human speech most power in modulation spectrum is concentrated between 1 and 10Hz [8]. In effect influence of the system on processing speech depends on level of interference. Clear speech is almost not affected by the system.

Signal in bottom path after filtration with tunable, second order, bandpass filter is rectified (half-wave rectifier) to eliminate samples with negative values and back transformed to linear scale (T^{-1}) as follows:

$$y = \frac{e^y - 1}{1000}. \quad (5)$$

Characteristic of the tunable bandpass filter is computed independently for each channel from the power spectrum of the signal in the channel. Its characteristic can change in the range shown in figure 5.

Output signal is reconstructed from filtered power spectrum and original phase of the input signal. Because input signal is real, the magnitude of signals after polyphase analysis filter for channels 34 to 64 are the mirror of channels 2 to 32, so processing is only necessary in the first 33 channels. The number of channels was selected experimentally.

2.2 Speech processor

After enhancement, speech signal is passed to the speech processor where signal is split in a bank of tunable bandpass filters (see Fig. 4). Each filter consists of two identical second-order sections which coefficients are shown in Tab. 1. This processor is designated for CIS stimulation strategy [11]. Signals from bandpass filters are full-wave rectified and low pass filtered with a 200 Hz cutoff frequency (second-order Butterworth). The envelope of each frequency band can be used for signal reconstruction to test speech processor or after compression

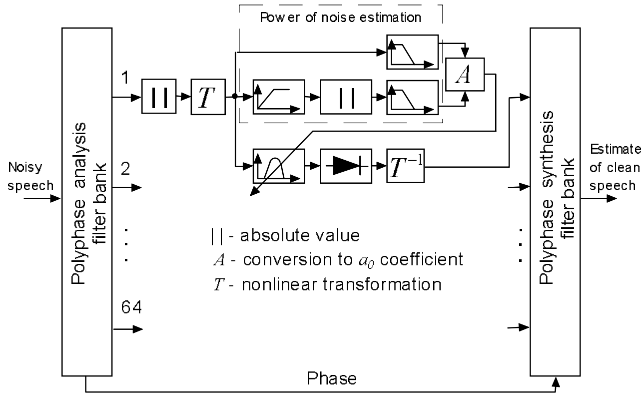

Figure 2: Speech enhancement system block diagram

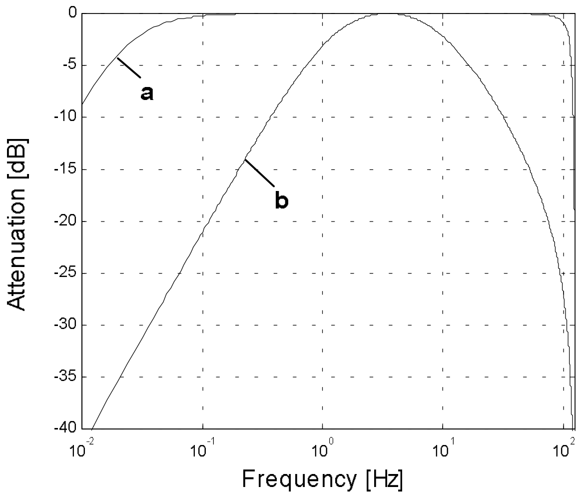

Figure 3: Maximal (a) and minimal (b) pass band of tunable filters in speech enchancement procedure

to control electrodes in the implant in case of real time implementation on the experimental device (see Fig. 1). Tuning properties of the analysis filter bank simplify adjusting of frequency characteristics of processor in individual cases.

2.3 Speech reconstruction

In order to verify the correctness of the designing processor and for experimental purposes testing procedure of the system was implementing [12]. This procedure allows to reconstruct speech signal after its passing through speech processor. For speech reconstruction eight sinusoidal generators were used. Frequency of each generator is equal to central frequency of appropriate bandpass filter in speech processor. The signal from speech processor is divided into 2ms frames. For each frame root-mean-square (RMS) energy is computed and used to set amplitude of the generators. Concurrently

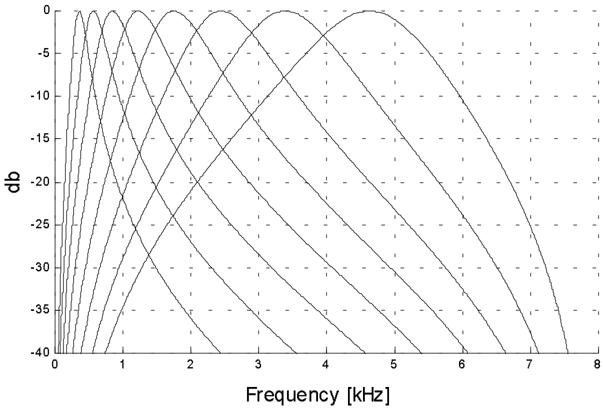

Figure 4: Magnitude response of the cochlear filter bank. Distance between electrodes is 2.27 mm. Sampling frequency is 16 kHz

	f_0	Δf	a_0	g
1	335	132	0.0482	1.9806
2	559	213	0.0756	1.9521
3	840	312	0.1074	1.8920
4	1226	431	0.1425	1.7727
5	1748	570	0.1800	1.5470
6	2450	727	0.2189	1.1434
7	3389	904	02583	0.4851
8	4638	1100	02976	-0.4962

Table 1: Parameters of the digital cochlear filters from Fig. 4

discrete Fourier transform (FFT) is used to compute phase of the original signal for frequencies equal to central frequencies of the bandpass filters in the processor. The sinusoids from each generator are finally summed, low pass filtered and presented to listeners.

3 SYSTEM TESTING

The speech enhancement system was tested both on noisy and reverberant signals. An example with noisy speech (10 dB SNR) is illustrated in Fig. 5. Intelligibility test was carried out for clean speech and noisy speech with -5, 0, 5, 10, 20 dB SNR. Speaker independent speech recognition system was used. In each case approximately 30% improvement in recognition was observed.

The speech processor was tested using speech reconstruction procedure. Special tests for vowels, consonants and sentences recognition was carried out to assess quality of designed processor. Experiments show that optimal number of channels for this system is eight. Further increasing number of channels do not improve intelligibility of reconstructed speech. High differences in

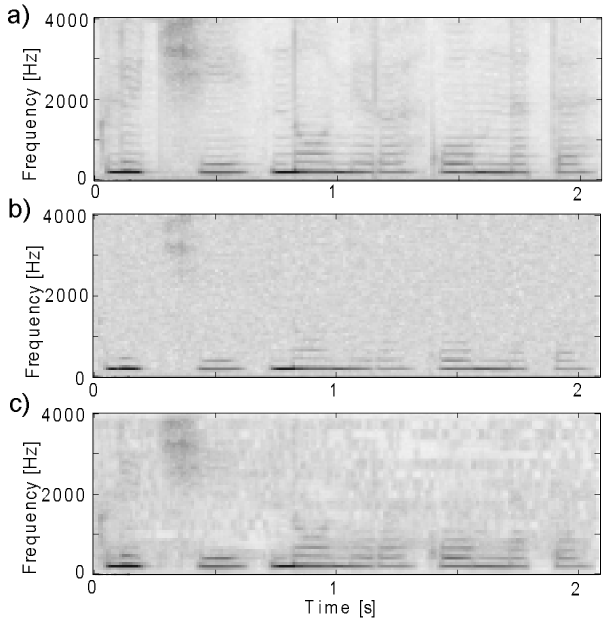

Figure 5: Spectrogram of a) the clean speech; b) speech with noise at an input SNR of 10 dB; c) results after processing with the system from Fig. 2

recognition rate between the listeners was observed in the experiment.

The influence of quantization effects on tunable filters in case of fixed point implementation was derived for bank of cochlear filters and adaptive speech enhancement procedure. Computations show that 16 bit representation of tunable filters coefficients ensure more than 70 dB dynamic range of the filters what is sufficient for presented application.

4 Conclusion

Complete speech processing system for develope cochlear prothesis was presented. Early tests shown that bank of filters designed on the basis of the created model is fit for design efficient speech processors. Their tuning property simplify adjusting of frequency characteristics of processor in individual cases. Test procedures allows to preliminary assess performance of the processor. Procedure of enhancement of noisy and reverberant speech allows to operate the system in adverse acoustical conditions. The main adventure of presented technique is lack of speech pause detection. This technique do not require reference channel and in contrast to classical RASTA processing of speech can adapt to different acoustical environments. Low distortions of speech make this technique useful for mobile communication systems.

References

[1] H. J. McDermott, C. M. McKay, and A. E. Vandali, "A new portable sound processor for the university of Melbourne/Nucleus multielectrode cochlear implant", *J. Acoust. Soc. Am.*, vol. 91, no. 6, pp. 3367–3371, 1992.

[2] H. Bogli and N. Dillier, "Digital speech processor for the nucleus 22-channel cochlear implant", in *Ann. Int. Conf. Of the IEEE Eng. In Med. and Biol. Soc.*, USA, 1991, vol. 13, pp. 1901–1902.

[3] H. G. Hirsh, "Intelligibility impovement of noisy speech for people with cochlear implants", in *Proceedings of the ESCA Workshop, "Speech Processing in Adverse Conditions"*, Cannes, Frankreich, 1992, pp. 98–103.

[4] W. Wan. and A. A. Petrovsky, "A two-dimensional nonlinear cochlear model for speech processing:response to pure tones", in *Proceedings, 6th International Fase - Congress*, Zurich, Switzerland, 29-31 July 1992, pp. 233–236.

[5] A. Petrovsky, "The synthesis of high order digital bandpass filters with tunable center frequency and bandwidth", in *VIII Europ. Sig. Proc. Conf. (EUSIPCO 96)*, Trieste, Italy, 10-13 September 1996, pp. 1527–1530.

[6] T. Houtgast, H.J.M. Steeneken, and R. Plomp, "Predicting speech intelligibility in rooms from the modulation transfer function", *Acoustica*, vol. 46, pp. 60–72, 1980.

[7] H. Hermansky, N. Morgan, and H. G. Hirsch, "Recognition of speech in additive and convolutional noise based on RASTA spectral processing", in *Proceedings of ICASSP-93*, 1993, pp. II–83–86.

[8] H. Hermansky and N. Morgan, "RASTA processing of speech", *IEEE Transaction on Speech and Audio Processing*, vol. 2, no. 4, pp. 587–589, October 1994.

[9] J. Baszun and A. Petrovsky, "Speech processor applicable to hearing prosthesis in noise and reverberant environment", in *8th International Symposium on Sound Engineering and Mastering, ISSEM'99*, Gdansk, Poland, September 09-11 1999, pp. 37–42.

[10] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, chapter 4, pp. 125–134, Prentiece Hall, 1993.

[11] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implant", *Nature*, vol. 352, pp. 236–238, 18 July 1991.

[12] J. Baszun and A. Petrovsky, "Design of speech processor for cochlear implants and method of its testing", in *Proc. of the 1st Russian Int. Conf. Digital Signal Processing and its Applications*, Moscow, Russia, June 30 - July 3 1998, vol. 1, pp. 181–189.