SPEECH ENHANCEMENT SYSTEM FOR HANDS-FREE TELEPHONE BASED ON THE PSYCHOACOUSTICALLY MOTIVATED FILTER BANK WITH ALLPASS FREQUENCY TRANSFORMATION [#]

Krzysztof Bielawski [1] *Alexander A. Petrovsky* [1,2]
[1]Department of Computer Science, Bialystok University of Technology
Wiejska 45A, 15-351 Bialystok, Poland, e-mail: biel@ii.pb.bialystok.pl
[2]Belarusian State University of Informatics and Radioelectronics
6, P. Brovky Str., 220027, Minsk, Republic of Belarus, e-mail: palex@bsuir.by

ABSTRACT

In this paper the application of multirate acoustic echo canceller derived from cochlea model and psychoacustically motivated noise and residual echo attenuation system operating on the signal decomposed in cochlear spaced subbands are described. Polyphase realisation of the FIR filter bank with non-uniform frequency resolution achieved by the frequency transformation of filter characteristic using recursive allpasses is presented as a starting point to construct of the analysis and synthesis filter bank used in proposed speech enhancement system. Human auditory perception model delivered by the psychoacoustic as well cochlear model with discrete time and discrete space are used to create this allpass transformation filter bank. Performing the acoustic echo cancellation by adaptive filtering in subbands we get echo compensated signal, which is forwarded to the noise and residual echo suppression unit. Presented idea of the system combines frequency warping and acoustic echo and noise control techniques to serve in application of communication hands-free device.

1. INTRODUCTION

In the hands-free telephone environment the reduction of acoustic echo is the main problem, which is successfully solved by the well-known acoustic echo canceller with adaptive algorithm performed in full-band or subbands schemes. But environmental and echo compensator system properties show that further problem is to reduce the residual echo leaven by the canceller as well the background noises. Most systems employ the spectral subtraction or statistic estimator on the signal to solve this problem. However, the recent studies employ the psychoacoustic consideration for speech enhancement system [1-4]. But this is an early stage of a system, which combine echo compensator and psychoacoustically motivated noise and residual echo reduction system operated in full-band.

In this paper we propose the non-uniform subband acoustic echo canceller and noise reduction system derived from cochlear model [5-7]. The main advantages of using the non-uniform frequency resolution in analysis filter bank is possibility of exploit the perception property of the human ear. A starting point to constructing of the analysis and synthesis filter bank used in proposed speech enhancement system is polyphase realisation of the FIR filter bank with non-uniform frequency resolution achieved by the frequency transformation of filter characteristic using recursive allpasses [8]. Human auditory perception model delivered by the psychoacoustic is used to create this allpass transformation filter bank [7] according to the cochlea model and critical band. Performing the acoustic echo cancellation by adaptive filtering [9-10] in subbands we get echo compensated signal, which is forwarded to the noise and residual echo suppression unit.

Most approaches exploit phenomenon of auditory masking instead of the masking threshold of a standard spectral weighting rule to improve noise reduction algorithms. This is build upon the fact that human listener will not perceive any additive signal components as long as their power spectral density lies completely below the so called masking threshold [2]. But this conclusion about listener perception can not be easily transfer to the system. Good balance between noise estimation method and weighting role is needed to prevent "musical noise" appear.

Another approach is to adopt the signal analysis to the human ear. Upon the fact that the human ear performs a nonuniform frequency on Bark scale the method of the signal decomposition allow an analysis adapted to the human perception. For this reason in some application of speech recognition, coding as well in speech enhancement systems the techniques of signal decomposition which approximate Bark scale is used in [4, 11-13]. Mainly frequency subbands derived from FFT analysis of input signal is used, for Bark scale by mapping a set of the N frequency components to M subband signals where (M<N). Even logarithmic scale decomposition can be in some use in speech

[#] This work was supported by Bialystok University of Technology under the grant W/II/1/99

enhancement system when we want to achieve some estimation of the auditory system, do not mention the straight realisation of the filter bank design especially to perform signal analysis as the ear this do.

Approach presented in this paper also adopts the signal analysis to the human ear, but we try to use the different tool with the quite new approach. Idea of presented system came from the assumption that:

- echo compensation in auditory bands can give the same advantages like uniform subband schemes, however the subjective filling of residual echo can be improve;
- noise reduction can be more efficient if perform in order to evaluate the psychoacoustic properties of the human ear.

These assumption results in system constructed upon the bands spread according to the cochlear model, where we exploit the advantages of echo cancellation in subbands and the noise reduction. So we do not expect high objective improvements of the presented system comparing to the existing solution but subjective outcome from the system can be much higher. What's more, we can adjust the subjective preference simple by manipulating the decomposition.

2. FILTER BANK

It is obvious that filter bank with uniform frequency resolution can be implemented very efficiently by using polyphase network for modulated bandpass filters, which have been derived from a common prototype lowpass and the Discrete Fourier Transformation [7,8]. If the magnitude response of the prototype lowpass FIR filter is denoted as $H(e^{j\omega})$ then the frequency response of this uniform bank are given by equation:

$$H_m(e^{j\omega}) = H(e^{j(\omega - \frac{2\pi}{N}m)}), \tag{1}$$

where $H_m(e^{j\omega})$ is m-th filter of the DFT filter bank has centre frequencies

$$\omega_m = \frac{2\pi}{M}m, \qquad m = 0,\dots,M-1, \tag{2}$$

where M is a number of bands.

2. ALL-PASS FREQENCY TRANSFORMATION

Proposed modification of this uniform filter bank (1), leads to the non-uniformly spaced frequency filter bank using allpass filter instead of delay elements at signal input. The basis of allpass frequency transformation can be found in [4, 14]. In proposed filter bank the allpass transformation of first order is used. The delay elements are substituted by the casual and stable allpass,

$$A(z) = \left(\frac{z^{-1} - \underline{a}}{1 - \underline{a}^* z^{-1}}\right), \tag{3}$$

where $\underline{a} = ae^{j\alpha}$ is a complex parameter with $|\underline{a}| < 1$, and phase of this allpass

$$F(\omega) = -\omega + 2arctan\frac{a\sin(\omega - \alpha)}{a\cos(\omega - \alpha) - 1}. \tag{4}$$

This leads to the following bandpass filters

$$H_m(e^{j\omega}) = H(e^{j(-F(\omega) - \frac{2\pi}{N}m)}), \tag{5}$$

similar with human auditory systems, which cochlea acts electrically as a highly overlapped filter bank. Solving equation (4) adequately to the smooth and monotonic warping (mapping) function $F(\omega)$ we get desired frequency warping transformation. So we can map the uniform frequency resolution to the new representation of the uniform frequency resolution on the new scale (e.g. Bark, ERB-scale) which has a non-uniform resolution when observed from old scale (Hz).

Please note that useful characteristic of the allpasss mapping is that inverse transformation is produced simply by changing λ to $-\lambda$.

3. MODIFIED FILTER BANK WITH A COCHLEAR SPACED FREQUENCY RESOLUTION

For making allpass transformation useful to create the cochlea spaced filter bank we decide to use the real parameter $\underline{a} = a$, to make a compromise between the complexity introduce by the allpass and perfect approximation of the human inner ear model. Concerning presented cochlear model [6] and idea of critical bands [5] we created the frequency resolution of the human ear in band of telephone speech (300-3400Hz) for 8kHz sampling frequency.

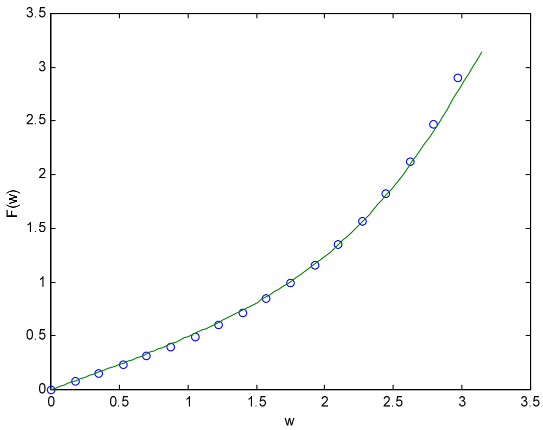

Figure 1. Approximation of the cochlea resolution frequency (solid line), same scaled freq. (points)

Knowing limitation of the assumed approach (real value of the parameter $\underline{a}$), we must decide which frequencies (lower or upper) of the model will be approximated better by the function of the allpass transformation. We decide that lower frequencies resolution must be more exact, so parameter $\underline{a} = 0.37$ is chosen.

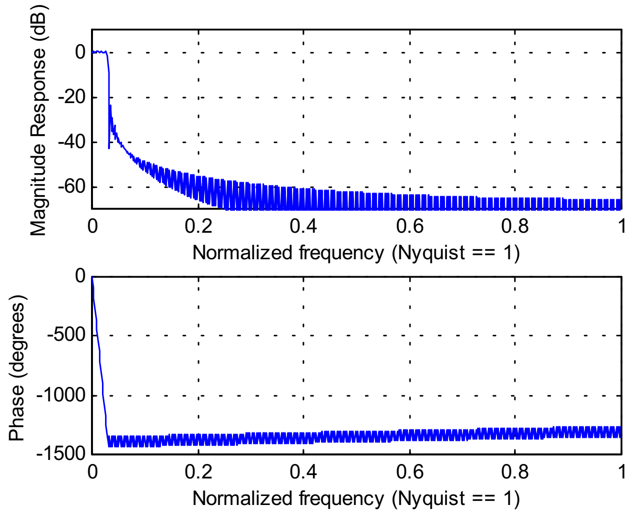

Figure 2. Prototype filter characteristic

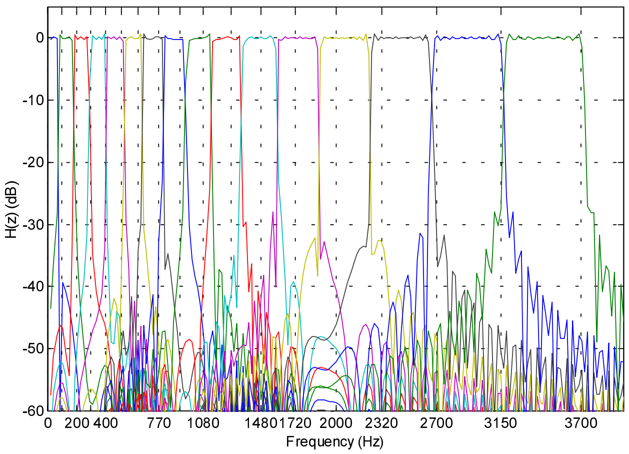

Figure 3. Nonuniform filter bank with allpass transformation based on cohlear model. Vertical dashed lines show critical bands according to Zwicker [5].

The advantage of that assumption is based on perception property of the human ear which is able to separate low frequencies more precisely than high frequencies. Comparing to the outcome $\underline{a} = 0.4034$ from formula of the Hertz-Bark-mapping presented in [15] with the same parameter of the allpass the difference can be notice, which lies in the fact that we used slightly different model of the cochlea and assume some frequency approximation. The interest in low frequencies are also motivated by the acoustical environment properties in moving vehicle [16, 17], where the noise field is highly energetic in low frequencies and most noise power is concentrated below 500Hz. It is related to the road nature, so in case of moving car can not be predicted by statistic tool. Please notice that in low frequencies human auditory system has high resolution (100 Hz band width) in terms of the critical bands. So evaluating noise reduction technique in close manner to the ear can be worth the effort.

For example lets consider a 256 taps FIR filter as prototype to projected cochlea spaced filter band (see Fig. 2), with DFT length of 32, then we get filter bank presented in Fig. 3 for selected even numbered bands.

4.SUBBAND SPEECH ENHANCEMENT

Acoustic echo cancellation is realised by naive structure [10]. The structure can be seen in Fig. 4. The loudspeaker and microphone signals are fed into identical cochlear spaced M-band analysis filter bank. Adaptive filtering by NLMS algorithm [9] is performed separately in each subband and finally outputs of subband adaptive filters are passed to subband noise reduction unit.

The FFT analysis commonly used in noisy speech enhancement systems (e.g. spectral subtraction) results in large store requirements and processing delay, so in presented system Fourier analysis is replaced by the subband structure discussed in section 2 and 3. Standard spectral subtraction time-domain algorithm [3] is applied to cochlear spaced subbands in order to attenuate the noise. Then method of postprocessing of the musical residual noise presented in [18] is used to improve overall system performance.

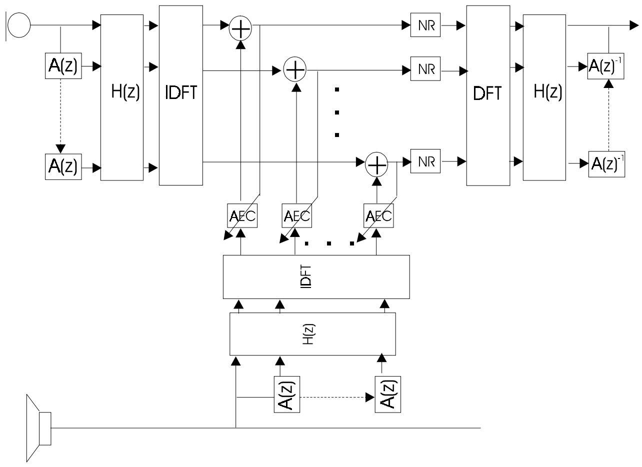

Figure 3. Speech enhancement system for hands-free devices based on non-uniform filter bank with allpass transformation

5. RESULTS

Preliminary results obtained from MATLAB simulations shows:

- easy modification and adjustment of the auditory system filter bank approximation to the specific listener and desired resolution,
- simple design technique of the filter bank,
- system are not subsampled so can not maximally exploit subband properties in acoustic echo cancellation unit,
- noise reduction unit operate on the subbands is working expected well,
- distortion to the signal made by the all-passes need pre-filter to preserve the energy of the original signal.

However further tests and research are needed to save the computation load of the method because of the all-pass transformation used and still high complexity in AEC operation because of no down sampling used in subbands. Also distortion of the amplitude appear so the design of filter must be concern.

6. CONCLUSION

We have shown that auditory model can be exploit not only from algorithmic point of view but like a dynamic tools for signal analysis and synthesis. Presented filter bank is only one from many of possible perceptual decomposition of signal, but elasticity in rapid design can make it very useful eve if the conception of the system still need improvement and solution of appearing problems.

REFERENCES

[1] S.Gustafsson, R.Martin.: Combined Acoustic Echo Control and Noise Reduction for Mobile Communications, Proc. EUROSPEECH '97, Rhodes, Greece, September 1997, pp. 1403-1406

[2] S.Gustafsson, P. Jax.: Combined Residual Echo and Noise Reduction: A Novel Psychoacustically Motiveted Algorithm, - EUSIPCO'98, Greece, Sept. 1998, pp. 961-964

[3] Sen M. Kuo, Ming Dai.: Multirate Subband Techniques for Speech Enhancement, International Conference on Signal Processing Applications and Technology (ICSPAT'98), September 13-16, 1998, Toronto, Canada

[4] A. Miksic, B. Horvat.: Subband echo cancellation in automatic speech dialog systems, EuroSpeech'97, Rhodes, September 1997, Greece, Vol. 5, pp. 2579-2582

[5] E. Zwicker, H. Fastl.: Psychoacoustics. Facts and models, Springer-Verlag Berlin Heidelberg 1990

[6] W. Wan, A. Petrovsky, C. Fan.: Two-Dimensional nonlinear cochlear model for speech processing: response to pure tones, Proceedings 6th International Fase -Congress 1992, 29-31 July, Zurich, Switzerland, pp. 233-236

[7] K. Bielawski, J. Baszun, A.A. Petrovsky.: Cochlear spaced filter bank using allpass frequency transformation, EURASIP Conf. DSP for Multimedia Communication and Services (ECMCS'99), Krakow, Poland, June 1999, (to be published)

[8] M. Kappelam, B. Straus, P. Vary.: Flexible nonuniform filter banks using allpass transformation of multiple order, - EUSIPCO'96, Italy, Sept. 1996

[9] S. Haykin.: Adaptive filter theory, Third edition, Prentice-Hall, 1996

[10] J.H. Husoy, J.E. Christiansen, F.Barstad.: Adaptive filtering in subbands: a tutorial overview, Norvegian Signal Processing Symposium 1995, Stavenger, Sept. 1-2, 1995, pp. 3-9

[11] U.K. Laine, A. Härmä.: On the design of Bark-FAMlet filtersbanks, Proc. Of NAM'96, Helsinki, Finland, 1996, pp.277-284.

[12] U.K. Laine.: Critically Sampled PR Filterbanks of Nonuniform Resolution Based on Block Recursive FAMlet Transform, Proc. EuroSpeech'97, Rhodos September 1997, Greece, Vol. 2, pp. 697-700

[13] P. Dreiseitel, H. Puder.: Speech Enhancement for mobile telephony based on non-uniformly spaced frequency resolution, EUSIPCO'98, Greece, Sept. 1998, pp. 965-968

[14] B.E. Sarroukh, A.C. den Brinker,S.J.L. van Eijndhoven.: Perfect reconstruction filter banks induced by windowed unitary transformations, PRORISC/IEEE Workshop on circuits, systems and signal processing 1997, pp.100-106

[15] J.O. Smith, J.S. Abel.: The Bark Bilinear Transform, Proc of IEEE ASSP Workshop Mohonk, New Paltz, 1995

[16] P. Scalart, A. Benamar.: A system for speech enhancement in the context of hands-free radiotelephony with combine noise reduction and acoustic echo cancelation, Speech Communication 20 (1996), Elsevier Science, Holand, pp.203-214

[17] K. Bielawski, A.A. Petrovsky.: Speaker's acoustic environment in hands-free communication device, Proceedings of the 7th Structural Acoustics Conference, Zakopane, Poland, April 1998, pp.21-27

[18] T. Haulick, K. Linhard, P. Schrögmeier.: Residual Noise suppression Using Psychoacoustic criteria, EuroSpeech'97, Rhodos September 1997, Greece, Vol. 3, pp.1395-1398