ACOUSTIC ECHO CONTROL AND NOISE REDUCTION IN NON-UNIFORM FILTER BANK: AN APPLICATION OF OVERSAMPLING MULTIRATE SYSTEMS AND ALL-PASS TRANSFORMATION *

Krzysztof Bielawski and Alexander A. Petrovsky

Department of Computer Science, Białystok University of Technology,
Wiejska 45A, 15-351 Białystok, POLAND, biel@ii.pb.bialystok.pl

ABSTRACT

This paper propose the non-uniform subband scheme with combine acoustic echo canceller and noise reduction system. Bands of the presented multirate systems are spread according to the cochlear model by the all-pass transformation of polyphase filter bank. Oversampled approach was assumed for solution because of down-sampling used, which reduce complexity with relation to full-band schemes, also adaptive algorithms used for acoustic echo cancellation converge faster, as well noise reduction algorithms operating on the signal compensated in cochlea spaced subband concern the acoustic environmental properties.

1. INTRODUCTION

Communication by the speech is often disturbed by the noise and can be influenced by the acoustic environment. Yet the human auditory system is remarkably robust in a lot of common oral communication situation. This is to a large extent due to the binaural nature of our hearing and to the (nonlinear) adaptive processing in our inner-ear and brain. Situation appears differently if communication equipment are used, especially if it is hands-free device where systems in general are not robust against adverse conditions. Noise and acoustic will degrade communication quality and efficiency especially where single microphone hands-free device is used. On top of all local noise, reverberation and mobility problems comes the need for acoustic echo cancellation to avoid that the remote correspondent will hear his own voice coming back through the communication channel. So far, the well-known acoustic echo canceller successfully solves it with adaptive algorithm performed in full-band or sub-bands schemes. But environmental and echo compensator system properties shows that remaining problem is to reduce the background noises which should significantly increase speech intelligibility. Most systems employ the spectral subtraction like method to the signal in speech enhancement system. However, the recent studies employ the psychoacoustic consideration .

In this paper the new approach to the combined echo and noise reduction system for used with the hands free devices will discuss. Polyphase filter bank transformed to auditory model is starting point to the critical band operation system for acoustic echo and noise reduction due to increase the speech intelligibility. Additionally the algorithm of noise reduction used in system is based on the utilization of the well known auditory mechanism and noise masking.

2. NON-UNFORMLY FREQUENCY SPACED FILTER BANK

Filter bank with uniform frequency resolution can be implemented very efficiently by using polyphase network for modulated bandpass filters [1], which have been derived from a common prototype lowpass $H(e^{j\omega})$ and the Discrete Fourier Transformation (DFT). The frequency response of this uniform bank are given by (1):

$$H_m(e^{j\omega}) = H(e^{j(\omega - \frac{2\pi}{N}m)}), \quad m = 0,...,M-1, \qquad (1)$$

All-pass transformation of this filter bank, leads to the non-uniformly spaced frequency filter bank where dalays were replaced by the allpass filter (2) of the first order. The basis of this transformation can be found in [2-4],

$$H_{AP}(z) = \left(\frac{z^{-1} - \underline{a}}{1 - \underline{a}^* z^{-1}} \right), \text{ where } \underline{a} = ae^{j\alpha}. \qquad (2)$$

This leads to the following filter bank filters

$$H^*_m(e^{j\omega}) = H(e^{j(-F(\omega) - \frac{2\pi}{N}m)}). \qquad (3)$$

Where the smooth and monotonic warping (mapping) function (4), which is phase of allpass

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

determine desired frequency warping transformation. So the mapping of the uniform frequency resolution (e.g. Hz) to the new representation of the uniform frequency resolution on the new scale (e.g. Log, Bark, ERB-scale) can be done which has a non-uniform resolution when observed from old scale (Hz) see fig 2.

* This work was sponsored by Białystok Univ. of Tech. grant no. W/II/2/00

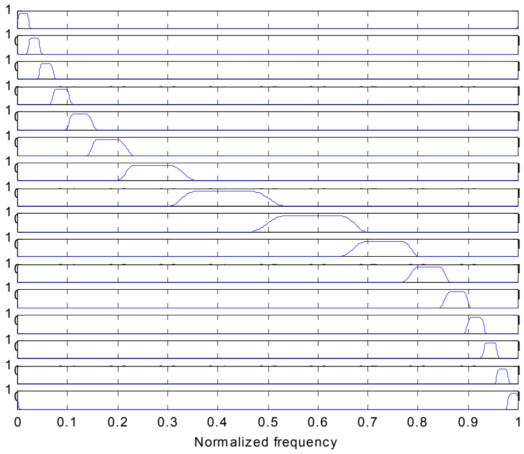

Figure 1. Filter bank achieved by the allpass transformation of the DFT polyphase filter bank.

3. COCHLEAR SPACED FILTER BANK

Since the human perception of audio signals works in Bark scale, it is appropriate to analyse the signal by a non-uniformly (cochlear) spaced filter bank [4]. This way of signal processing seems to be a perfect solution for speech enhancement system based on psychoacoustical properties of the human ear.

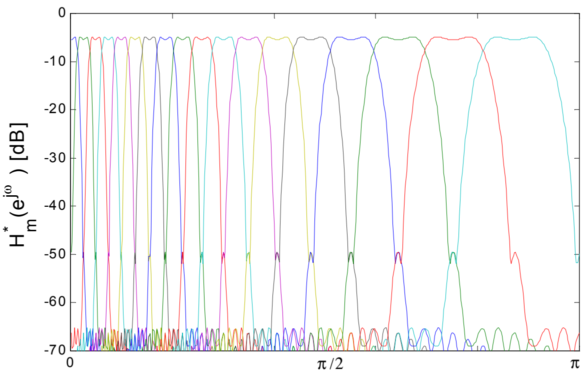

Figure 2.Approximation of the critical band frequency resolution by the allpass transformed filter bank.

For making allpass transformation useful to create the cochlea spaced filter bank the real parameter (dependent on the sampling frequency)

$$a_{f_s} \approx 1.0211\left(\frac{2}{\pi}\arctan(0.076 f_s)\right)^{\frac{1}{2}} - 0.19877 \qquad (5)$$

of the all-pass is used, to make a compromise between the complexity introduce by the allpasses and perfect approximation of the human inner ear model. The 18 critical bands filter bank for an 8kHz sampling rate and coefficient $a_{f_s} = 0.4034$ is depict in fig. 2. The flexibility of the transformation is presented in figures 1 and 2, where the filter bank according to the critical band is presented.

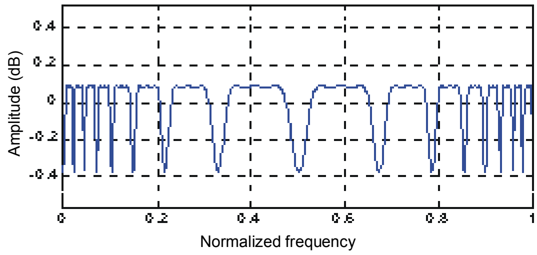

Figure 3. Logarithmic Fourier transformation of the impulse response of the analysis- synthesis structure for bank from fig. 2.

The computation complexity introduce to the polyphase filter bank is given by $2(L-1)+\frac{L}{R}+16\log_2 M$ real operation per sample in analysis part and the same number operation in synthesis part, assuming M-as a band number, L-prototype filter length, R-down-sampling factor. Future reduction can be made by the use of the DCT like filter bank with all-pass transformation

4. PSYCHOACOUSTIC MODEL

Firstly, the traditional Human auditory system model proposed by Zwicker [5] is used in this paper. Non-uniform filter bank with bands spreading function according to the cochlear model try to substitute the (non-linear) adaptive processing in our inner-ear and compensate lost of the binaural hearing in single microphone hands-free device. It has been exploit in cochlear spaced filter bank where observation on masking tones by noise led to modeling the peripheral auditory analysis by a bank of critical band filters. The postulates that sound are pre-process by these filters, whose critical bands bandwidths as a function of frequency is given by

$$BW_f = 25 + 75\left(1 + 1.4\left(\frac{f}{kHz}\right)^2\right)^{0.69} \qquad (6)$$

Secondly, Auditory Masking Threshold (AMT) defined as a spectral amplitude threshold below which all frequency component are masked in the presence of the masker signal is used. The detailed derivation of used combined methods of estimation can be found in [6,7]. Here, the brief algorithm of AMT estimation is describe where i denote Bark,

a) windowing data per Critical Band (CB)
b) total window power per CB is found $X_{CB}(i)$
c) Spectral Flatness Measure $SFM_{CB}(i)$ for each CB is found as a ratio of the geometric to the arithmetic mean of signal power
d) tonality and offset per CB is estimated according to ISO/MPEG model [7],

$$t_{CB}(i) = \min\left\{\frac{10*\log_{10}(SFM_{CB}(i))}{SFM_{MAX}},1\right\} \qquad (7)$$

$$O_{CB}(i) = t_{CB}(i) * (-0.275 * (i-1) - 15.025) - ... \\ - 9 * (1 - t_{CB}(i)) \quad (8)$$

e) Find shifted Bark energy spectrum

$$D_{CB}(i) = X_{CB}(i) * 10^{(O_{Cb}(i)/10)} \quad (9)$$

f) Find spreading of Bark energy $C_{CB}(i)$ as a convolution of linear version of Spreading Function [5,6] with shifted bark energy

g) Find temporal masking coefficient for each critical band $F_{CB}(i)$

h) Finally AMT in dB is

$$AMT_{CB}(i) = C_{CB}(i) * \max\left(\frac{F_{CB}(i)}{X_{CB}(i)}, 1\right) \quad (10)$$

i) Normalization AMT to absolute threshold of hearing

5. SPEECH ENHANCEMENT SYSTEM

Speech enhancement system depict in fig. 4 is build based on polyphase realisation of the filter bank where the non-uniform frequency resolution is achieved by the frequency transformation of FIR filter prototype characteristic using recursive allpasses [3,4]. Human auditory perception model delivered by the psychoacoustic is used to create this allpass transformation (see sec. 3). Performing the acoustic echo cancellation by adaptive filtering in subbands we get echo compensated signal, which is forwarded to the noise suppression unit also operating on subband signal each.

5.1. Acoustic echo canceller

The AEC task is to identify of the environmental impulse response by mean of an adaptive filter. We used a transversal structure FIR filter with normalised LMS algorithm [8]. The global speech activity detector is used for driving the adaptation process. AEC operate on subband signal with reduce sampling rate comparing to the full band input signal. According to this reduction the length of adaptive filters can be also truncated for detailed comparison of the full and sub-band approach [9].

5.2. Noise reduction system

Speech enhancement algorithm concerning additive noise is proposed in [10], which used a parametric nonlinear gain in minimization process of audible noise. Assuming the clean speech spectrum is known the audible noise spectrum is defined as difference between an audible spectrum of noisy speech and clean.

$$A_{CBn}(i) = A_{CBy}(i) - A_{CBx}(i), \quad (11)$$

with assumption that the noisy speech consist of the sum of the clean speech and noise

$$CB_y(i) = CB_x(i) + CB_n(i). \quad (12)$$

Therefore the criterium for enhancement is,

$$A_{CBn}(i) \le 0, \text{ for each CB.} \quad (13)$$

Solution which satisfy is presented in [10] but brief derivation follows,

$$CB_{\hat{x}}(i) = \frac{CB_y(i)}{a_{CB}{}^{v_{CB}(i)}(i) + CB_y^{v_{CB}(i)}(i)} CB_y(i), \quad (14)$$

where $a_{CB}(i)$ and $v_{CB}(i)$ are time-critical band varying parameters defining the threshold for inaudible components and controls the rate of suppression.

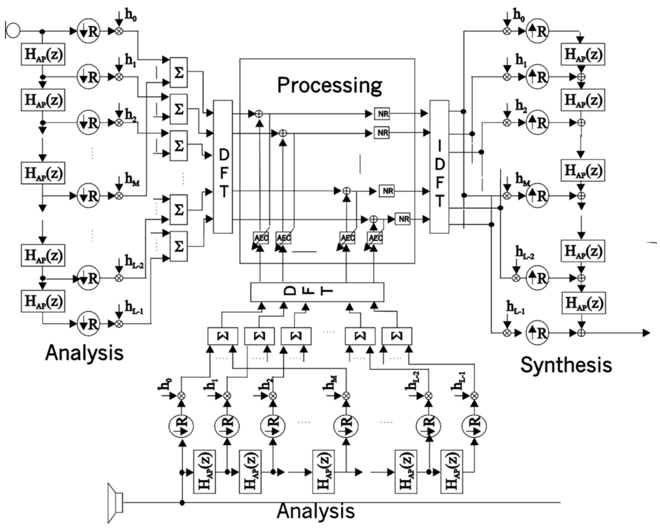

Figure. 4. Speech enhancement system block diagram

6. IN BAND PROCESSING

Basically, all components needed to in band processing were presented earlier in this paper. The noise and echo corrupted signal is fed to the analysis polyphase modify filter bank to get the critical band signals Y_{CB} then AEC with control mechanism used for NLMS algorithm of adaptation is used with help of speech detector (VAD) which operate on full band incoming signal. For more precise description of detector see [11].

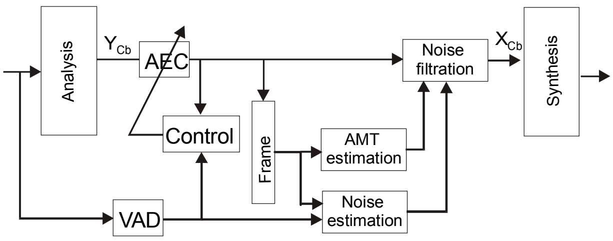

Figure 5. Block diagram of in band processing

Then the echo compensated signal is fed to the noise reduction system operating on framed data to calculate the denoise filter. Then the noise filtration is made to get the enhanced signal in band X_{CB}.

7. EXPERIMENTS

Early experiment made for system are shown in figures 6, 7 and 8. They have been made with following parameters: number of bands 36 (18 bark for 8 kHz sampling rate), filter prototype length 144 taps, downsampling ratio 4, frame of 64 samples (due to the sampling reduction of 32ms for assumed speech stationary characteristic).

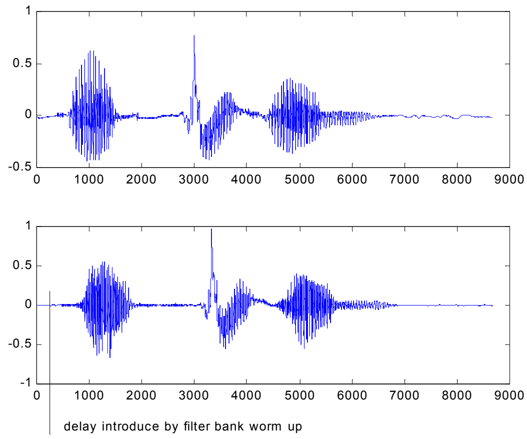

Figure 6. Clean speech signal reconstruction based on psychoacustical information in presented filter bank.

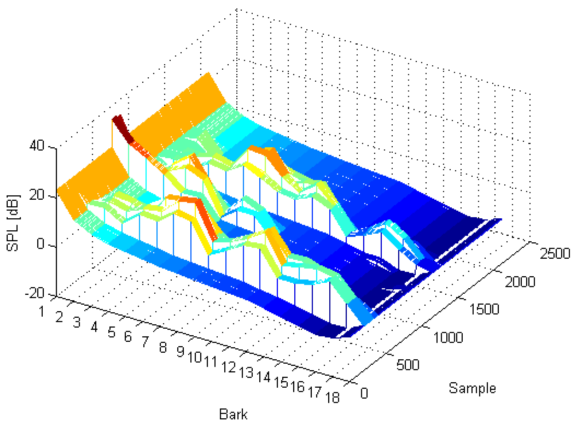

Figure 7. AMT dynamic of the clean speech signal processed in filter bank.

The experiments measured the distortion caused by the all-pass transformation were made in order to reconstruct psychoacusticaly equivalent representation of the signal. The subjective listener test shows no significant change in acoustic perception of signals. Figures 8 shows the results with the noisy speech resulting in perceptual equivalent representation.

8. CONCLUSION

For large adaptive filter length, subband processing is superior to full-band processing in terms of computational complexity. The proposed system exploit this property although designing procedure of the modify filter bank, must find a compromise between the prototype filter length, subsampling rate, stop-band attenuation and aliasing. If this is done in pragmatic way we get the next advantage: signal is split into bands according to the mapping function so our design can concern uniform logarithmic, Bark, ERB or other frequency resolution in our research.

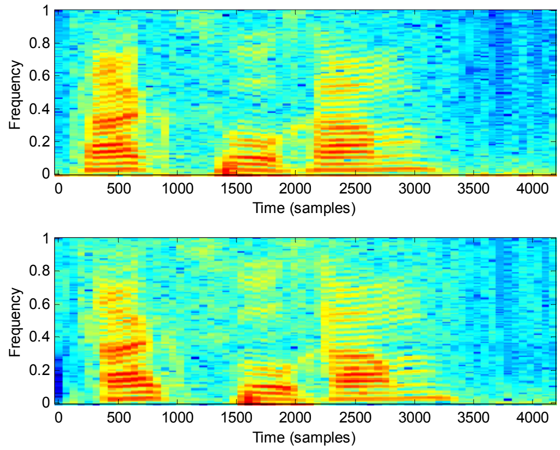

Figure 8. Noisy speech signal reconstruction based on psychoacustical information in presented filter bank.

9. REFERENCES

[1] G. Wackersreuther, “On the design of filter for ideal QMF and polyphase filter banks,” *AEÜ,* vol. 39, no. 2, pp. 123-130, April 1985

[2] J.O. Smith and J.S. Abel, “Bark and ERB Bilinear Transforms,” *IEEE Trans. Speech and Audio Processing*, vol. 7, no. 6, pp. 697-708, Nov. 1999.

[3] M. Kappelan and B. Strauß and P. Vary, ”Flexible nonuniform filter banks using allpass transformation of multiple order,” *in Proc. EUSIPCO'96,* Trieste, Italy, Sep. 1996, pp.

[4] K. Bielawski and J. Baszun and A.A. Petrovsky, “Cochlear spaced filter bank using allpass frequency transformation,” *in Proc. EURASIP Conf. DSP for Multimedia Commun. and Services (ECMCS’99),* Krakow, Poland, June 1999

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

[6] D.J. Johnson, “Transform coding of audio signals using Perceptual Noise Criteria”, *IEEE Journal on Selected Areas of Communication,”* vol. 6, no.2, pp. 314-323, Feb. 1988

[7] ISO/IEC IS13818-3, Information technology – Generic Coding of moving pictures and associated audio information, Part 3. Audio, 1994

[8] K. Bielawski and A. Petrovsky, “Joint system for acoustic echo and noise control in hands-free communication devices,” *Electronics and Telecommunication,* vol. 44, no. 4, Warsaw, 1998, pp.115-136

[9] G. Schmidt, “Acoustic Echo Control in subbands – an application of multirate systems,” *in Proc. EUSIPCO'98,* Rhodes, Greece, Sep. 1998, pp. 1961-1964

[10] D.E. Tsoukalas, J.N. Mourjopoulos and G. Kokkinakis, “Speech Enhancement based on Audible noise suppression,” IEEE Trans. Speech and Audio Processing, vol. 7, no. 6, pp.497-513, Nov. 1997.

[11] P. Sovka and P. Pollak, “The study of speech/pause detectors for speech enhancement methodos”, *in Proc. Eurospeech’95,* pp.1-4