AUDITORY MODEL BASED SPEECH ENHANCEMENT SYSTEM FOR HANDS-FREE DEVICES†

Alexander Petrovsky, Krzysztof Bielawski

Białystok Technical University
Faculty of Computer Science, Real Time Systems Department
Wiejska 45A, 15-351 Białystok, Poland
e-mail: palex@it.org.by

Abstract: *This paper introduce the new solution of integrated system to acoustic echo and noise reduction in order to improve the speech intelligibility in communication with used of the hands-free devices. The filter bank is proposed with minimum requirement of the band quantity to approximate psychoacoustic scale with nonuniform bands design using the first order all-pass transformation. Proposed solution fit in application of the 16-bit signal and sampling rate from 8 kHz.*

1 INTRODUCTION

The last decade shows that the noise reduction in single microphone device is still an open problem, especially concerning the hands-free devices. The well known Spectral Subtraction rules [1,2] and Minimum Mean-Square Spectrum enhancement method [3,4] are just the starting point to develop the improved and quite new approach. To satisfy the listener expectation of signal quality and eliminate the annoying musical noise and unnatural sounding of the processed speech the new solution searches the inspiration in an auditory property of the human inner ear and brain processing resulting in new class of the psychoacoustic motivated noise reduction system [5-10] and improvement of the well known methods [11-14].

Mainly all novel solutions exploit the widthband Acoustic Echo Canceller (AEC) and then the noise reduction system which exploit masking property of the human ear with calculation of the excitation pattern and masking threshold based on Fourier analysis.

In this paper the multirate system based on filter bank with the quantity of bands equal the number of Barks for assumed sampling frequency is proposed. Solution is mentioned for the 8 kHz sampling frequency in hands-free device, which is a minimum cost AEC and noise reduction system considering the number of bands for sampling frequency, and complexity reduction due to the downsampling operation, but still satisfying the intelligibility improvement of the noisy speech with speech distortion on the acceptable level.

2 NONUNIFORM FREQUENCY RESOLUTION POLYPHASE FILTER BANK

The uniform polyphase filter bank has been used in proposed solution [15-17]. Designing the bandpass filter as a complex shifted version of the prototype filter $H_0(e^{j\omega})$. and using the polyphase decomposition of order R, where $R \leqslant M$, the polyphase filter bank with analysis and synthesis stage is constructed as it is depicted in Fig. 1.

The polyphase nonuniform filter bank approximating the Bark scale can be constructed with the use of the first order bilinear mapping, where the uniform Hz scale z are map to uniform Bark scale, which is nonuniform when seen in Hz ζ. The bilinear transform for three point from both Z domain defines:

$$z = A_\alpha(\zeta) = \frac{\zeta + a}{1 + \zeta a}, \qquad a = \frac{\zeta_3 - z_3}{1 - z_3 \zeta_3}, \tag{1}$$

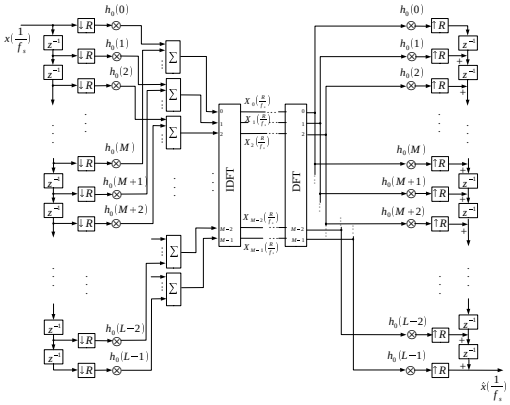

Fig. 1. Direct structure with delay chain of polyphase filter bank

where a define the freedom of mapping frequency $0 \leqslant \omega \leqslant 2\pi$ related with the point $e^{j\omega}$ on unit circle to a new location $A_a(\omega)$. From Eq. (1) it can be noticed that bilinear map has the same form as a first order all-pass filter function

$$A_a(\zeta) \equiv H_{AP}(z) = \frac{z^{-1} + a}{1 + a z^{-1}}, \tag{2}$$

with all-pass $H_{AP,a}(z)$ filter phase function:

$$\phi(\omega) = 2 \arctan\left(\frac{1-a}{1+a} \tan\frac{\omega}{2}\right), \tag{3}$$

which determines the the frequency warping. Founding on work [18] and taking the arctangent coefficient approximation of the map from Hz to Bark represented in Hz, dependent of sampling frequency, the following formula for mapping coefficient can be used:

$$a_{Bark} = 1.048 \left[\frac{2}{\pi} \arctan\left(0.07212 \frac{f_s}{1000}\right)\right]^{\frac{1}{2}} - 0.1957. \tag{4}$$

To obtain the nonuniform frequency resolution in polyphase filter bank just the replacement of the delay chain by the all-pass chain with propriety filters coefficient is needed Then the following passband filter characteristic of the transformed filter bank can be achieved:

$$H_m(e^{j\omega}) = H_m(e^{-j\phi(\omega)}) = H_0(e^{j[-\phi(\omega) - 2\pi m/M]}), \tag{5}$$

where the influence of the all-pass filter phase function is clearly visible. Fig. 2 shows the described mapping process.

The mentioned filter bank must be designed in pragmatic way to find the compromise between the overall filter bank

† This work was supported by Białystok Technical University under the project W/II/2/00

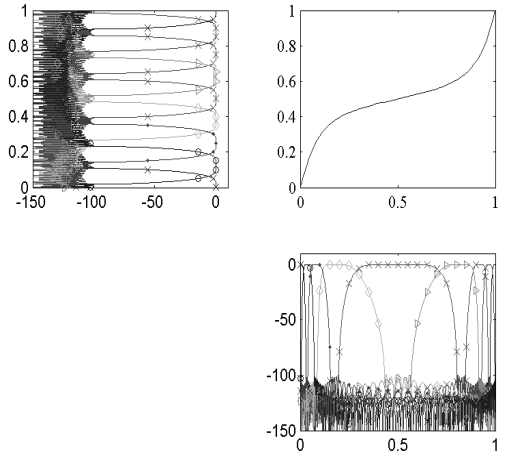

Fig. 2. All-pass transformation of polyphase filter bank for $a=-0.6$

performance and nonuniform spreading band usage. Additionally concerning the tuning the filter bank spread by modification of the mapping coefficient. Such tuning in design must be concerned with preserving filter's characteristic of synthesis filter bank successor used for phase correction, which is affected by cascade of the allpass. The design example of the Bark spreaded filter bank is presented in Fig. 3. Also its magnitude characteristic for tuning of the filter bank map coefficient and phase characteristic are depicted in Fig. 4, where the mapping spread of uniform alliasing is visible for higher frequency. Such carefully constructed filter bank can achieve only the the nearly perfect reconstruction property, which results in non-audible distortion of signal.

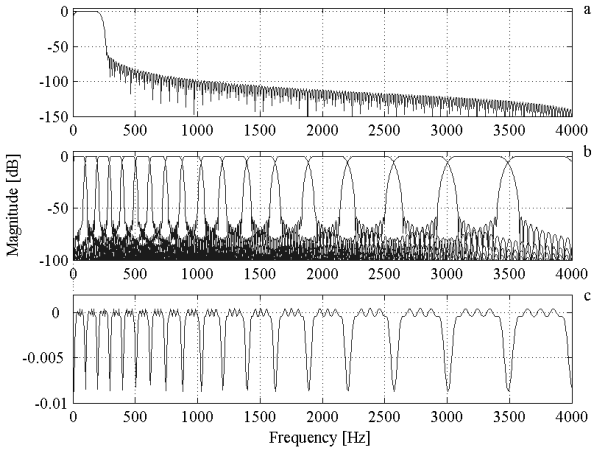

Fig. 3. Filter bank characteristic (a) prototype filer (b) analysis filter bank stage (c) sum characteristic of the analysis stage

3 ACOUSTIC ECHO CONTROL

AEC is set up with used of NLMS adaptation algorithm applied separately for each band with control mechanism operate on full band incoming microphone and loudspeaker signals. The different filter length in each band is used depending of the statistical property of the environment dedicated for use of this hands-free device.

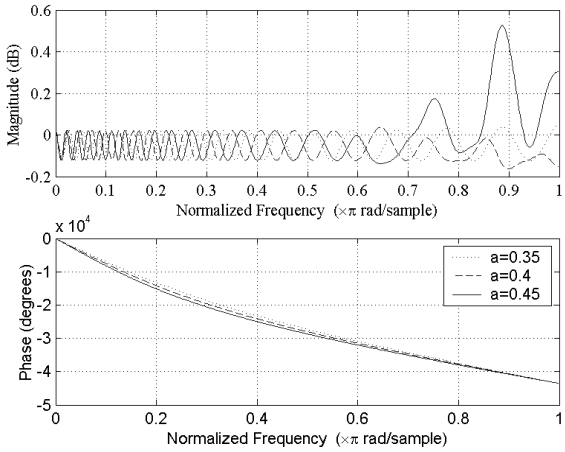

Fig. 4. Overall transfer characteristic for tuned filter bank

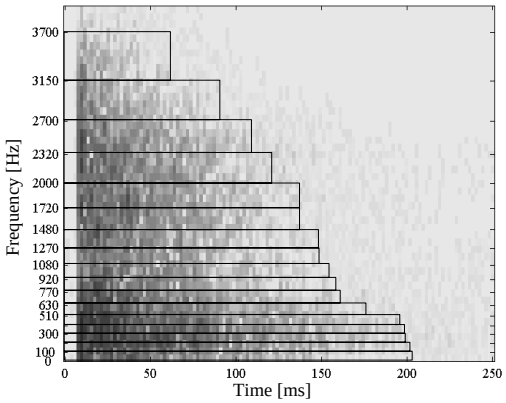

Fig. 5. Visualization of the adaptive filter length for band dependent of environment characteristic

4 AUDIBLE NOISE SUPPRESSION RULE

Proposed solution uses also the audible noise suppression rule shown in [8]. Assuming that the speech $s(k)$ is corrupted by the additive noise $n(k)$ resulting in the noised signal with the same relation of the signals in the bands:

$$y_m(k)=s_m(k)+n_m(k), \qquad 0\leqslant m\leqslant M-1, \tag{6}$$

where the subband signal $y_m(k)$ is a result of bandpass filtering of the filter $h_m(k)$ and M is the quantity of the bands. Assuming the power estimators of the signals $y_m(k), s_m(k), n_m(k)$ in the processing frame of the length W as $P_{y,m}(k), P_{s,m}(k), P_{n,m}(k)$ and the audible fragments of the band signals can be defined as $S_{y,m}(k_b)$, $S_{s,m}(k_b)$, where auditory masking threshold $T_m(k_b)$ for band m and block k_b is calculated according to the [19].

The audible noise is defined as

$$S_{n,m}(k_b)=S_{y,m}(k_b)-S_{s,m}(k_b)= \tag{7}$$

$$\begin{cases} P_{y,m}(k_b)-P_{s,m}(k_b) & \text{if } P_{y,m}(k_b)\geqslant T_m(k_b) \text{ and } P_{s,m}(k_b)\geqslant T_m(k_b) \quad (I) \\ P_{y,m}(k_b)-T_m(k_b) & \text{if } P_{y,m}(k_b)\geqslant T_m(k_b) \text{ and } P_{s,m}(k_b)<T_m(k_b) \quad (II) \\ T_m(k_b)-P_{s,m}(k_b) & \text{if } P_{y,m}(k_b)<T_m(k_b) \text{ and } P_{s,m}(k_b)\geqslant T_m(k_b) \quad (III) \\ 0, & \text{if } P_{y,m}(k_b)<T_m(k_b) \text{ and } P_{s,m}(k_b)<T_m(k_b) \quad (IV) \end{cases}$$

ant it is used in the suppression rule defined as

$$S_{n,m}(k_b) \leqslant 0, \qquad 0 \leqslant m \leqslant M-1, \tag{8}$$

where the signal weighting coefficient of the method is estimated for each processing block as

$$G_m(k) = \frac{1}{\left(\frac{a_m(k_b)}{P_{y,m}(k_b)}\right)^{v_m} + 1}, \quad \begin{matrix} k_b \leqslant k \leqslant k_b + W \text{ and }, \\ 0 \leqslant m \leqslant M-1 \end{matrix} \tag{9}$$

where the time-variable $a_m(k_b)$ and $v_m \in \Re^+ \leqslant 1$ are the parameters which determine the level of audible noise suppression defined by :

$$a_m(k_b) = [T_m(k_b) + P_{n,m}(k_b)] \left[\frac{P_{n,m}(k_b)}{T_m(k_b)}\right]^{\frac{1}{v_m}}, \tag{10}$$

For more precise evaluation of algorithm see papers [8,20].

5 SYSTEM EVALUATION

The presented earlier AEC method and weighting rule are combine in system based on nonuniform filter bank according to the Fig. 6.

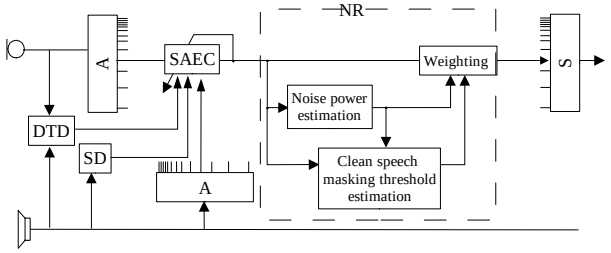

Fig. 6. In band processing schema

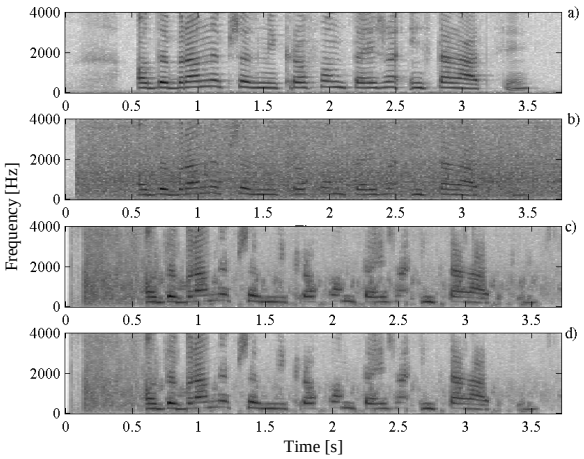

Fig. 7. Spectrograms: a) clean speech, b) noised speech at $SEGSNR_n^s = 0$ dB , c) PSS clean speech estimation, d) speech enhancect by proposed method

System setup provides only the corrupted signal, so the intermediate technique is used to estimate the masking threshold of the clean speech. Nonlinear noise estimate tracking method [21] with linear averaging for each frame, as in case of signal power spectrum is used. Also the rough speech enhancement method to get the estimate of the clean speech based on power spectral suppression rule [2] is used. Examples of system work results are depicted in Fig. 7. However, the speech enhancement quality strongly depends on noise tracking and masking threshold estimation so the careful design of estimation method must be conducted. To test the system derived in the previous sections, it was implemented in analysis/synthesis structure of polyphase all-pass transformed filter bank with band number M=36 (Bark scale), decimation ratio $R=9$, all-pass coefficient $a=-0{,}4$, frame length $W=16$, and prototype filter of the bank length of 180 coefficients. Test was conducted to 8kHz, 16-bit wav file recorded in car cabin (with engine off and on at speed 100 km/h). The signals was mixed together to provide the test signal at different signal-to-noise ratio calculated at speech activity [22-23]according to average value of i frame ratio:

$$SEGSNR_n^{s,i} = 10\log_{10}\left(\frac{\sum_{k=0}^{W-1} s^2(k+iW)}{\sum_{k=0}^{W-1} n^2(k+iW)}\right), \tag{11}$$

where i is a frame index of speech activity, W - frame length, s - speech, n -noise signals.

The objective measure of noise attenuation NA was taken and also $SEGSNR_{\hat{s}-s}^s$ difference of speech and speech distortion after processing which has the high correlation with results from auditive tests.

$$NA = 10\log_{10}\left(\frac{1}{O(K_n)}\sum_{k\in K_n}\frac{n^2(k)}{\hat{n}^2(k)}\right), \tag{12}$$

where K_n is a set of speech pauses, $\hat{n}(k)$ attenuated noise, $O(K_n)$ number of samples in set.

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

where $\hat{s}(k)$ is an enhanced speech the $SEGSNR_{\hat{s}-s}^s$ is average value over the set of speech activity from $SEGSNR_{\hat{s}-s}^{s,i}$.

According to informal listening test with various speech material, presented system offers a performance which fulfils the ITU norm and the NR method is the superior to conventional spectral subtraction rule used in it to get the rough clean speech estimate, and with minimum requirements needed provides close performance of other DFT based psychoacoustical methods, what has been shown in objectives tests depicted in Fig. 8. However the small amount of the noise must be left in enhanced signal to mask the unnaturalness of residual echo left after AEC and NR, what is depicted in Fig. 9..

CONCLUSION

Proposed combine system fulfils the ITU requirements for hands free devices for echo attenuation as well the proposed NR rule improve the speech intelligibility, what was confirmed but the objective and informal subjective listener test. The system has been proposed for real time implementation based on join architecture of DSP TMS320C31 and two FPGA XC4000 processors presented in [24].

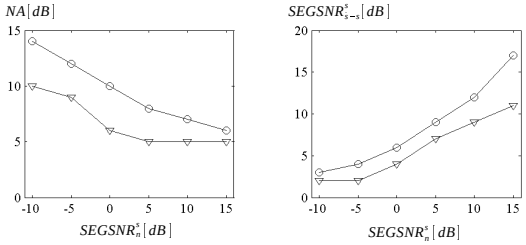

Fig. 8. Instrumental measurement data obtained from simulation of proposed NR rule (triangle) and the psychoacoustical most advanced rule presented in [9] results provided by author (circle) in his Ph.D. thesis.

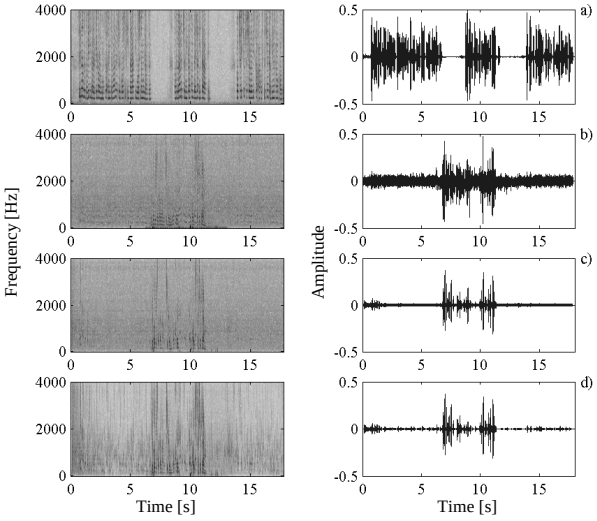

Fig. 9. Combined system output - spectrogram and corresponding time plots: a) loudspeaker signal, b) microphone signal noise degraded at $SEGSNR_n^s = 5\,\mathrm{dB}$, c) enhanced signal with preserving predefined background noise as comfort noise, d) enhanced signal without comfort noise

REFERENCES

[1] Boll S.F., *Suppression of acoustic noise in speech using spectral subtraction*, IEEE Trans.on Acoustic Speech and Signal Processing, vol. 27, no 2, 1979, pp. 113-120

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

[3] Ephraim Y., Malah D., *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*, IEEE Trans. on Acoustic Speech and Signal Processing, vol. 32, no 6, 1984, pp.1109-1121

[4] Ephraim Y., Malah D., *Speech enhancement using a minimum mean-square log-spectral amplitude estimator*, IEEE Trans. on Acoustic Speech and Signal Processing, vol. 33, no 2, 1985, pp. 443-445

[5] Jebara S.B., Benazza-Benyahia A., Khelifa A.B., *Reduction of musical noise generated by spectral subtraction by combining wavelet packet transform and Wiener filering*, in X European Signal Processing Conf. Proc., 2000, CD

[6] Gulzow T., Engelsberg A., Heute U., *Comparison of a discrete wavelet transformation and nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement*, Signal Processing, vol. 64, 1998, pp.5-19

[7] Dreiseitel P., Puder H., *Speech Enhancement For Mobile Telephony Based on Non-Uniformly Spaced Frequency Resolution*, in Proc. IX European Signal Processing Conf., 1998, pp. 965-968

[8] Tsoukalas D.E., Mourjopoulos J.N., Kokkinakis G., *Improving the intelligibillity of noisy speech using an audible noise suppresion technique*, in 5th European Conf. on Speech Communication and Technology Proc.,1997, pp. 1415-1418

[9] Gustafsson S., Jax P., Vary P., A *Novel Psychoacoustically Motivated Audio Enhancement Algorithm Preserving Background Noise Characteristic*, In Proc. International Conf. on Acoustic, Speech and Signal Processing, 1998, CD

[10]Haulick T., Linhard K., Schrogmeier P., *Residual Noise Suppression Using Psychoacoustic Criteria*, in 5th European Conf. on Speech Communication and Technology Proc., 1997, pp. 1395-1398

[11]Cappe O., *Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor*, IEEE Trans. on Acoustic Speech and Signal Processing, vol. 2, 1994, pp. 345-349

[12]Beaugeant C., Scalart P., *Noise reduction using perceptual spectral change*, in 6th European Conf. on Speech Communication and Technology Proc., 1999, pp. 2543-2546

[13]Quatieri T.F., Baxter R.A, *Noise reduction Based on spectral change*, IEEE ASSP Workshop on Application of Signal processing to Audio and Acoustics, 1997, CD

[14]Virag N., *Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System*, IEEE Trans. on Speech and Audio Processing, vol. 7, no 2. 1999, pp. 126-137

[15]Crochiere R.E., Rabiner L.R., *Multirate Digital Signal Processing*, Prentice Hall, 1983

[16]Vary P., Heute U., Hess W., *Digitale Sprachsignal-verarbeitung*, B.G. Teubner Stuttgart, 1998

[17]Akansu A.N., Haddad R.A., *Multiresolution signal Decomposition: Transforms, Subbands, and Wavelets*, Academic Press, INC., 1992

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

[19]Johnston J. D., *Transform Coding of Audio Signals Using Perceptual Noise Criteria*, IEEE Journal on Selected Areas in Communications, vol. 6, no 2, 1988, pp. 314-323

[20]K. Bielawski, A.A. Petrovsky, *Proposition of minimum bands multirate noise reduction system which exploits properties of the human auditory system and all-pass transformed filter bank*. IEEE Workshop SIGNAL PROCESSING'2001, Poznań, Poland, 2001, pp. 65-70

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

[22]Quackenbush S.R.,et al. *Objective measures of speech Quality* Printice Hall, Engelwood Clifs, New Jersey, 1988

[23]Heute U., *Objektive Sprachqualitatsmessungen: Vergleichende Ubersicht und ein neuer Ansatz*, In Proc. 8 Aachener Kolloquium Signaltheorie, VDE-Verlag Berlin, 1994, pp.21-28

[24]K. Bielawski, Al. Petrovsky, *Dynamic non-uniform filter bank constructing algorithms for reconfigurable speech processing system based on the FPGA-device and TMS320C31*, 3rd Euro-pean DSP Education and Research Conf., Paris, 2000, (CD)