
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Author(s): J. Rämö, V. Välimäki, M. Alanko, and M. Tikander

Title: Perceptual Frequency Response Simulator for Music in Noisy Environments

Year: 2012

Version: Final published version

Please cite the original version:

J. Rämö, V. Välimäki, M. Alanko, and M. Tikander. Perceptual Frequency Response Simulator for Music in Noisy Environments. In Proc. AES 45th Int. Conf., 10 pages, Helsinki, Finland, March 2012.

Note: © 2012 Audio Engineering Society (AES)

Reprinted with permission.

Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society (www.aes.org).

This publication is included in the electronic version of the article dissertation:
Rämö, Jussi. Equalization Techniques for Headphone Listening.
Aalto University publication series DOCTORAL DISSERTATIONS, 147/2014.

All material supplied via Aaltodoc is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Perceptual Frequency Response Simulator for Music in Noisy Environments

Jussi Rämö¹, Vesa Välimäki¹, Mikko Alanko¹, and Miikka Tikander²

¹*Aalto University, Department of Signal Processing and Acoustics, P.O. Box 13000, FI-00076 AALTO, Espoo, Finland*

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

Correspondence should be addressed to Jussi Rämö (jussi.ramo@aalto.fi)

ABSTRACT

A perceptual simulator for music in noisy environments is described. The listening environments where people use their headphones have changed from quiet indoor environments to more noisy outdoor environments. The perceptual simulator utilizes auditory masking models and the isolation capabilities of different headphones to simulate the auditory masking phenomenon. A real-time demonstrator using Matlab and Playrec was implemented, which illustrates how the background noise alters the timbre of the music. It can be used with either headphones or loudspeakers. Informal listening tests showed that the simulator is operating correctly in most cases. However, when there is great amount of energy at the lowest frequencies of the background noise, the masking threshold is predicted to be too high.

1. INTRODUCTION

Over the years headphone listening has become more and more mobile, hence listening environments have also changed dramatically. This sets novel challenges to headphone listening, especially in noisy environments, including public transportation, restaurants, and places with heavy traffic. The main problem when listening to music in noise is that the ambient noise masks parts of the music signal. In other words, the noise affects the perceived timbral balance of the music signal.

A masking threshold refers to a level under which the music signal is inaudible, whereas partial masking reduces the loudness of the music but does not mask it completely [1]. The masking effect is often analyzed in critical bands (Bark bands) [2], i.e., the masking threshold and partial masking are defined separately to each critical band. There are known models for predicting the masking threshold and partial masking, e.g., [3]–[5], however, due to the complex signals used in the proposed simulator, these models could not be utilized directly.

This article introduces a real-time demonstrator, which simulates the perceived audio performance of headphones in different noisy listening situations. The perceived frequency response is achieved by applying a real-time equalization to a music signal. Furthermore, the demonstrator operates with different background noises

and adapts according to the noise isolation capabilities of different headphones.

This paper is organized as follows. Section 2 describes the measurements needed to simulate the isolation properties of the headphones. Section 3 presents the auditory masking models. Section 4 focuses of the implementation of the simulator and the real-time demonstrator. Section 5 discusses the evaluation listening tests and results, and Section 6 concludes the paper.

2. HEADPHONE MEASUREMENTS

Measurements were conducted to derive the ambient noise isolation capability of various headphone types. Six different headphones were measured: two in-ear headphones (IE1, IE2), two intra-concha headphones (IC1, IC2), and two closed-back supra-aural headphones (SA1, SA2). Furthermore, SA2 headphones had also active noise control (ACN) option, which is denoted as SA2-ANC.

The headphone measurements were conducted in an acoustically treated listening room. A diffuse sound field was created by reproducing pink noise with four Genelec loudspeakers and one subwoofer. The measurement equipment included Matlab and Playrec software accompanied with a MOTU UltraLite mk3 audio interface. Playback of the multi-channel pink noise signals was re-

alized with Audacity software and reproduced with the audio interface.

The isolation curve was measured so that first the noise from the Genelec loudspeakers was measured with the ear microphone of a Brüel&Kjær HATS (model 4128C type 3.3 ear simulator) mannequin torso at the drum reference point (DRP) without headphones. Then the headphones were installed on the HATS and the noise from the Genelec loudspeakers, attenuated by the headphones, was measured. The isolation result was then obtained as a deconvolution between the two recorded noise signals. Thus, the derived isolation curve illustrates the ambient sound isolation of the headphones as a function of frequency.

Figure 1 shows the measured isolation curves (dB re 20 μ Pa). As can be seen, intra-concha headphones (IC1, IC2) have the worst isolation of the measured headphone types. In fact, it is almost non-existent even at high frequencies. Furthermore, frequencies around 1 - 2.5 kHz are amplified. On the other hand, in-ear headphones (IE1, IE2) have clearly the best passive isolation of the measured headphone types. However, the fitting of the in-ear headphones in HATS is slightly tighter than with real human ears, which may result in excessive isolation at the lowest frequencies [6]. The supra-aural headphones introduced fairly good isolation at frequencies over 1 kHz. However, when the ANC of the SA2 headphones was turned on (SA2-ANC), it clearly improved the isolation at low frequencies.

3. AUDITORY MASKING MODELS

Auditory masking is a common phenomenon that occurs in our everyday life. By definition auditory masking occurs when one sound affects the perceived loudness of another sound. Basically, a masker (i.e., the masking sound) can hide a maskee (i.e., the sound that is being masked) completely or partially. In the former case the maskee sound becomes inaudible while in the latter case its loudness is reduced.

3.1. Masking Threshold

In order to determine the effect of masking, a masking threshold is usually calculated. The masking threshold is the sound pressure level of a maskee tone necessary to be just audible in the presence of a masker [1]. The threshold of masking can be calculated in steps as follows [2, 7]:

1. Windowing the masker signal and calculating the short-time Fourier transform (STFT).

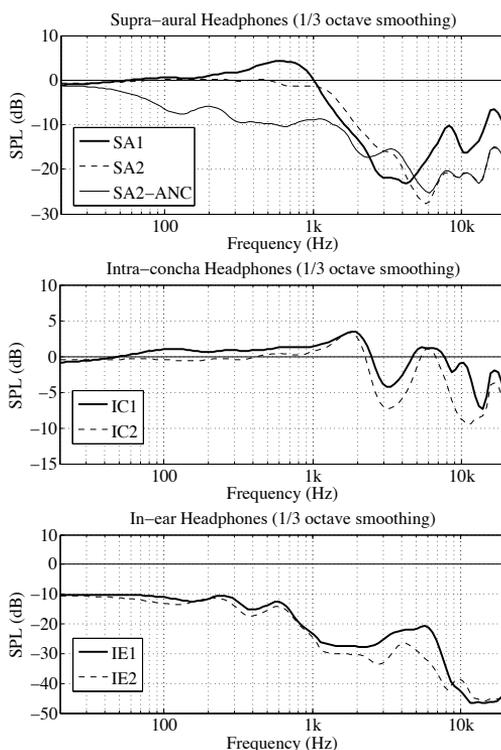


Fig. 1: The isolation curves of the supra-aural, intra-concha, and in-ear headphones, respectively.

2. Calculating the power spectrum of each discrete Fourier transform (DFT).
3. Mapping the frequency scale into the Bark domain and calculating the energy per critical band.
4. Applying the spreading function to the critical band energy spectrum.
5. Calculating the spread masking threshold.
6. Calculating the tonality dependent masking threshold.
7. Calculating the final masking threshold.

3.1.1. Power Spectrum and Bark Mapping

First the audio signal is analyzed using the STFT, which

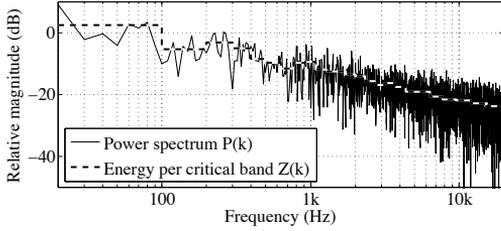


Fig. 2: A power spectrum $P(k)$ and energy per critical band $Z(k)$ of a 100 ms excerpt of pink noise.

consists of windowing short signal segments and computing the 8820-point DFT. For example, a Hamming window of 200 ms can be used with 200 ms window hops to select the next segment. The lack of overlap in spectral analysis does not lead to disturbances, such as musical noise, since the analysis data will be used for controlling the target gains of the equalizer. Then, each DFT is converted to the power spectrum

$$P_m(k) = \text{Re}\{T_m(k)\}^2 + \text{Im}\{T_m(k)\}^2 = |T_m(k)|^2, \quad (1)$$

where $T_m(k)$, and $P_m(k)$ are the calculated m^{th} DFT and power spectrum, respectively.

After that, the frequency scale is mapped into the Bark scale by using an approximation [1]

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

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

$$Z_m(v) = \sum_{k=B_l(v)}^{B_h(v)} P_m(k)/N_v, \quad v = 1, 2, \dots, N_c, \quad (3)$$

where $B_l(v)$ is the lower boundary of the critical band v , $B_h(v)$ is the upper boundary of the critical band v , N_v is the number of data points at each critical band v , and N_c is the number of critical bands, which depends on the sampling rate. For example, when the sampling rate is 44.1 kHz, N_c is 25 and the lowest and highest bounds are 5 Hz and 22 kHz, respectively. Figure 2 shows an example of a power spectrum $P(k)$ and energy per critical band $Z(k)$ calculated from a 100 ms excerpt of pink noise.

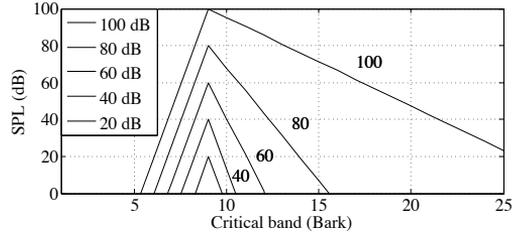


Fig. 3: Two-slope spreading function for different levels of the masker L_M .

3.1.2. Spreading Function

The effect of masking in each critical band spreads across all critical bands. This is described by a spreading function. One possibility for the spreading function model was presented by Schroeder [8]. It should be noted that the Schroeder spreading function is independent of the masker's sound pressure level (SPL). This allows the computation of the overall masking curve with a convolution between the critical band energy function and the spreading function [5].

A better approximation of the spreading function, which takes the SPL of the masker into account, is described in [5]. It is called the two-slope spreading function and it is written in terms of the Bark scale difference between the maskee and masker frequency $\Delta v = v(f_{\text{maskee}}) - v(f_{\text{masker}})$ as follows:

$$10 \log_{10}[B(\Delta v, L_M)] = [-27 + 0.37 \max\{L_M - 40, 0\} \theta(\Delta v)] |\Delta v|, \quad (4)$$

where L_M is the SPL of the masker and $\theta(\Delta v)$ is a step function equal to zero for negative values of Δv and equal to one for positive values of Δv . Figure 3 illustrates the above described two-slope spreading function.

The computation of the overall spread masking curve $S_{P,m}$ of the two-slope spreading function is not as straightforward as it is with the Schroeder spreading function. The following equation shows the summation formula:

$$S_{P,m} = \left(\sum_{v=1}^{N_c} B_v^\alpha \right)^{\frac{1}{\alpha}}, \quad 1 \leq \alpha \leq \infty, \quad (5)$$

where $S_{P,m}$ represents the intensity of the masking curve resulting from the combination of N_c individual mask-

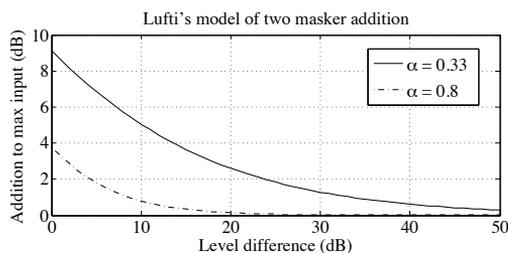


Fig. 4: Lufti's model for addition of two masking curves, with values of $\alpha = 0.33$ and 0.8 .

ing curves with intensities B_v and α is a parameter that defines the way the curves are added. Setting $\alpha = 1$ corresponds to intensity addition, while taking the limit $\alpha \rightarrow \infty$ corresponds to using the highest masking curve. Furthermore, it is possible to choose α to be lower than one, in which case the combined effect of two equal maskers is greater than the sum of their intensities [5].

Lufti [9] has suggested that the addition of masking for maskers of comparable intensities is best described using a value of $\alpha \approx 0.33$. Thus, two equal masking curves have a combined effect equal to a single masking curve with an intensity eight times that of a single curve (i.e., 9 dB when the dB difference in inputs is zero). Figure 4 illustrates Lufti's model of two masker addition, calculated using Equation (5), while $\alpha = 0.33$ and $\alpha = 0.8$.

3.1.3. Tonality and Offset

The masking threshold depends on the characteristics of the masker and the masked tone. Johnston [2] has introduced two different thresholds; a tone masking noise threshold and a noise masking tone threshold. For the tone masking noise the threshold is estimated as $14.5 + v$ dB below the overall spread masking curve $S_{P,m}$, and for noise masking the tone it is estimated as 5.5 dB below the $S_{P,m}$. Spectral flatness is used to determine the noise-like or tone-like characteristics of the masker. The spectral flatness V_m in decibels is defined as [2]

$$V_m = 10 \log_{10} \frac{[\prod_{k=0}^{N-1} P_m(k)]^{\frac{1}{N}}}{\frac{1}{N} \sum_{k=0}^{N-1} P_m(k)}, \quad (6)$$

which is the ratio of the geometric and arithmetic mean of the power spectrum. The tonality factor α_m is defined

as

$$\alpha_m = \min\left(\frac{V_m}{V_{max}}, 1\right), \quad (7)$$

where $V_{max} = -60$ dB, which means that if the masker signal is entirely tone-like, $\alpha_m = 1$, and if the signal is pure noise, $\alpha_m = 0$. The tonality factor is used to geometrically weight the above mentioned thresholds to form the masking energy offset $U_m(v)$ for each band:

$$U_m(v) = \alpha_m(14.5 + v) + (1 - \alpha_m)5.5. \quad (8)$$

The offset is then subtracted from the spread masking threshold $S_{P,m}$ to estimate the raw masking threshold R_m .

$$R_m(v) = 10^{\log_{10}(S_{P,m}(v)) - \frac{U_m(v)}{10}}. \quad (9)$$

The final masking threshold is calculated by comparing the raw masking threshold to the absolute threshold of hearing and mapping from Bark to the frequency scale. The absolute threshold of hearing is used when the masking threshold is below it.

A listening test with complex test sounds was arranged in order to validate that these psychoacoustic methods are applicable for the purposes of the perceptual frequency response simulator (see Section 5).

3.2. Partial Masking

Partial masking reduces the loudness of a target tone but does not mask it completely. This means that the masking sound does not only produce a shift of the absolute threshold to the masked threshold, but it also produces a masked loudness curve (or a masked loudness-matching function - MLMF) [1, 10].

The MLMF shows the level of the target tone alone as a function of the level of the target tone in noise. The general shape of the MLMF can be described as follows: When the target sound is close to its threshold in the masker, the level of the target in the masker is much higher than the level of the target alone. When the level of the target in the masker increases, the matched level of the target alone also increases, but at a faster rate. At a sufficiently high level, the level of the target in the masker equals that of the target alone. Furthermore, this equality then persists at higher levels [1, 10].

There was no existing partial masking model that was applicable for the perceptual frequency response simulator. Thus, a partial masking model for complex sounds, such as music, had to be constructed.

3.2.1. Partial Masking Model

A short-scale listening test with complex test sounds was arranged in order to obtain a model for partial masking. The used synthesized tonal test sounds were two bass sounds, two pad tones, and the used atonal test sounds were a synthesized bass drum sound, a snare drum sound, and a hihat sound. The main idea with the test tones was that they should have a realistic envelope and harmonic structure, i.e., not just plain sine tones. Furthermore, the tones should resemble musical sounds, such as electric bass and synthesizer sounds (tonal), and percussions (atonal). A more detailed description of the test sounds can be found in the appendix.

First, the loudness of the complex test signals had to be matched. Three persons participated in a listening test in which they compared the test sounds to a 1-kHz sine signal and adjusted the loudness of the test signals to match the loudness of the 1-kHz sine signal.

The masker was uniformly masking noise, which has a flat spectrum below 500 Hz and pink spectrum above that. The masker levels were 50 dB and 70 dB. The listening test was conducted as follows: two samples were played sequentially; first the test tone in quiet and then in noise. The tantee used a slider to adjust the level of the tone in quiet to match the level of the tone in noise.

Figure 5 shows the combined results of two subjects. The data points represent the mean of two subjects and all of the seven test tones in 50-dB and 70-dB noise, while the whiskers illustrate the standard deviation. The fact that the data points appear below the equality line (dash-dot line) shows that partial masking is in effect.

4. IMPLEMENTATION

The masking threshold is calculated as illustrated in Section 3.1. Figure 6 shows an example analysis of a 200 ms signal frame, where the squares represent the power spectrum of the noise, the dots represent the calculated masking threshold, the solid line is the power spectrum of the music signal (calculated in Bark bands), and the dashed line is the resulting perceived frequency response of the music signal, including complete and partial masking. As can be seen in the figure, the music spectrum is below the masking threshold at Bark bands 2–7 and therefore inaudible, thus, the processed music signal is attenuated completely. Furthermore, partial masking occurs at Bark bands 8–17 and 20–25, and thus, the processed music signal is attenuated according to the partial masking model.

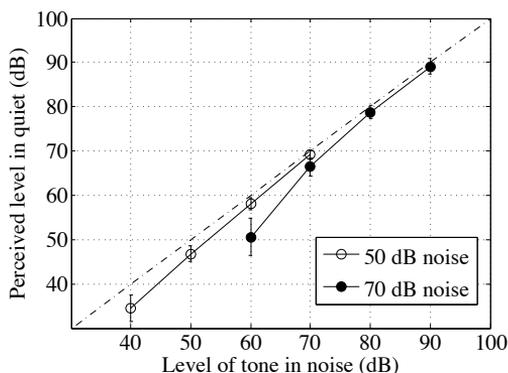


Fig. 5: Loudness of test tones as a function of their level. Whiskers extending from the data points represent the standard deviation.

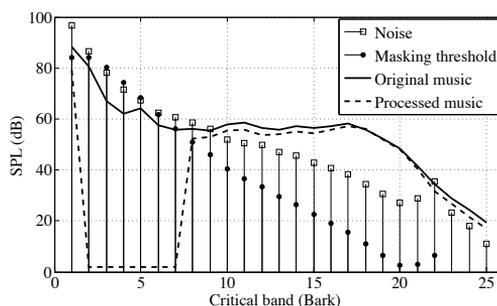


Fig. 6: An example analysis of the music and noise signal (200 ms frame).

The masking effect was implemented by using a high-order graphic equalizer [11]. With this technique the gain in one band is almost completely independent from the gain in adjacent bands. The equalizer consists of twenty-five 12th-order filters where each 12th-order filter is composed of three cascaded fourth-order sections. Figure 7 shows the block diagram of one fourth-order section of the filter. The blocks $A(z)$ contain a second-order allpass filter having the transfer function

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

The bandwidths of the highest frequency bands (21 -

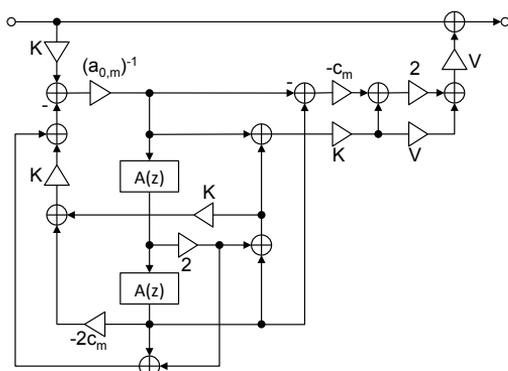


Fig. 7: Block diagram of a fourth-order section of the graphic equalizer (adapted from [11]).

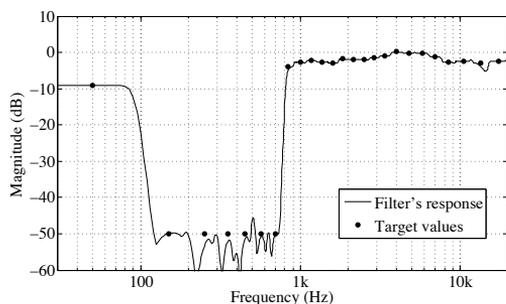


Fig. 8: An example response of the graphic equalizer. The black dots are the target gain values, and the solid line is the frequency response of the equalizing filter.

25 Bark bands) was made slightly wider than the actual Bark bands should be in order to obtain a smoother overall response. Furthermore, the order of the filters can be adjusted. The maximum cut, i.e., when the music signal is under the masking threshold, is set to be -50 dB. Figure 8 shows an example frequency response of the graphic equalizer and the target gain values for each Bark band. Figure 8 corresponds to the filter needed to create the processed music signal in Figure 6.

The partial masking was implemented based on the listening test described in Section 3.2.1. The partial masking curves for different background noise levels are interpolated according to the results shown in Figure 5. Fur-

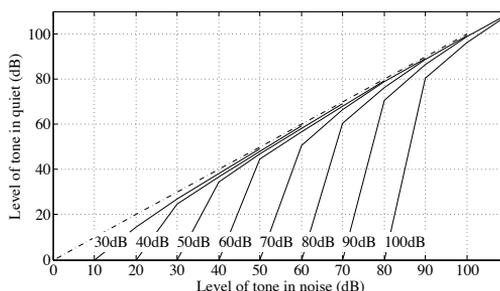


Fig. 9: Partial masking curves for different background noise levels (in uniformly masking noise). These curves are interpolated based on the data of Fig. 5 and an additional point at 0 dB, which is set 20 dB below the noise level.

thermore, background noise levels less than 60 dB use the shape of the 50 dB curve and background noise levels higher than 60 dB use the shape of the 70 dB curve. Figure 9 illustrates the interpolation of partial masking curves.

A temporal masking model was not included due to the 200 ms frame size. The implementation of the temporal masking would require a better time resolution, i.e., smaller frame size.

4.1. Calibration of the Headphones

Since both the partial masking and the masking threshold depend on the sound pressure, the system had to be calibrated so that the output level could be controlled. The calibration was performed in an acoustically treated listening room.

Sennheiser HD 650 headphones were calibrated using a Brüel&Kjær HATS model 4128C mannequin torso with type 3.3 ear simulator. The HATS ear microphone was connected to a Brüel&Kjær NEXUS microphone amplifier. First a Brüel&Kjær sound level calibrator model 4231 with a UA154 adapter was fitted to the ear microphone of the HATS. Then the RMS level of the 97.1 dB calibration signal was measured with the Audio Precision AP2700 v. 3.3 program.

Next, the Sennheiser HD 650 headphones were put on HATS and a 1-kHz sine signal generated with Matlab (with a peak amplitude of 0.01) was played through a MOTU UltraLite mk3 audio interface. The RMS level

of the sine signal was again measured with the calibrated Audio Precision program.

The measured RMS level of the 1-kHz sine signal with this measurement setting was 75.5 dB. The simulator was calibrated using this headphone calibration result. The calibration was performed so that the energy of the reference 1-kHz sine signal was calculated at the single Bark band in order to obtain the reference energy level E_{ref} . This reference was set to correspond $-L_{\text{cal}} = 75.5$ dB. For calculating the sound pressure levels (SPL) of the noise and music signal Bark bands, the following equation was used

$$Z_{\text{dB}}(v) = 10 \log_{10} \left(\frac{Z(v)}{E_{\text{ref}}} \right) + L_{\text{cal}}, \quad (11)$$

where E_{ref} is the reference energy level and L_{cal} is the measured SPL level of the 1-kHz sine signal, and $Z_{\text{dB}}(v)$ and $Z(v)$ are the SPL and the energy of the signal at the v :th Bark band, respectively.

4.2. Real-Time Demonstrator

A real-time demonstrator was constructed based on the above models in order to illustrate the auditory masking phenomenon and the advantage of the ambient noise isolation of different headphones. The filters modeling the ambient noise isolation were implemented with FIR filters of order 200. The demonstrator has a graphical user interface, which allows the user to choose different headphone types, background noises, and music tracks. Furthermore, the user can control the volume levels of the background noise and the music independently. Different listening options are music only, background noise only, music and noise together, and the processed music signal where the masking phenomenon is taken into account and all of the components that are masked by the background noise are suppressed.

The demonstrator was implemented with Matlab and Playrec software using a MOTU UltraLite mk3 audio interface and Sennheiser HD 650 headphones.

5. EVALUATION OF THE SIMULATOR

A set of listening tests was conducted in order to evaluate that the simulator is operating properly. Two listening tests using the above described signal processing and Sennheiser HD 650 headphones were conducted to evaluate the masking threshold and partial masking models. Four different noise maskers were used in these tests: uniformly masking noise, car noise, bus noise, and babble noise. The last three noise signals were produced by

filtering white noise with a 10th-order all-pole IIR filter calculated from a set of noise recordings. Linear prediction of order 10 was used to determine the coefficients of the all-pole filter.

5.1. Evaluation of the Masking Threshold Model

The masking threshold model was evaluated with an adaptive listening test adopted from [12] and [13]. The method is based on an up-down procedure in which the level of the test signal is varied with a predetermined amount of decibels either up or down. The test subject is given a single button ('Sound is audible') and instructed so that when he/she is confident to hear the test signal, the button is to be pushed. An algorithm either increases the test signal level, when the test subject has not pushed the button and therefore not detected the signal or decreases the signal level, when the test subject has pushed the button and therefore detected the test signal from the noise masker.

The objective of the listening test was to determine the masking thresholds of the noises with different test sounds. This result was then compared with the output of the simulator. All of the eight complex test signals (described in Section 3.2.1) were tested with the uniformly masking noise, and four test signals (namely 125-Hz bass, 1024-Hz pad tone, 1-kHz sine, and 5-kHz hihat sound) with each of the three LPC-filtered noises; that is, a total of 20 masking threshold levels was tested. The RMS level of the noise signals was 70 dB.

The test was conducted in the Aalto University listening room with six test subjects. The test results of the uniformly masking noise are shown in Figure 10. The variation between the results of the different test subjects is small, mostly between 4 - 8 dB. The median values are shown as filled black rectangles. The variation in median values between different test sounds with uniformly masking noise is possibly due to different frequency spectra and envelopes of the test signals.

To evaluate the correct operation of the simulator, the test sounds with RMS levels equal to the median values of the listening test results were fed through the signal processing. When the signal remained at the masking threshold of the respective noise, it could be said that the simulator is working properly.

When the system is working properly the level of the test tone was at the same level as the masking threshold with a few exceptions. Especially with the bass drum

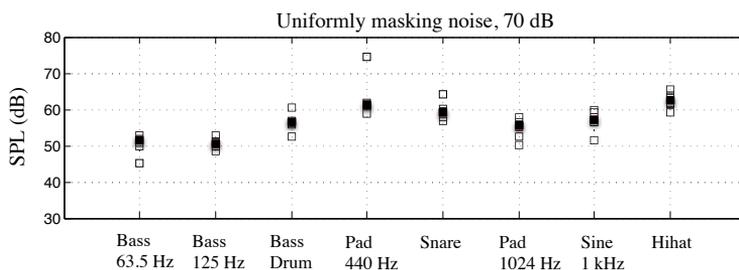


Fig. 10: Hearing threshold result in 70 dB uniformly masking noise with complex test sounds. White boxes represent six individual subjects, while black boxes represent the median of six subjects in each case.

sound the calculated sound level was much higher than the masking threshold, yet the simulator was perceived to be performing correctly. When the system was not working properly, the level of the test sound was either well below or much higher than the masking threshold level.

Figure 11 shows a successful test and an unsuccessful test cases. As can be seen in the rightmost subfigure, the system was not working correctly with bus noise, since the level of the 1024-Hz pad tone is well below the calculated masking threshold. Although, the spectrum of the bus noise is similar to car noise, the bus noise signal contains so much energy at frequencies under 50 Hz that the addition of the masking thresholds at different Bark bands is not working properly. However, with the car noise (the leftmost subfigure), it can be seen that the same test tone is at the same level as the calculated masking threshold, thus, the simulator is working properly.

5.2. Evaluation of the Partial Masking Model

The partial masking model was evaluated as follows: First the test tones were fed through the signal processing of the simulator. Then the obtained processed sounds, where the partial masking effect was taken into account, were scaled +4 dB and -4 dB. The reference sample was always the noise signal plus the original test tone, and that was compared with either the correctly processed signal, the +4 dB signal, or the -4 dB signal. The task of the testee was to judge which one of the test samples was louder, or whether they were equally loud. The test was conducted with four different noises; uniformly masking noise, car noise, bus noise, and babble noise. The sample pairs were played back randomly in both orders, processed signal – reference and reference – processed signal.

Four test subjects participated in the test. All of the testees were familiar with this project and therefore suitable for this slightly difficult listening test. Informal tests showed that 6 dB level differences are needed in order to achieve reliable distinction (almost 100 % correct) between the processed partially masked samples. However, when the processed signals were compared to the actual test tone signals in noise, it was found that 4 dB level differences were sufficient.

Figure 12 shows the results of the listening test. Each data point corresponds to the mean of the four testees including the six test cases, i.e., the reference signal compared to the real and ± 4 dB signals in both orders.

One obvious drawback was noticed when creating the test samples for this listening test. The calculation of the masking threshold for the bus noise is not accurate. This is due to the high energy at low frequencies in the bus noise. The energy in the first two Bark bands is so dominant that it affects the masking threshold of other bands. This can be seen from the results in Figure 12 as the majority of the partially masked signals were perceived to be too low in the bus noise case (unfilled circles).

Other noise types came through quite well. However, the results indicate that with the uniformly masking noise (unfilled squares) the partially masked signals was generally perceived to be too loud, with the exception of the 125-Hz bass sound, which was perceived correctly. Furthermore, there were a couple of test sounds that were perceived to be slightly too loud in the car noise and babble noise cases as well. It is fair to assume that ± 2 dB differences are so small that it will not deteriorate the user experience of the simulator.

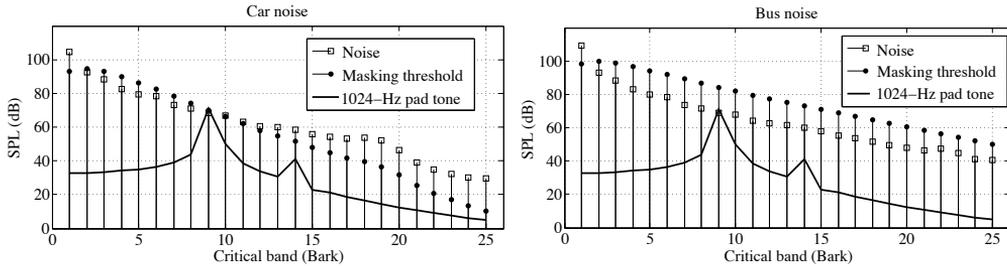


Fig. 11: Results of a successful hearing threshold evaluation test (left subfigure) and an unsuccessful test (right subfigure). The leftmost subfigure illustrates the 1024-Hz pad tone played at the level corresponding to the masking threshold of the car noise signal, whereas the rightmost subfigure illustrates the same sound played at the level corresponding to the masking threshold of the bus noise signal.

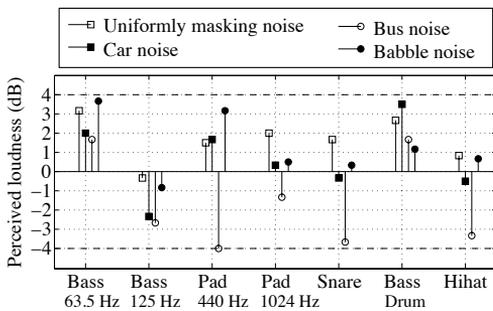


Fig. 12: Results of the partial masking evaluation test. Cases appearing below 0 dB were perceived too soft (i.e., too much simulated partial masking was applied) and cases above 0 dB were perceived too loud. Cases appearing close to 0 dB were processed correctly.

6. CONCLUSIONS

In this article a real-time demonstrator was designed and implemented to simulate the perceived music in a noisy listening environment considering the isolation capabilities of different headphones. Three different headphone types and four different background noise signals were considered. The perceived frequency response simulator takes as input a music and a background noise signal, and the user can adjust the playback level of both signals. It is then possible to listen to each signal separately, their mix, or a processed music signal from which all components masked by the background noise and the noise itself are suppressed.

This processing is implemented by running a spectral analysis for both the noise and the music signal, leading to a spectral representation on 25 critical frequency bands. A masking threshold is then calculated for the noise signal using psychoacoustic models, and the auditory spectrum of the music signal is compared against the threshold at each Bark band. A high-order graphical equalizer is used for implementing the masking and partial masking effects so that each Bark band can be attenuated between 0 dB and 50 dB.

This unique real-time demonstrator of the auditory masking phenomenon can be used for showing to a broader audience how background noise renders part of the music signal inaudible. This can be performed for various noise types, signal levels, and different headphone types. The need for prominent attenuation in headphones used in noisy environments is thus convincingly demonstrated. Furthermore, the clear advantages and superior performance of in-ear headphones can be easily shown; they are currently used in many mobile phones and offer remarkable passive attenuation.

7. ACKNOWLEDGMENT

The authors would like to thank Mr. Julian Parker for proofreading the paper.

APPENDIX: TEST SOUNDS

A set of synthetic tonal and atonal test sounds with a realistic envelope and harmonic structure was created for the listening tests. The objective was to use short signals that resemble musical sounds. The tonal test sounds

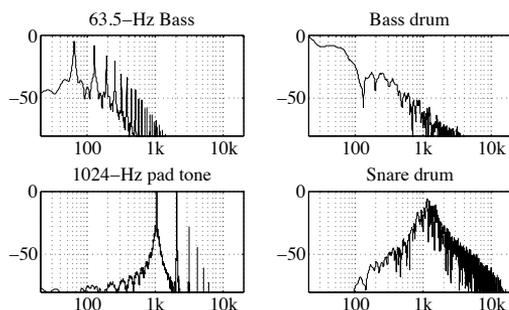


Fig. 13: Frequency responses of the test tones. Horizontal axis depicts frequency (Hz) and vertical axis depicts magnitude (dB).

were two bass tones, with fundamental frequencies of 63.5 Hz and 125 Hz, and two synthetic pad tones (440 Hz and 1024 Hz). The bass and pad sounds were synthesized by filtering a sawtooth waveform with a third- and sixth-order Butterworth lowpass filter, respectively. The temporal envelope of the bass tones had a 10-ms linear attack part and an exponential decay (time constant was 1 s), whereas the pad tones had a 50-ms linear attack part and an exponential decay (time constant was 1 s).

The noisy sounds were reminiscent of a bass drum, a hi-hat, and a snare drum. They were all synthesized by filtering an exponential decaying white noise sequence. The bass drum sound was synthesized by filtering the noise sequence with a second-order Butterworth lowpass filter (cutoff at 100 Hz). The hi-hat signal used a third-octave Butterworth bandpass filter (centered at 5 kHz) and the snare drum had a second-order Butterworth bandpass filter (cutoff frequencies 500 Hz and 2 kHz). Figure 13 shows the frequency responses of four test sounds.

8. REFERENCES

- [1] E. Zwicker, and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, New York, 1990.
- [2] J. D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," *IEEE J. Sel. Areas Comm.*, 6(2):314–323, Feb. 1988.
- [3] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," *J. Acoust. Soc. Am.*, 45(4):224–240, Apr. 1997.
- [4] 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.*, 53(10), pp. 906–918, Oct. 2005.
- [5] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*, Kluwer, 2003
- [6] ITU-T, *Recommendation P.380. Electro-Acoustic Measurements on Headsets*, Series P: Telephone Transmission Quality, Telephone Installations, Local Line Networks. ITU, 11/2003.
- [7] J. Riionheimo and V. Välimäki, "Parameter Estimation of a Plucked String Synthesis Model Using a Genetic Algorithm with Perceptual Fitness Calculation," *EURASIP J. Appl. Signal Processing*, vol. 8, pp. 791–805, 2003.
- [8] M. R. Schroeder, B. Atal, and J. L. Hall, "Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear," *J. Acoust. Soc. Am.*, 66(6):1647–1652, Dec. 1979.
- [9] R. A. Lufti, "Additivity of Simultaneous Masking," *J. Acoust. Soc. Am.*, 73(1):262–267, Jan. 1983.
- [10] H. Gockel, B. C. J. Moore, and R. D. Patterson, "Asymmetry of Masking Between Complex Tones and Noise: Partial Loudness," *J. Acoust. Soc. Am.*, 114(1):349–360, July 2003.
- [11] 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, Sept. 2006.
- [12] H. Levitt, "Transformed Up-Down Methods in Psychoacoustics," *J. Acoust. Soc. Am.*, 49 (2):467–477, 1970.
- [13] D. Isherwood and V.-V. Mattila, "Objective Estimates of Partial Masking Thresholds for Mobile Terminal Alert Tones," presented at *the AES 115th Convention*, New York, USA, 2003 Oct. 10–13.