PSYCHOACOUSTICALLY MOTIVATED NONUNIFORM COSINE MODULATED POLYPHASE FILTER BANK*)

A. Petrovsky, M. Parfieniuk, K. Bielawski
Department of Real Time Systems, Bialystok Technical University, Poland
e-mail: palex@it.org.by

ABSTRACT

Nonuniform cosine modulated filter bank is presented as valuable solution for perceptual sound processing systems. Proposed structure connects polyphase concept with idea of warping frequency by all-pass transformation. So it has uncommon flexibility of bandwidths shaping, relatively easy design and good performance. System is considered in context of critical-bands model approximation as front-end for audio coding and enhancement. Theoretical basics of fundamental ideas are reviewed and connecting them into final solution is shown. Issues of the design and details of the implementation are discussed, basing on examples of filter banks approximating well-known nonlinear Bark and ERB psychoacoustic scales. Perspectives of real-time dynamic tuning up the bandwidths are also considered.

1 INTRODUCTION

Perceptual processing takes very important place in contemporary acoustic signal processing. It bases on inherent properties of human auditory system. Some boundaries of human's hearing mechanisms allow to significantly improve both efficiency and performance of algorithms such as audio compression, transmission and enhancement. One of main results in this field is revealing of critical-bands existence [1]. This phenomenon consists in that so our ears analyze sounds with sensitivity nonlinearly (logarithmically) distributed over frequency - low tones are separated more precisely than high ones. It is possible to point out bands representing the different levels of selectivity. Thus audio processing system should approximate that model to emphasize features which are significant for auditor ears and ignore negligible components.

Direct approach is to use filter bank with bandwidths approximating critical-bands and then process each band separately by appropriate algorithm.

There are many kinds of non-uniform filter banks differencing in structural and functional properties. For example, in [2], unequal bandwidths are achieved by connecting the outputs of uniform filter banks. Disadvantage is obvious – bandwidth distribution is constrained to regular grid fixed by underlying uniform subband decomposition. Approximation of more sophisticated distributions requires increasing the density of this grid – number of channels in uniform filter bank. Moreover, there is no possibility to, even slightly, change bandwidths without restructuring of the filter bank. The same applies to well-known tree structured (octave) filter banks build by pyramidal stacking of two channels systems. Another approaches are the transforms such as the wavelet packet transform – but idea of these solutions is rather different from direct subband decomposition considered here.

The frequency warping by all-pass filters is the technique allowing to embed nonlinearity inherently into arbitrary DSP system. Idea basing on bilinear mapping was shown in [3] in the seventies, but its popularity grown in last ten years – for good survey see [4].

As a pioneering work on the warping of the entire filter bank can be treated [5], where it was described that well-known uniform polyphase DFT filter bank [6] can be transformed into nonuniform one by all-pass transformation. Solution is elegant - relatively simple and robust [7]. But main disadvantage of this approach lies in fact that channel signals are complex. This forces further processing algorithm to deal with complex numbers. So as structural as computational complexity grows. Moreover transition from real signal (for example speech) to complex sequences is often somewhat artificial.

In this article exchange the DFT for cosine modulation is presented as solution for above trouble. This is equivalent to warping of cosine modulated filter bank. In the resulting system, channel signals are purely real. This comes at cost of higher complexity of the bank, but now subband processing algorithms dealing with real numbers have decreased computational load. System remains common advantages of its former and common polyphase systems i.e. easy design depending on one prototype filter and computational profits given by decimation of channel signals.

2 COSINE MODULATED BANK BASICS

Uniform cosine modulated filter banks [6] are commonly known and used in digital signal processing. The genesis of their name is in fact that band filters are cosine modulated versions of one prototype. The most basic and simplest are pseudo QMF filter banks. In these structures phase distortions are completely eliminated, aliasing is

*) This work was supported by Bialystok Technical University under the grant W/WI/3/02

suppressed only approximately and amplitude distortions can be reduced by proper design of prototype filter. It is evident that they aren't exactly perfect reconstruction (PR) systems but their features are satisfactory in common applications.

Band filters in K-channel pseudo QMF system are related to prototype of length N in the following manner

$$\begin{aligned} h_k(n) &= 2p(n)\cos\left(\frac{p}{K}\left(k+\frac{1}{2}\right)\left(n-\frac{N}{2}\right)+q_k\right) && \text{analysis filters} \\ f_k(n) &= 2p(n)\cos\left(\frac{p}{K}\left(k+\frac{1}{2}\right)\left(n-\frac{N}{2}\right)-q_k\right) && \text{synthesis filters} \end{aligned} \quad (1)$$

$$\text{where} \quad k = 0..K\text{-}1, \quad n = 0..N, \quad q_k = (-1)^k \frac{p}{4}$$

Prototype $p(n)$ is supposed to be even order real-coefficient FIR filter with cut-off frequency $p\,/\,2K$.

Extensive derivation of these formulas and analysis of overall system can be found in [6]. We focus only on the details of the implementation of cosine modulation, equivalent to discrete cosine transform. The book [8], classical on this subject, gives extended review of the algorithms for DCT. Here solution using FFT is applied. Idea is in treating of real-coefficient filter as combination of two filters with complex coefficients.

It is well known that cosine can be represented as sum of two conjugate exponentials so

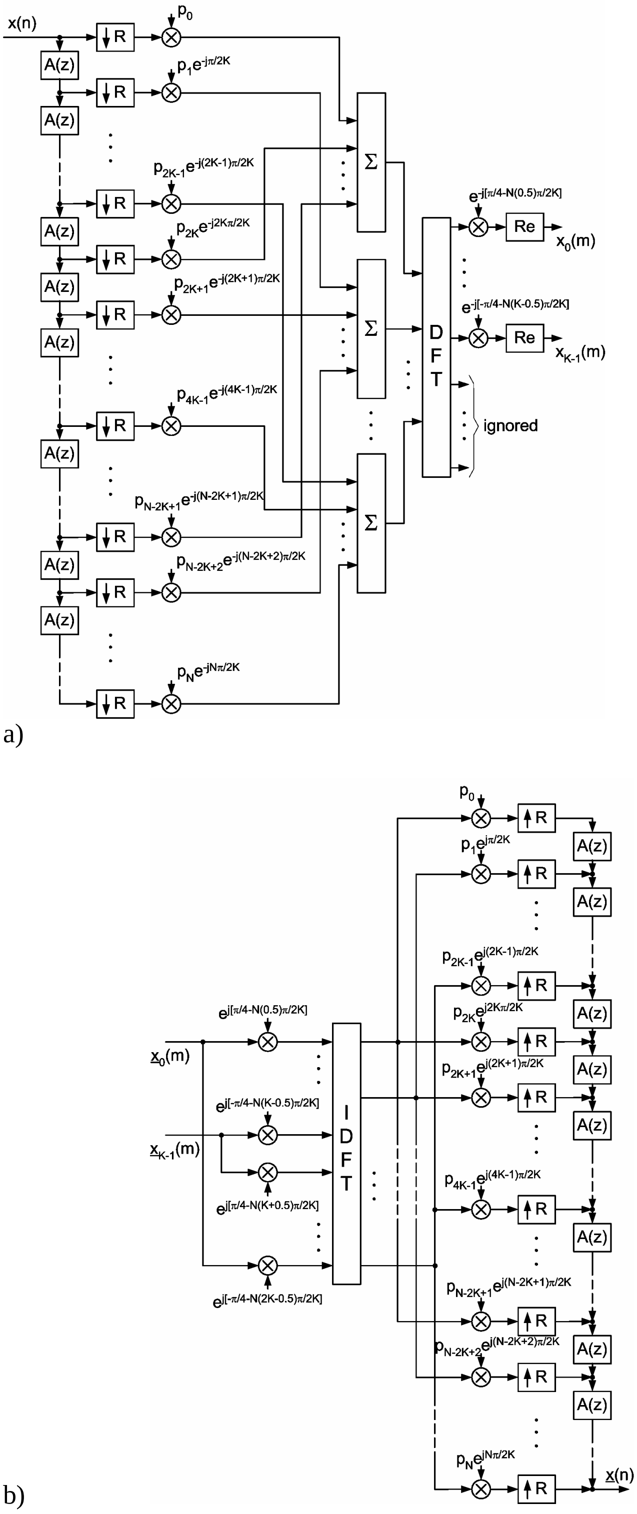
$$\begin{aligned} h_k(n) &= p(n)\left[e^{j\left[\frac{p}{K}\left(k+\frac{1}{2}\right)\left(n-\frac{N}{2}\right)+q_k\right]}+e^{-j\left[\frac{p}{K}\left(k+\frac{1}{2}\right)\left(n-\frac{N}{2}\right)+q_k\right]}\right] \\ &= 2\mathrm{Re}\left[e^{-j\left[q_k-\frac{p}{K}\left(k+\frac{1}{2}\right)\frac{N}{2}\right]}p(n)e^{-j\frac{p}{K}kn}e^{-j\frac{p}{2K}n}\right] \end{aligned} \quad (2)$$

Overall complex modulation function in last expression is the product of three simpler functions – each of them has different character and can be implemented in specific manner. Factor $e^{-j\frac{p}{2K}n}$ is independent from channel number so it may be connected with prototype's coefficients by their premodulation to complex form. Factor $e^{-j\frac{p}{K}kn}$ seems to be similar to Discrete Fourier Transform coefficients. Exactly, it can be computed by DFT of doubled size 2K. Good way to do this involves Fast Fourier Transform algorithm. Last modulation factor can be realized by multiplying FFT outputs by proper complex constants. Then getting doubled real parts of products gives desired results. Analogue is true for the synthesis filters.

Thus cosine modulation can be done at cost

$$\underbrace{N}_{\substack{\text{multiplication}\\ \text{by FIR coeffs}}} + \underbrace{2K\log_2 2K}_{\text{FFT}} + \underbrace{K}_{\substack{\text{FFT outputs}\\ \text{multiplication}}} \quad (3)$$

complex multiplications, in opportunity to direct matrix multiplication at cost KN real operations. Further computational benefits can be achieved by decimation of channel signals. Following the multirate rules decimation factor R can be up to K (critical sampling case).



a)

b)

Fig. 1 Overall structure of nonuniform cosine modulated filter bank for analysis (a) and synthesis (b) case

3 ALL-PASS TRANSFORM IDEA

As it was mentioned earlier, the fundamentals of frequency transform in all-pass filter chain can be found in [3, 4, 5]. Modification of uniform cosine modulated filter bank by substituting all delay elements by first order all-pass filter (causal and stable)

$$z^{-1} \Rightarrow \frac{z^{-1}+\underline{a}}{1+\underline{a}z^{-1}} \qquad \underline{a} = ae^{ja} \quad \wedge \quad |\underline{a}| = a < 1 \tag{4}$$

gives system with deformed nonuniform bandwidths

$$\begin{aligned} H_k(e^{jw}) &= e^{jq_k} W^{(k+0.5)N/2} P\left(e^{jw} W^{k+0.5}\right) \\ &\quad + e^{-jq_k} W^{-(k+0.5)N/2} P\left(e^{jw} W^{-(k+0.5)}\right) \\ &\Rightarrow \\ H_k(e^{-jf(w)}) &= e^{jq_k} W^{(k+0.5)N/2} P\left(e^{-jf(w)} W^{k+0.5}\right) \\ &\quad + e^{-jq_k} W^{-(k+0.5)N/2} P\left(e^{-jf(w)} W^{-(k+0.5)}\right) \end{aligned} \tag{5}$$

The factor $f(w)$ in the above expression is the all-pass phase response, which is the following nonlinear function of frequency

$$f(w) = -w + 2\arctan\frac{a\sin(w-a)}{a\cos(w-a)-1} \tag{6}$$

It is obvious that nonlinearity depends directly on filter coefficient. For the case $a = 0$ function become linear and filter is equivalent to pure delay.
Final structure of filter bank is shown in Figure 1.

All-pass chain computation must be done before decimation so computational load significantly increases

$$\underbrace{RN}_{\text{all-pass chain computation}} + \underbrace{N}_{\text{multiplication by FIR coeffs}} + \underbrace{2K\log_2 2K}_{\text{FFT}} + \underbrace{K}_{\text{FFT outputs multiplication}} \tag{7}$$

Moreover restrictions on decimation factor R arise. Because of phase distortions and changes (particularly extension) of bandwidths with regard to uniform case risk of aliasing is dangerous. Critical sampling is impossible without errors – optimal results were observed for $R < K/4$. So computational profits given by decimation are smaller.

4 PSYCHOACOUSTIC MODELS

Most popular approach in modeling of human auditory system is to analyze signals in nonuniformly (cochlear) spaced filter bank [9]. There are known [1] two alternative propositions of psychoacoustic scales e.g. distributions of bandwidths such a filter bank. Classical approach of Bark scale was determined empirically in masking experiments. It defines 24 critical bands on frequency range 0 – 15.5 kHz and may be extrapolated to higher sampling rates. Analytically it is expressed by the following formula

$$Bark(f) = 13\arctan(0.76f) + 3.5\arctan\left[\left(\frac{f}{7.5}\right)^2\right] \tag{8}$$

$$f \text{ in kHz}$$

The second, improved Equivalent Rectangular Bandwidth (ERB) scale is defined analytically as

$$ERB(f) = 21.4\log_{10}(0.00437f) \tag{9}$$

$$f \text{ in Hz}$$

Frequency warpings defined by both these expressions can be compared in Figure 2.

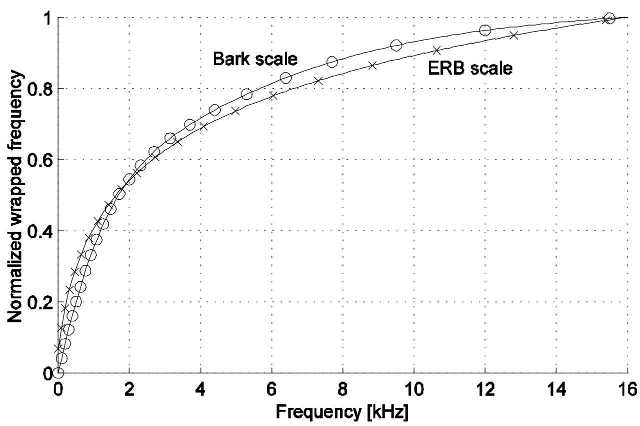


Fig. 2 Bark and ERB frequency warpings for a sampling rate of 16 kHz

To build a filter bank approximating such a scale it is needed to determine proper all-pass coefficient. All-pass transform mathematically is interpreted as bilinear conformal mapping (with one degree of freedom), which maps the unit circle to itself. For $0 < a < 1$ it is done in such a way that low frequencies are stretched and high ones are compressed. Noticing the above facts, basing on arctangent approximation, two formulas were proposed in work [1] .

$$a_{Bark} = 0.1957 - 1.048\left[\frac{2}{p}\arctan\left(0.07212\frac{f_s}{1000}\right)\right]^{\frac{1}{2}} \tag{10}$$

$$a_{ERB} = -0.7164\left[\frac{2}{p}\arctan\left(0.09669\frac{f_s}{1000}\right)\right]^{\frac{1}{2}} - 0.08667 \tag{11}$$

They both give all-pass transform coefficient for selected sampling frequency in Hz, what is shown in Figure 3 for $f_s = 4..48\text{kHz}$.

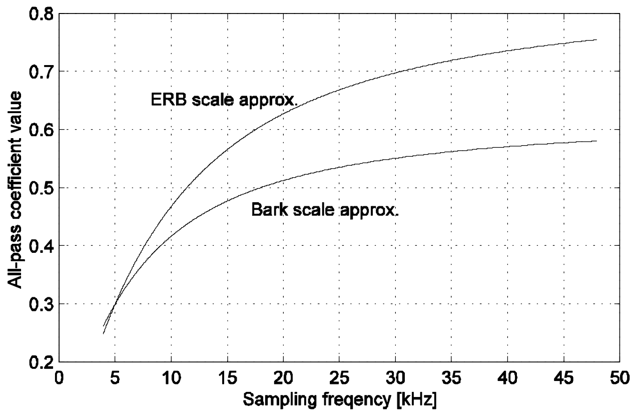


Fig. 3 Approximations of all-pass coefficient giving Bark and ERB scales

4 DESIGN OF FILTER BANKS APPROXIMATING PSYCHOACOUSTIC SCALES

Following the rules from previous sections, two banks are designed. The first approximates Bark scale, the second ERB approach. Sampling frequency was 8kHz (for processing of telephony speech 300 – 3400 Hz). Used all-passes' characteristics are shown in Figure 4.

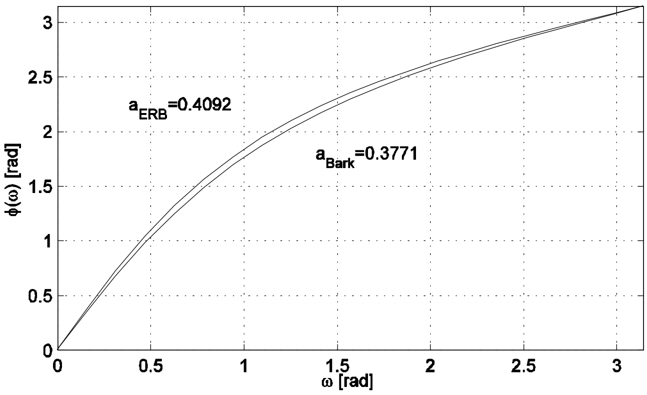


Fig. 4 Phase responses of all-pass giving approximation of Bark and ERB scales

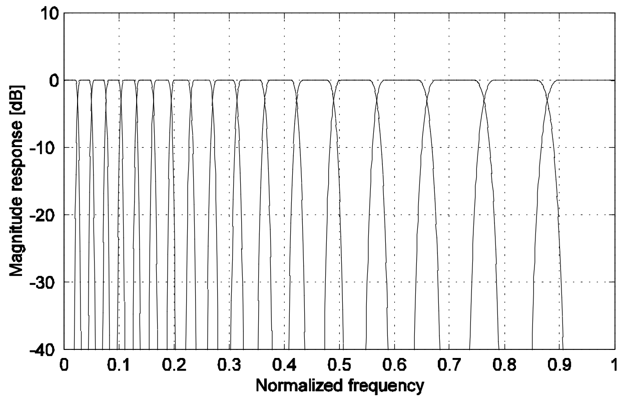
For selected sampling frequency, there are 18 critical bands so that is the number of channels in the Bark filter bank. In ERB case bandwidths are narrower so there are 19 channels. Figure 5 presents resulting bandwidths.

The prototypes were designed by standard simple windowing procedure – according to the formula

$$h(n) = \frac{\sin\left(\left(n - \frac{N}{2}\right) \cdot w_c\right)}{p \cdot \left(n - \frac{N}{2}\right)} w(n) \quad n = 0..N \qquad (12)$$

The cut-off frequency w_c was initially set to $p/2K$ and then optimized to give minimal amplitude distortions of overall analysis-synthesis system. Windowing function $w(n)$ was chosen appropriately to acceptable tradeoff between stopband attenuation and filter order. The procedure isn't sophisticated but gives prototypes with satisfactory quality.

a)



b)

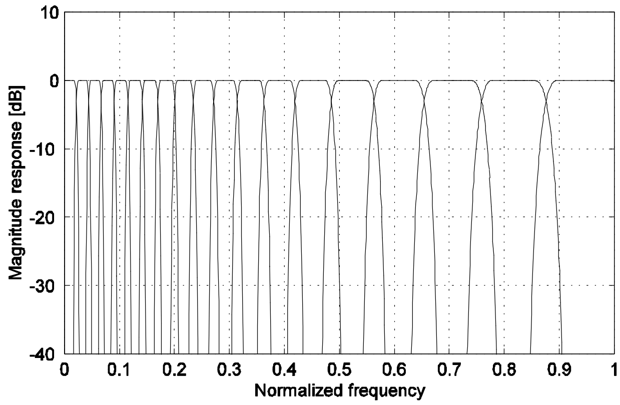
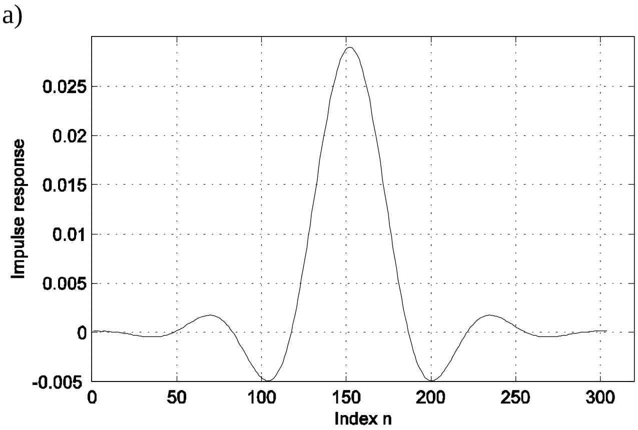


Fig. 5 Bandwidths corresponding to Bark (a) and ERB (b) scales

Sample response of filter of order 304, generated using Hamming window and $w_c = 0.0289p$ is presented in Figure 6.

a)



b)

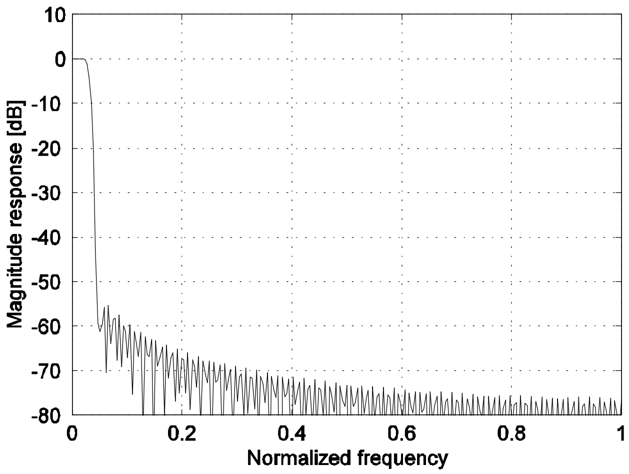
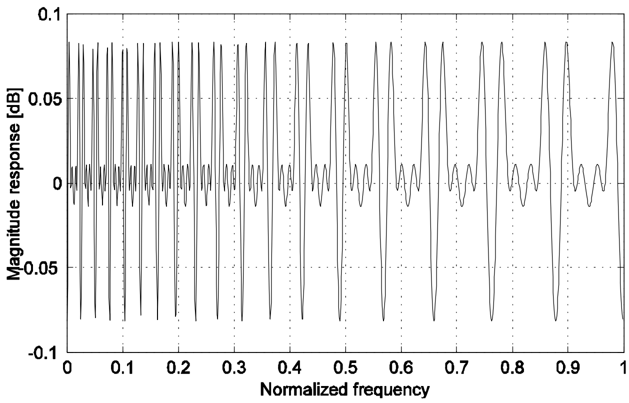


Fig. 6 The impulse (a) and frequency (b) response of sample prototype filter

Resulting amplitude distortions of system's response not exceed 0.08 dB. The same prototype remains valid in approximation as Bark as ERB scale. This is shown in Figure 7.

a)



b)

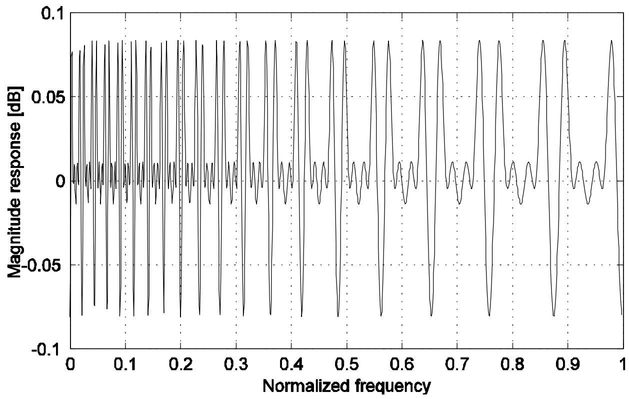


Fig. 7 Errors in the response of overall analysis-synthesis system - Bark (a) and ERB (b) approximation cases

More advanced designs using Chebyshev or Least Squares error measures and respecting PR requirements undoubtedly would give better results.

5 EXAMPLE OF PRACTICAL APPLICATION IN SPEECH ENHANCEMENT SYSTEM

The filter bank presented earlier can be used as a psychoacoustic frequency decomposition tool in combined system of acoustic echo control (AEC) and noise reduction (NR) for telephonic devices. The processing is performed according to the schema sketched in Figure 8.

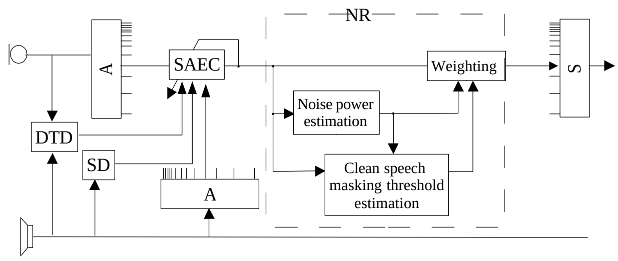


Fig. 8 In band processing schema of speech enhancement system

Echo control is set up with the use of NLMS adaptation algorithm applied separately to each band with control mechanism operating on full band of incoming microphone and loudspeaker signals. NR solution bases on the audible noise suppression rule shown in [10]. Because system setup provides only the corrupted signal, the intermediate technique must be used to estimate the masking threshold of the clean speech. Nonlinear noise estimate tracking method [11] with linear averaging for each frame, as in case of signal power spectrum is applied. Also the rough speech enhancement method giving the estimate of the clean speech based on power spectral suppression rule [12] is used.

Example plots and spectrograms approaching processing of noised speech are depicted in the Figure 9.

Proposed combined system meets the ITU requirements for hands free devices for echo attenuation. Additionally, the proposed rule of noise reduction improves the speech intelligibility, what was confirmed by the objective listener test and informal subjective measures.

To the evaluation of the system performance, the following measures are used

a) Segmental signal-to-noise ratio was calculated at speech activity according to average value of frame ratio:

$$SEGSNR_n^{s,i} = 10\log\left(\frac{\sum_{k=0}^{W-1} s^2(k+iW)}{\sum_{k=0}^{W-1} n^2(k+iW)}\right) \tag{13}$$

where i - frame index of speech activity

W - frame length

s - speech signal n - noise signal

b) The objective measure of noise attenuation was taken as

$$NA = 10\log_{10}\left(\frac{1}{O(K_n)}\sum_{k\in K_n}\frac{n^2(k)}{\hat{n}^2(k)}\right) \qquad (14)$$

where K_n - set of speech pauses

$\hat{n}(k)$ - attenuated noise

$O(K_n)$ - number of samples in set

c) The objective measure of noise attenuation difference of speech and speech distortion after processing which has the high correlation with results from auditive tests.

$$SEGSNR^{s,i}_{\hat{s}-s} = 10\log\left(\frac{\sum_{k=0}^{W-1} s^2(k+iW)}{\sum_{k=0}^{W-1}\left(\hat{s}(k+iW)-s(k+iW)\right)^2}\right) \qquad (15)$$

where $\hat{s}(k)$ - enhanced speech

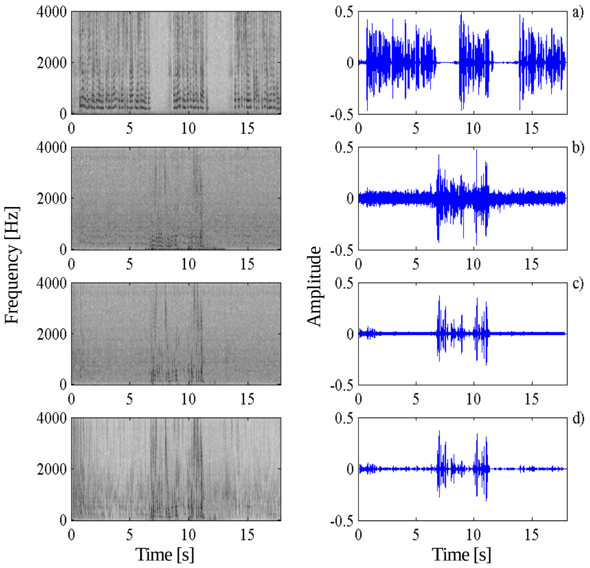


Fig. 9 Spectrograms and time plots of signals in considered system: a) loudspeaker signal, b) microphone signal noised at $SEGSNR = 5dB$, c) signal enhanced with preserving predefined background noise as comfort one, d) enhanced signal without comfort noise

The considered system was compared to that described in [13], where echo control and noise reduction are done in similar manner but psychoacoustic is implemented by weighting the spectrum given by commonly known filter bank overlap-and-add method. The results of comparison, in terms of above measures (SEGSNR is averaged), are presented in Figure 10.

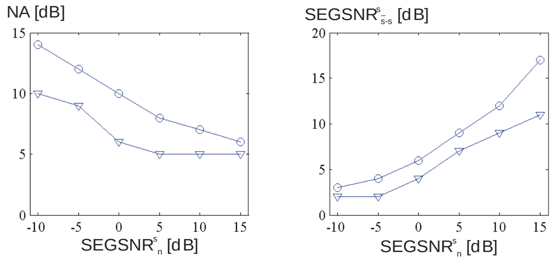


Fig. 10 Instrumental measurement data obtained from simulation of proposed NR rule (triangle) and the psychoacoustical rule presented in [13] (circle)

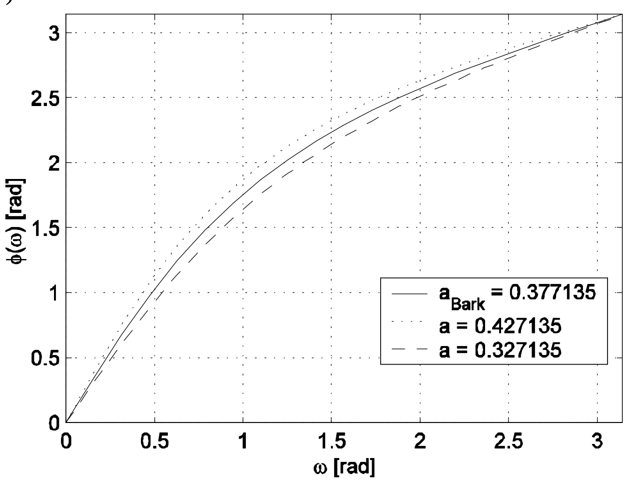
More details of analogous audio enhancement solution, but based on the DFT modulated non-uniform filter bank, are shown in [14] and author's Ph. D. thesis. Another examples of use of frequency warped filter banks in sound enhancement are also mentioned in [4].

6 TUNING UP OF THE BANDWIDTHS

The need for the tuning of the bandwidths distribution arises from the fact that psychoacoustic possibilities of different people also differ because individualities. So models such as Bark scale is only generalization with error even about 20% for particular individuals. The tuning allows to adjust filter bank characteristics and therefore overall audio processing system to personal preferences of its users. Experiments show that this approach often gives real benefits in quality.

Frequency warped filter bank can be tuned up by simple change of all-pass transform coefficient. Moreover this change can be done in real time – by interaction of the operator with working device.

a)



b)

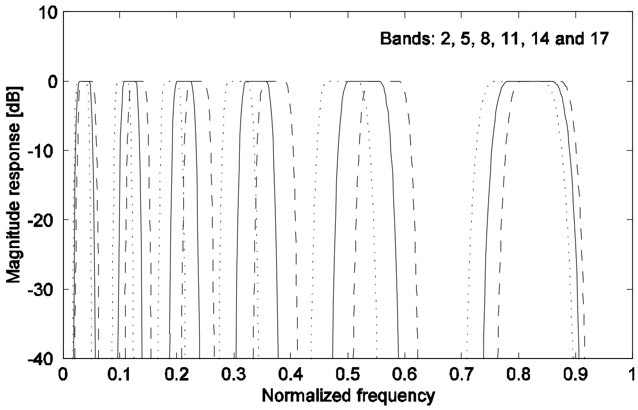


Fig. 11 Influence of coefficient change on all-pass phase response (a) and bandwidths (b) for filter bank approximating Bark scale

Even relatively slight modification of *a* results in significant changes in bandwidths. It is shown in Figure 11, that change in *a* of ±0.05 shifts the bands' center frequencies of 10 - 50% their bandwidths, which simultaneously squeeze / spread of 5 - 10%.

This reasonable case doesn't require simultaneous modifications in other parts of system (mainly phase corrector). Otherwise, large deviation in transform coefficient requires deeper changes in entire processing schema.

7 CONCLUSIONS

Cosine modulated filter bank was configured to approximate nonuniform bandwidths corresponding to psychoacoustic Bark and ERB scales. Transition from uniform to nonuniform system by bilinear transformation, done in first-order all-pass chain, was explained. Possible ways of the design and parameters selection were revealed. Example of practical application in sound enhancement system was given. Finally advances and limitations of bandwidths tuning were considered.

REFERENCES

[1] J.O. Smith, J.S. Abel, "Bark and ERB Bilinear Transforms", IEEE Trans. of Speech and Audio Processing, Vol. 7, No. 6, 1999, pp. 697-708

[2] F. Argenti, E. Del Re, "Non-Uniform Filter Banks Based on a Multi-Prototype Cosine Modulation", Proc. International Conference on Acoustic, Speech and Signal Processing ICASSP'96, Vol. 3, Atlanta, Georgia, 7-9 May 1996, pp. 1511-1514

[3] A. Oppenheim, D. Johnson, K. Steiglitz, "Computation of spectra with unequal resolution using Fast Fourier Transform", Proc. of IEEE, Vol. 59, No 2, 1971, pp. 299 - 301

[4] A. Härma, M. Karjalainen, L. Savioja, V. Välimäki, U.K. Laine, A.J. Huopaniemi, "Frequency-Warped Signal Processing for Audio Applications", J. Audio Eng. Soc., Vol. 48, No 11, November 2000, pp. 1011 – 1030

[5] M. Kappellan, B. Strauss, P. Vary, „Flexible nonuniform filterbanks using allpass transformation of multiple order", Proc. European Signal Processing Conf. EUSIPCO'96, Vol. 3, Trieste, Italy, September 1996, pp. 1745 – 1748

[6] P. P. Vaidyanathan, "Multirate Systems and Filter Banks", PTR Prentice Hall Englewood Cliffs, New Jersey, 1993

[7] A. Petrovsky, K. Bielawski, M. Parfieniuk, "Nonuniformly frequency spaced reconfigurable filter bank and its application to speech processing", 7th Int. Conf. MIXDES 2000, Gdynia, Poland, 2000, pp. 461 - 466

[8] K.R. Rao, P. Yip, "Discrete Cosine Transform – Algorithms, Advantages and Applications", Academic Press, 1990

[9] K. Bielawski, J. Baszun, A.A. Petrovsky, "Cochlear spaced filter bank using allpass frequency transformation", Proc. EURASIP Conference DSP for Multimedia Communi-cation and Services, Cracow, Poland, June 1999, CD

[10] D.E. Tsoukalas, J.N. Mourjopoulos, G. Kokkinakis, "Improving the intelligibility of noisy speech using an audible noise suppression technique", Proc. 5th European Conf. on Speech Communication and Technology EURO-SPEECH'97, 1997, pp. 1415-1418

[11] G. Doblinger, "Computationally Efficient Speech Enhancement by Spectral Minima Tracking in Subbands", Proc. 4th European Conf. on Speech Communication and Technology EUROSPEECH'95, Madrid, Spain, 18-21 Sept. 1995, pp. 1513-1516

[12] M. Berouti, R. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise", Proc. IEEE International Conf. on Acoustic, Speech and Signal Processing, 1979, pp. 208-211

[13] S. Gustafsson, P. Jax, P. Vary, A Novel "Psychoacoustically Motivated Audio Enhance-ment Algorithm Preserving Background Noise Characteristic", Proc. International Conf. on Acoustic, Speech and Signal Processing, 1998, CD

[14] K. Bielawski, A.A. Petrovsky, "Speech enhancement system for hands-free telephone based on the psychoacoustically motivated filter bank with allpass frequency transformation", Proc. 6th European Conf. on Speech Communication and Technology EURO-SPEECH'99, Budapest, Hungary, Sept. 1999, pp. 2555-2558