Audio Engineering Society

Convention Paper

Presented at the 110th Convention
2001 May 12–15 Amsterdam, The Netherlands

Sound Equalization in a Noisy Environment

Meir Tzur (Zibulski), and Alexander A. Goldin
BIT Innovation Technologies Inc.
P.O.B. 431 Tirat HaCarmel 39032, Israel

ABSTRACT

This paper addresses the problem of equalizing an audio signal in a constantly changing noisy environment. The purpose of equalization is to provide perceptually equal loudness of sound regardless of the environmental conditions. Based on an automatic estimation of noise level and its spectral content, selective amplification of frequencies masked by noise is performed. In the case of speech signals, the result is intelligible speech regardless of the surrounding noise. For musical signals, an improved comprehension of the musical content is achieved.

1. INTRODUCTION

In many scenarios it is required to play an audio source in a noisy environment where the level and spectral content of the surrounding noise is constantly changing. For example, in an automotive environment the noise level inside a standing car is much lower than in a moving car and its spectral content is different. The audio source may be a car radio, hands free communication kit or GPS system with text-to-speech capabilities. Another example for a changing noisy environment is cellular communication outdoors.

In order to keep equal loudness of sound, while the noise level is changing, a human listener manually adjusts the volume of the audio source. Automatic solutions that change the volume of the audio source according to the noise level were proposed. However, adjusting the volume does not take into account the spectral content of noise and audio signals, i.e., no equalization is performed. For speech, this leads to either insufficient intelligibility or exaggerated volume. For music, it causes "sound coloration".

In this paper a scheme for automatic equalization and amplification of sound in a constantly changing noisy environment is presented. The scheme is based on selective amplification of frequencies masked by the surrounding noise. Factors such as noise spectral content, audio spectral content and psychoacoustics are considered by the equalization process in order to provide perceptually equal loudness of sound.

2. EQUALIZATION PRINCIPLE

The objective of the equalization procedure is to amplify frequencies masked by the surrounding noise. The equalization principle is based on the noise masking property of the human auditory system [1]. This property is used in the design of audio coders [2]. Masking occurs whenever a strong audio spectral component makes a spectral neighborhood of weaker audio signals inaudible. Roughly speaking the auditory system consists of a filter bank of overlapping band-pass filters called critical bands. The noise-masking threshold depends on the signal energy within the critical band [1]. If the noise level is below the masking threshold noise is inaudible. As the amount of noise in the environment increases, we propose to increase the energy of the signal such that in each critical band the masking threshold is above the noise level. Thus, frequencies that were originally masked by the surrounding noise are amplified and are not masked anymore.

3. DESCRIPTION OF THE EQUALIZATION SCHEME

The proposed scheme accounts for the properties of the auditory system and the spectral content of noise and signal, as described above. Furthermore, in order to present a comprehensive solution, the characteristics of the loudspeaker and the acoustic environment are considered as well.

Beside a loudspeaker, the scheme contains a reference microphone located close to the listener. Therefore, it is assumed that the signal received by the microphone represents the signal heard by the listener. In many cases, for example in case of a hands-free communication car kit or a cellular phone, such a microphone already exists.

Fig. 1 depicts the proposed scheme for sound equalization. The input from the microphone is fed into an echo canceller, which removes possible echo generated by the loudspeaker. The output of the echo canceller is divided into independent frequency bands by an analysis filter bank and noise levels are estimated for each frequency band. The sound source signal is divided into equivalent frequency bands by another analysis filter bank. Equalization gains for every frequency band are calculated based on the estimated noise levels and the sound source frequency bands. After applying the equalization gains to the corresponding frequency bands of the sound source, the signal bands are fed into an acoustic path equalizer, which compensates for the loudspeaker frequency response and the acoustic path between the loudspeaker and the microphone. The signal bands are then fed into a distortions limiter, which possibly restricts the power of every frequency band in order to limit loudspeaker distortions caused by its overload. Finally, the signal to be reproduced by the loudspeaker is reconstructed by means of a synthesis filter bank.

3.1 Echo Canceller

Since the microphone picks both the environmental noise and the output of the loudspeaker, the echo-canceling block is essential for correct noise level estimation. Efficient echo canceling algorithms can be realized in a subband structure. In such a realization the output of the microphone is fed into an analysis filter bank followed by a subband echo cancellation.

3.2 Filter Bank

An analysis filter bank is designed to split the signal into critical bands. Usually, Bark scale [3] is used for such a design. A synthesis filter bank combines the subband signals. It is desired for the filter bank to have a perfect reconstruction property, in which case if there is no processing the sound source is reproduced perfectly. The filters may be designed directly based on a frequency-warped transform [4] or indirectly based on a DFT filter bank [5].

3.3 Equalization Gains Calculation

Calculation of the equalization gains is based on the noise masking property of the human auditory system. A single gain is calculated for each critical band such that the noise-masking threshold is above the corresponding estimated noise level. A procedure for calculating a noise-masking threshold for the case of audio coders is presented in [2]. The threshold calculated therein accounts for the transitory spectral structure of the audio signal. In our case the procedure is modified in order to calculate a slowly varying threshold. The noise-masking threshold is a function of the signal level received by the microphone, i.e., it is a function of the sound source signal level and the equalization gains. Thus, in order to find the equalization gains an equation should be solved, where the equalization gains are expressed as a function of themselves. A simplified scheme can be used for calculating the equalization gains G_i as follows:

$$G_i = \begin{cases} 1, & N_i \le T_i \\ f\left(N_i/T_i\right), & N_i > T_i \end{cases},$$

where T_i is the noise masking threshold, N_i is the estimated noise level and $f(\cdot)$ is a monotonically increasing function such that $f(1)=1$, found experimentally.

An external volume input, controlled by the user, can change the gains. In this way, the user can change the loudness of the output signal while the system operates.

The "tonal balance" of the signal is also considered while calculating the equalization gains. In order not to modify the "tonal balance" sounds should be reproduced at the same level as the original [1]. This phenomenon is explained by the so-called equal-loudness contours. A typical way to minimize tonal distortions due to this phenomenon is by boosting low and high frequencies for low level sounds and boosting middle range frequencies for high level sounds. Such a boost should be made carefully in order not to harm the noise-masking threshold. Therefore, it is performed only if it does not reduce the noise-masking threshold below the noise level. Note that this function cannot be done accurately, since the true level of the original sound is not known. However, the system can be calibrated for nominal levels of sound, i.e., no boosting for a level of 70-80 dB SPL. The signal that produces such a sound level should be measured during an initial calibration procedure.

3.4 Noise Level Estimation

A slowly changing noise level is estimated for each critical band, assuming that the noise is a stationary process. In case the sound source is a pure speech signal, the echo canceller block can be omitted. The noise levels, in such a case, are estimated during non-speech sections determined by a voice activity detector. If however, the sound source is a musical signal, or a noisy speech, the echo canceller block is essential.

3.5 Acoustic Path Equalizer

This block is designed to account for the properties of the loudspeaker and the acoustic environment, i.e., equalization is done in order to flatten the joint frequency response of the loudspeaker and the acoustic environment. An initial calibration procedure is

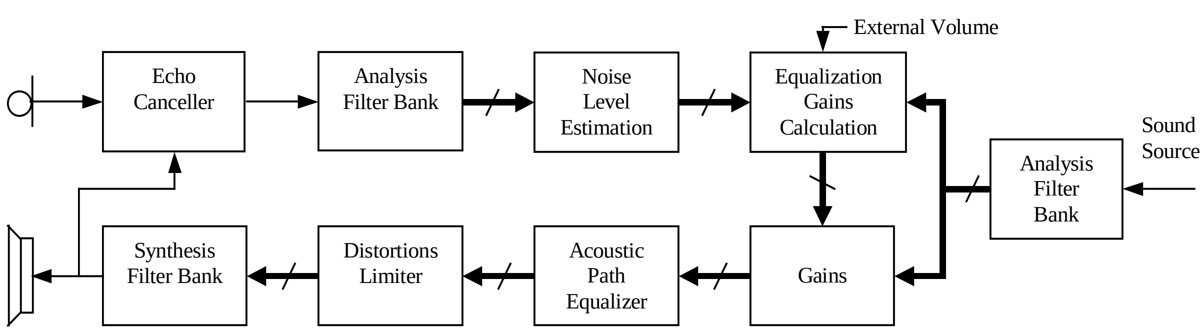

Fig. 1. A scheme of a sound equalizer in a noisy environment

performed in order to measure the frequency response and to design the equalization gains.

3.6 Distortions Limiter

The output power at each frequency is limited in order to minimize harmonic distortions of the loudspeaker. The limit values are measured during an initial calibration procedure. Note that the performance of the echo canceller degrades if the response of the loudspeaker is non-linear. Therefore, minimization of harmonic distortions is important for good performance of the echo canceller.

4. HANDS-FREE CAR KIT IMPLEMENTATION

This section describes an implementation of the equalization scheme for a hands-free car kit developed at BIT Innovation Technologies. The main goal of the system is to improve speech intelligibility in a noisy environment. The sound source is a speech signal sampled at 11025 Hz.

Fig. 2 depicts the implemented scheme. For the scheme to be efficient echo canceling is realized in a subband structure. An analysis DFT filter bank is utilized for the subband decomposition. The sub-bands do not correspond to critical bands, therefore, the concept of critical bands is realized indirectly before processing by the noise level estimator. Similarly, critical bands analysis is performed on the sound source signal analyzed by the DFT filter bank. Since the main goal of the system is speech ineligibility improvement, the acoustic path equalizer block is omitted for simplicity reasons. Furthermore, loudness equalization in the equalization gains calculation block is omitted for the same reason.

4.1 DFT Filter Bank

The advantage of using an analysis/synthesis DFT filter bank is that the filter bank can be designed to have the perfect reconstruction property and can be realized efficiently [5], [6]. An analysis DFT filter bank is realized using windowing and FFT methods and a synthesis DFT filter bank is realized by IFFT and an overlap-add methods. Basically, during the analysis procedure the input signal is divided into overlapping time frames, each frame is multiplied by a window function followed by an FFT. The output of the FFT is usually regarded as the short-time spectrum of the signal.

4.2 Echo Canceller

Echo canceling can be realized either in full-band or subband structures. As shown in Fig. 2, echo canceling is implemented in a subband structure, since the subband structure offers a reduced computational complexity and a better speed of convergence [7]. The input to the synthesis filter bank is used as a reference signal to the echo canceller at each subband. Ideally, the reference signal should come form an analysis filter bank located at the output to the loudspeaker. However, if the filter bank is designed properly there is no difference between the two signals. Description of the design algorithm is beyond the scope of this paper.

4.3 Critical Bands Analysis

Critical band analysis is performed on the short time spectrum of the signal or noise by adding up energies in each critical band. A Bark scale based critical band filter bank is described in [3].

4.4 Equalization Gains Calculation

As previously described, equalization gains are calculated for each critical band in order to maintain the noise-masking threshold above the estimated noise level. Since the number of frequency bins, which is determined by the FFT length, is larger than the number of critical bands, and the frequency resolution of the FFT is finite, gains for each frequency bin of the analysis DFT filter bank are calculated by interpolating the critical bands equalization gains. These gains are applied to the corresponding frequency bins of the analysis filter bank.

4.5 Noise Level Estimation

Although the echo canceller reduces the amount of echo exists in the signal coming from the microphone, there always remains a residual echo. Therefore, noise estimation is performed during non-speech periods determined by a voice activity detector, which is part of the noise level estimation block.

5. CONCLUSIONS

The presented scheme addresses the problem of automatic equalization and amplification of sound in a constantly changing noisy environment. The proposed solution amplifies sound frequencies that are masked by the surrounding noise. Signal spectral content, noise spectral content, auditory system properties and characteristics of the loudspeaker and the acoustic path are taken into account in a comprehensive solution.

Specific implementation for a hands-free car kit is described where the sound signal is a speech signal. In this case the main goal of the system is intelligibility improvement.

Calculation of the noise-masking threshold is based on the procedure presented in [2] which is used in the case of audio coders. However, it is not clear that such a procedure is suitable in the case of speech signals, since a certain amount of noise does not degrade speech intelligibility. Therefore, calculation of a noise-masking threshold in the case of speech signals is a subject for further research.

REFERENCES

[1] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. London: Academic Press, 1997.
[2] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE J. Select. Areas Commun., vol. 6, pp. 314-323, 1988.
[3] E. Zwicker and E. Terhardt, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. , vol. 68, no. 5, pp. 1523-1525, 1980.

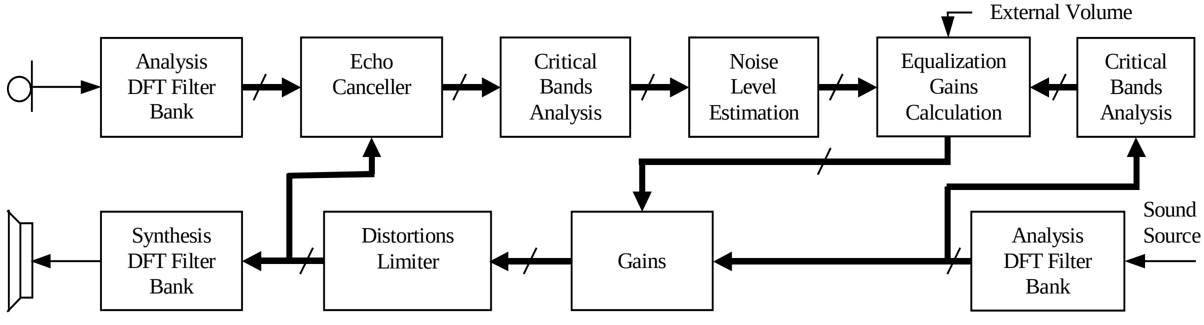

Fig. 2. Hands-free car kit implementation of an equalization scheme

[4] A. Härmä, M. Karjalainen, L. Savioja, V.Välimäki, U. K. Laine, and J.Huopaniemi, "Frequency-warped signal processing for audio applications", J. Audio Eng. Soc, vol. 48, pp. 1011-1031, 2000.

[5] R. E. Crochiere, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[6] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[7] C. Breining, P. Dreiscitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control. An application of very-high-order adaptive filters", IEEE Signal Processing Magazine, vol. 16, pp. 42-69, 1999.