PERCEPTUAL HEADPHONE EQUALIZATION FOR MITIGATION OF AMBIENT NOISE

Jussi Rämö and Vesa Välimäki

Aalto University School of Electrical Engineering
Dept. of Signal Processing and Acoustics
P.O. Box 13000, FI-00076 AALTO, Finland

Miikka Tikander

Nokia Corporation
Keilalahdentie 2–4, P.O. Box 226
FI-00045 NOKIA GROUP, Finland

ABSTRACT

An adaptive perceptual equalizer for headphones is introduced. It estimates the effect of auditory masking while considering the characteristics of the headphones, ambient noise, and music. The system utilizes a psychoacoustic masking model to estimate the level to which the music should be raised to have the same perceived tonal balance in noise as it has in a quiet environment. Prototype testing showed that the most important task is to make the music audible in each Bark band. The compensation of the partial masking further improves the perceived sound quality. The system uses a microphone of a headset to capture the ambient noise. The equalization is implemented using a high-order graphical equalizer that does not require subband decomposition of the music signal. The proposed equalizer also retains reasonable SPL levels: in an example case, the maximum gain in one Bark band was 11 dB while the overall SPL increase was only 2.5 dB.

***Index Terms*—**Acoustic noise, acoustic signal processing, audio systems, music, psychoacoustics

1. INTRODUCTION

Listening to music through headphones takes place mostly in noisy environments due to the vast success of portable music players and smartphones. Gartner, Inc. reported that worldwide sales of mobile phones reached almost 428 million units in the third quarter of 2012. The share of smartphone sales was 40 percent of total mobile phone sales, increasing 47 percent from the third quarter of 2011 [1].

Loud background noise is known to mask parts of the music signal and thus changes the perceived timbre of the music [2, 3]. By definition, auditory masking occurs when a sound affects the perceived loudness of another sound. The masking threshold represents the level under which a desired signal becomes inaudible, whereas partial masking only reduces the loudness of the desired signal [4].

The authors have previously presented a perceptual frequency response simulator which utilized elementary auditory masking models previously used in audio coding applications as well as the measured isolation capabilities of different headphones to estimate the auditory masking phenomenon when using headphones in a noisy environment [5]. The aim of this article is to utilize a low-complexity auditory masking model, although there are more recent masking models presented, e.g., by van de Par *et al.* [6] and Jepsen *et al.* [7], in order to design a real-time adaptive psychoacoustic equalizer for headphones, which takes the characteristics of the background noise and music into account. Ideally, the proposed equalizer compensates the masking effect (both complete and partial masking) caused by the background noise.

This paper is organized as follows. Section 2 describes the masking estimation algorithm for the ambient noise. Section 3 presents the proposed perceptual equalizer. Section 4 focuses on the results and Section 5 concludes the paper. Furthermore, Section 6 discusses the relation to prior research.

2. MASKING ESTIMATION

Figure 1 shows the block diagram of the masking estimation (top part). The threshold of masking is calculated as follows [5, 8]. The noise (masker) signal is first filtered with a headphone isolation curve $H_{\text{hi}}(z)$ to simulate the noise that is transferred through the headphone into the ear canal. Then the signal is windowed and the short-time Fourier transform (STFT) is calculated in order to derive the power spectrum $P_m(k)$ for the m^{th} noise signal frame. The frequency scale is then mapped onto the Bark scale with the approximation [4]

$$\nu = 13 \arctan\left(\frac{0.76f}{\text{kHz}}\right) + 3.5 \arctan\left(\frac{f}{7.5\text{kHz}}\right)^2, \qquad (1)$$

where f is the frequency in Hertz and ν is the mapped frequency in Bark units. The energy in each critical band is the partial sum

$$Z_m(\nu) = \frac{1}{N_\nu} \sum_{k=B_l(\nu)}^{B_h(\nu)} P_m(k), \qquad \nu = 1, 2, ..., N_c, \qquad (2)$$

where $B_{\text{l}}(\nu)$ is the lower and $B_{\text{h}}(\nu)$ is the upper boundary of the critical band ν, N_ν is the number of data points in each critical band ν, and N_c is the number of critical bands. Furthermore, the energy of the music signal is calculated in the same way.

After that, the spreading of the masking throughout the adjacent critical bands is approximated with a two-slope spreading function [9]

$$\begin{aligned} 10\log_{10}\left[B(\Delta\nu, L_{\text{M}})\right] = \\ \left[-27 + 0.37\max\left\{L_{\text{M}} - 40, 0\right\}\theta(\Delta\nu)\right]|\Delta\nu|, \end{aligned} \qquad (3)$$

where $\Delta\nu = \nu(f_{\text{maskee}}) - \nu(f_{\text{masker}})$, L_{M} is the SPL of the masker, and $\theta(\Delta\nu)$ is a step function equal to zero for negative values of $\Delta\nu$ and equal to one for positive values of $\Delta\nu$. The individual masking curves with intensities B_ν are then added using a summation formula

$$S_{P,m} = \left(\sum_{\nu=1}^{N_c} B_\nu^\alpha\right)^{\frac{1}{\alpha}}, \qquad 1 \le \alpha \le \infty, \qquad (4)$$

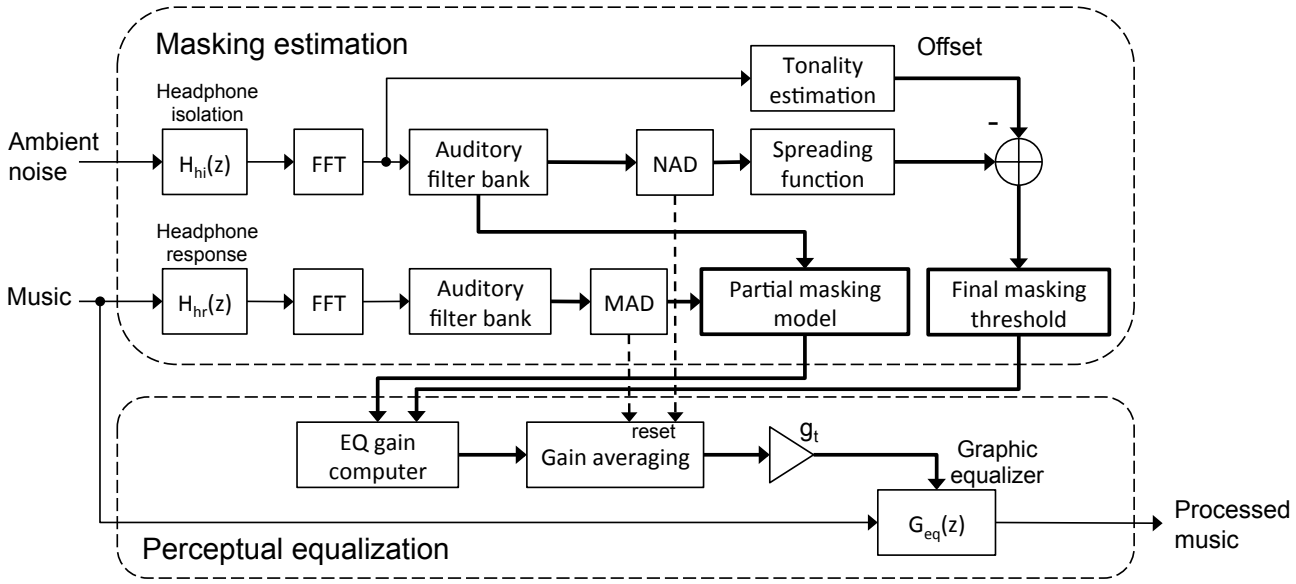

Fig. 1. Block diagram of masking estimation and perceptual equalization processes. Thick paths contain a multichannel signal in Bark bands.

where $S_{P,m}$ is the overall spread masking curve, which represents the intensity of the masking curve resulting from the combination of N_c individual masking curves and α defines the method according to which the curves are combined. Lufti [10] has suggested that α should be approximately 0.33, which is the value used in this work.

The tonality of the masker affects the degree of the masking effect. According to Johnston [8], the two extremes are a tone masking a noise and a noise masking a tone. The offset for a tone-like masker is $14.5+\nu$ dB and for a noise-like masker 5.5 dB. Spectral flatness is used to estimate the tonal characteristics of the masker. The spectral flatness V_m is defined as the ratio of the geometric and arithmetic mean of the power spectrum [8]:

$$V_m = 10\log_{10}\frac{\left[\prod_{k=0}^{N-1} P_m(k)\right]^{\frac{1}{N}}}{\frac{1}{N}\sum_{k=0}^{N-1} P_m(k)}, \tag{5}$$

The tonality factor α_m is defined as [8]

$$\alpha_m = \min\Big(\frac{V_m}{-60\text{dB}}, 1\Big), \tag{6}$$

which is used to geometrically weight the offsets for noise and tone to form the masking energy offset $U_m(\nu)$ for each critical band [8]:

$$U_m(\nu) = \alpha_m(14.5+\nu) + (1-\alpha_m)5.5. \tag{7}$$

The energy offset is then subtracted from the spread masking threshold $S_{P,m}$ to estimate the raw masking threshold R_m [8]:

$$R_m(\nu) = 10^{\log_{10}(S_{P,m}(\nu)) - \frac{U_m(\nu)}{10}}. \tag{8}$$

The partial masking model for complex musical sounds that is utilized in the masking estimation is described in [5], where masked loudness-matching functions were constructed for complex test tones, which had realistic envelopes and harmonic structures [11] .

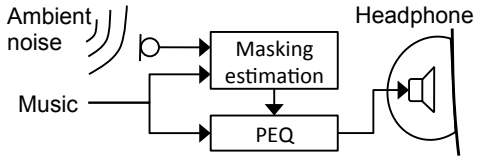

Fig. 2. Block diagram of the system, where PEQ is the proposed perceptual equalizer.

Both the masking threshold and partial masking estimations take the level of the music and the frequency response of the headphone $H_{\text{hr}}(z)$ into account. In practice, this implies that the sensitivity of the microphones and the headphones must be known.

3. PERCEPTUAL EQUALIZER

The perceptual equalizer (PEQ) proposed in this paper requires a pair of headphones with a mono microphone (see Figure 2). Fortunately, most modern headsets have an in-wire microphone built in which is suitable for the perceptual equalizer.

Ideally, the level of the music should be raised to a level where the noise has no influence on the listening experience, i.e., the music has the same perceived tonal balance in a noisy environment as in a quiet one. However, in practice the most important task is to make the music at least audible at all frequency bands. After that, it is advantageous to boost the partially masked components of the music as well.

Figures 1 and 2 show the block diagrams of the whole system, where the PEQ utilizes the masking threshold and partial masking estimations calculated with the method described in Section 2 and in Figure 1. The estimation of the frequency response $H_{\text{hr}}(z)$ and the isolation $H_{\text{hi}}(z)$ of the headphones has been conducted using a dummy head (head and torso simulator), and the filters are imple-

mented using FIR filters of order 100 and 200, respectively. The perceptual equalizer consists of the EQ gain computer function, which takes the masking information as an input; the gain averaging function, which smooths the gain variation; and the high-order graphic equalizer, which ultimately applies the equalization to the music signal, as shown in Figure 1.

3.1. Gain Calculation

After the entire masking information from one frame (the frame length is 1 second) is estimated, the gain computer estimates in which bands the energy of the music is below the masking threshold. It is possible to set a target value with respect to the masking threshold to which the masked sounds are boosted. Informal listening tests showed that already when the inaudible sounds are boosted 2 dB above the masking threshold, the emerging of the sounds that were masked clearly improves the perceived frequency response of the music.

Furthermore, the algorithm uses the partial masking information to estimate how much the music is being masked in each Bark band and uses these values to boost the partially masked components. The algorithm also checks how much headroom is left in the music signal and limits the boosting to that so as not to distort the music signal.

Moreover, controlling the amount of the partial masking effect included is possible in the proposed equalizer. This is implemented with an adjustable gain ($0 \leq g_p \leq 1$) that is applied to the calculated partial masking values. For example, for a g_p value of 0, the partially masked components are not boosted at all.

3.2. Gain Averaging

It was discovered that a large change in the gain value from one frame to another results in an audible pumping of the sound. Thus, a gain averaging function was implemented in order to limit the gain variation. The size of the averaging table N_{at} can be adjusted. The longer the average is, the smoother and slower the algorithm becomes. In other words, the size of the average table is a compromise between the adaptation speed of the algorithm and the sound quality of the equalized music.

Informal listening tests in this study showed that the averaging time can be quite long ($\geq$10 seconds), because human hearing adapts rather slowly, and it is insensitive to short noise bursts, especially when listening to musical content.

However, the slow adaptation speed becomes an issue when the noise in a noisy environment ends abruptly, such as when the user enters a quiet environment from a noisy one, e.g., from the street to indoors. Hence, a noise activity detection (NAD) was implemented, which resets the averaging table if the mean energy of the critical bands drops below a set threshold value λ_{noise}

$$\frac{1}{N_c}\sum_{\nu=1}^{N_c} 10\log_{10}\big(Z_m(\nu)\big) < \lambda_{\text{noise}}. \tag{9}$$

This way the equalizer is also reset, and the algorithm starts to build a new averaging table with new gain values for the changed environment.

Furthermore, a music activity detector (MAD) was implemented the same way as NAD, with an independent threshold λ_{music} to avoid boosting when there is not enough content in the music signal. Otherwise, silent parts in music would cause the equalizer to boost the signal because the music signal is below the masking threshold of the ambient noise. This would result in an exaggerated boost, e.g., between music tracks. Moreover, when the averaging table is reset between two songs, the algorithm starts to build a new set of gain values for the new song. People nowadays often listen to music from playlists, and hence consecutive songs may be dissimilar and therefore require different equalization.

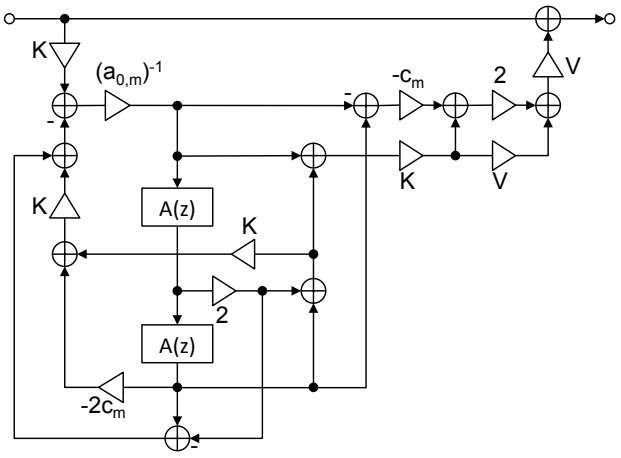

Fig. 3. Block diagram of a fourth-order section of the graphic equalizer [5, 13].

The output of the gain averaging block contains the target gains for the equalizer. A total effect depth control ($0 \leq g_{\text{t}} \leq 1$) was added to the signal chain so as to be able to adjust the effect of the equalizer (see Figure 1). When $g_{\text{t}} = 0$, the equalizer is turned off and when $g_{\text{t}} = 1$, it operates according to the implemented psychoacoustic models.

3.3. Graphic Equalizer Design

The graphic equalizer used in the proposed perceptual equalizer is based on the designs presented by Orfanidis [12] and Holters and Zölzer [13]. With this design the gain in one band is almost completely independent of the gain in adjacent bands. Figure 3 shows the block diagram of one fourth-order section of the graphic equalizer. The blocks $A(z)$ contain a second-order allpass filter having the transfer function

$$A(z) = \frac{a_2 + a_1 z^{-1} + z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{10}$$

where $a_2 = 0$ and $a_1 = \cos(\Omega_{\text{M}})$. The parameter Ω_{M} is the optimized center frequency of the equalizer

$$\Omega_{\text{M}}(\nu) = 2\arctan\left(\sqrt{\tan\left(\frac{\Omega_{\text{U}}(\nu)}{2}\right)\tan\left(\frac{\Omega_{\text{L}}(\nu)}{2}\right)}\right), \tag{11}$$

where $\Omega_{\text{L}}(\nu)$ and $\Omega_{\text{U}}(\nu)$ are the normalized lower and upper cut-off frequencies of the ν^{th} Bark band, respectively. Furthermore, the used orders were 16 and 12 for the first and second Bark bands, respectively, and 8 for all the others.

The main advantage of the high-order graphic equalizer is that it does not require a filter bank to first decompose the music signal into frequency bands and then obtain the processed time-domain signal by using the overlap-add method. With the current design, the

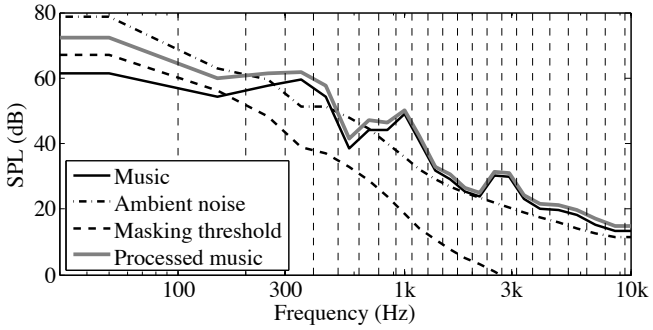

Fig. 4. Spectrum of the music and ambient noise signal in Bark bands, and the estimated masking threshold for the noise.

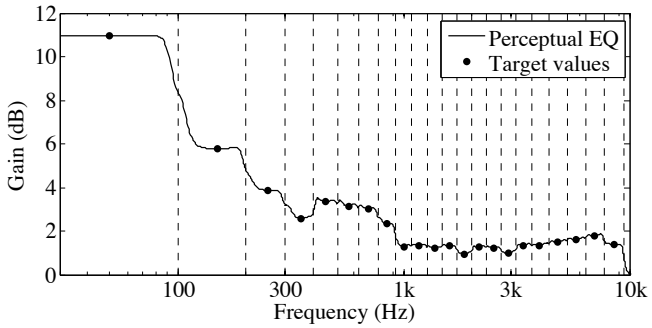

Fig. 5. Magnitude response of the perceptual equalizer closely matches the target gain values in Bark bands.

unprocessed music signal is just filtered with the graphic equalizer, whose parameters are adjusted based on the psychoacoustic masking estimation.

4. RESULTS

The proposed perceptual headphone equalizer was implemented in real time with Matlab [14] and Playrec [15]. Figure 4 shows the analysis of a one-second frame, where the spectrum of the music (solid black line) and ambience noise signal (dash-dotted line), as well as the estimated masking threshold (dashed line), which is calculated using the masking estimation (see Section 2 and Figure 1), are illustrated. Furthermore, the thick gray line shows the spectrum of the processed music. The used ambient noise was simulated bus noise: a white noise sequence was filtered with a linear predictive model of a bus noise recording.

Figure 5 shows the corresponding frequency response of the perceptual equalizer (solid line) and the target gain values (dots), averaged over ten frames (i.e., 10 seconds), for each Bark band. The maximum boost is limited to the available headroom in the music signal in order not to distort the music signal. The available headroom in this particular case was 13 dB. Furthermore, the signals from the last of the ten frames is shown in Figure 4.

One of the additional advances of the PEQ is that the SPL stays at a reasonable level, at least when compared to the typical volume boost. For example, the SPL increase that the PEQ introduces in the music signal in Figure 4 is 2.5 dB, whereas if the same low-frequency boost is acquired with a volume adjustment, the SPL is increased by 11 dB. Unfortunately, people often use the volume control to compensate for the masked parts of the music.

As can be seen in Figure 5, the need of the equalizer is concentrated at low frequencies ($<$ 1 kHz). This is often the case, since the isolation of headphones is usually poor below 1 kHz and the background noise generally has pronounced low-frequency content. Based on informal listening tests, it was observed that the proposed equalizer gain could be restricted to the first nine Bark bands, i.e., 20–1080 Hz, to still acquire effective results. Furthermore, when the gain of the equalizer is varied above this frequency range, the result can quite easily be an audible pumping sound, which deteriorates the listening experience. The limitation of the frequency range also greatly reduces the computational workload of the algorithm.

Interestingly, the overall shape of the particular PEQ in Figure 5 resembles the shape of a low-complexity loudness compensation used in early hi-fi systems [16]. However, the proposed adaptive system utilizes information about the content and SPL of the music and noise signal as well as the frequency response and isolation capabilities of the headphone, which allows the PEQ to adapt in numerous listening situations.

5. CONCLUSIONS

This article introduced an adaptive perceptual equalizer for headphones, which is based on auditory masking models. The system estimates the noise and music levels in the ear canal by taking the characteristics of the headphones into account. The equalization is implemented with a high-order graphic equalizer, which has almost completely independent gains in each Bark band. The proposed PEQ retains a tolerable SPL compared to the typical situation where people compensate the masking of the music by turning up the volume. In the example case, the PEQ increased the SPL of the music by less than 3 dB while the maximum boost in a single subband was 11 dB, which corresponds to the required SPL increase without the PEQ. Furthermore, the system adapts to different types of music, headphones, and listening environments. Sound examples are available online at http://www.acoustics.hut.fi/go/icassp13-peq.

6. RELATION TO PRIOR RESEARCH

In audio processing, models of auditory masking have previously been used in perceptual coding of music [8, 17, 18, 19], in perceptual evaluation of audio signals [20, 21, 22], and in the reduction of audio content for analysis and recognition [23, 24]. There have appeared efforts to estimate and compensate the masking caused by background noise in automotive audio [25, 26], in other noisy environments [27], and, in one previous study, in headphone listening in a train-cabin noise at a 70-dB SPL [28]. This work expands the latter study by providing a general framework in which noise is registered with a calibrated microphone, both the noise and music signal are analyzed, and, based on the estimated simultaneous masking, the music signal is corrected using a high-order adaptive equalizer to cancel the masking and partial masking effect for headphones.

7. REFERENCES

[1] Gartner, "Gartner says worldwide sales of mobile phones declined 3 percent in third quarter of 2012; smartphone sales increased 47 percent," http://www.gartner.com/it/page.jsp?id=2237315, November 2012.

[2] B. R. Glasberg and B. C. J. Moore, "Development and evaluation of a model for prediction the audibility of time-varying sounds in the presence of background sounds," *J. Audio Eng. Soc.*, vol. 53, no. 10, pp. 906–918, October 2005.

[3] D. Isherwood and V.-V. Mattila, "Objective estimates of partial masking thresholds for mobile terminal alert tones," in *AES 115th Convention*, New York, USA, October 2003.

[4] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, New York, 1990.

[5] J. Rämö, V. Välimäki, M. Alanko, and M. Tikander, "Perceptual frequency response simulator for music in noisy environments," in *Proc. AES 45th International Conference*, Helsinki, Finland, March 2012.

[6] S. van de Par, A. Kohlrausch, G. Charestan, and R. Heusdens, "A new psychoacoustical masking model for audio coding applications," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. II-1805-II-1808, 2002.

[7] M. L. Jepsen, S. D. Ewert, and T. Dau, "A computational model of human auditory signal processing and perception," *J. Acoust. Soc. Am.*, vol. 124, no. 1, pp. 422–438, 2008.

[8] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal of Selected Areas in Communications*, vol. 6, no. 2, pp. 314–323, February 1988.

[9] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*, Kluwer, 2003.

[10] R. A. Lufti, "Additivity of simultaneous masking," *J. Acoust. Soc. Am.*, vol. 73, no. 1, pp. 262–267, January 1983.

[11] H. Gockel and B. C. J. Moore, "Asymmetry of masking between complex tones and noise: Partial loudness," *J. Acoust. Soc. Am.*, vol. 114, no. 1, pp. 349–360, July 2003.

[12] S. J. Orfanidis, "High-order digital parametric equalizer design," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1026–1046, November 2005.

[13] M. Holters and U. Zölzer, "Graphic equalizer design using higher-order recursive filters," in *Proc. Int. Conf. Digital Audio Effects (DAFx-06)*, pp. 37-40, September 2006.

[14] MATLAB, *version 7.14.0.739 R2012a*, The MathWorks Inc., Natick, Massachusetts, http://www.mathworks.se.

[15] R. Humphrey, "Playrec, multi-channel Matlab audio, http://www.playrec.co.uk," 2006–2008.

[16] T. Holman and F. Kampmann, "Loudness compensation: Use and abuse," *J. Audio Eng. Soc.*, vol. 26, no. 7/8, pp. 526–536, July/August 1978.

[17] K. Brandenburg and M. Bosi, "Overview of MPEG audio: Current and future standards for low-bit-rate audio coding," *J. Audio Eng. Soc.*, vol. 45, no. 1/2, pp. 4–21, February 1997.

[18] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–515, April 2000.

[19] T. Hirvonen and A. Mouchtaris, "Top-down strategies in parameter selection of sinusoidal modeling of audio," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 273-276, March 2010.

[20] T. Sporer, U. Gbur, J. Herre, and R. Kapust, "Evaluating a measurement system," *J. Audio Eng. Soc.*, vol. 43, no. 5, pp. 353–363, 1995.

[21] H.-M. Lehtonen, J. Pekonen, and V. Välimäki, "Audibility of aliasing distortion in sawtooth signals and its implications for oscillator algorithm design," *J. Acoust. Soc. Am.*, vol. 132, no. 4, pp. 2721–2733, October 2012.

[22] C. H. Taal, R. C. Hendriks, and R. Heusdens, "A low-complexity spectro-temporal distortion measure for audio processing applications," *IEEE Trans. Audio Speech, and Language Processing*, vol. 20, no. 5, pp. 1553–1564, July 2012.

[23] P. Balazs, B. Laback, G. Eckel, and W. A. Deutsch, "Time–frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking," *IEEE Trans. Audio Speech, and Language Processing*, vol. 18, no. 1, pp. 34–49, January 2010.

[24] T. May, S. van de Par, and A. Kohlrausch, "Noise-robust speaker recognition combining missing data techniques and universal background modeling," *IEEE Trans. Audio Speech, and Language Processing*, vol. 20, no. 1, pp. 108–121, January 2012.

[25] T. E. Miller and J. Barish, "Optimizing sound for listening in the presence of road noise," in *AES 95th Convention*, New York, USA, October 1993.

[26] M. Christoph, "Noise dependent equalization control," in *AES 48th International Conference*, Munich, Germany, September 2012.

[27] D. Kleis and N. Öias, "Optimum transmission of speech and music into noisy environment," in *AES 59th Convention*, Hamburg, Germany, February 1978.

[28] J. T. Pedersen and L. B. Vestergaard, "Optimal music reproduction in noisy environments," M.S. thesis, Aalborg University, Department of Electronic Systems, 2010, http://projekter.aau.dk/projekter/files/32315329/report.pdf.