AN AUDITORILY MOTIVATED ANALYSIS METHOD FOR ROOM IMPULSE RESPONSES

Tapio Lokki

Telecommunications Software
and Multimedia Laboratory
Helsinki University of Technology
P.O.Box 5400 FIN-02015 HUT, FINLAND
Tapio.Lokki@hut.fi

Matti Karjalainen

Laboratory of Acoustics and
Audio Signal Processing
Helsinki University of Technology
P.O.Box 3000 FIN-02015 HUT, FINLAND
Matti.Karjalainen@hut.fi

ABSTRACT

In this paper a new auditorily motivated analysis method for room impulse responses is presented. The method applies same kind of time and frequency resolution than the human hearing. With the proposed method it is possible to study the decaying sound field of a room in more detail. It is applicable as well in the analysis of artificial reverberation and related audio effects. The method, used with directional microphones, gives us also hints about the diffuseness and the directional characteristics of the sound fields in the time-frequency domain. As a case study two example room impulse responses are analyzed.

1. INTRODUCTION

Traditionally, room impulse responses are analyzed with octave or one-third octave bands in the frequency domain. For visualization, a spectrogram which shows the temporal behavior of each frequency band, is often used. However, this analysis approach is not optimal from a perception point of view. This is the reason why perceptually more relevant way to analyze room impulse responses is presented in this paper.

In auditory modeling the aim is to find mathematical models which represent some physiological or perceptual aspects of human hearing. Auditory modeling is potentially very useful because, with a good model, audio signals can be analyzed in a similar way that our hearing does. The method presented in this paper is not an accurate auditory model, it is rather an audio engineer's approach to the modeling of perception. Also, we do not try to model the binaural properties of the auditory system, rather we use directional microphones for capturing the directional components of the sound field.

This paper is organized as follows. First, as a motivation, the time and frequency resolution of human hearing is discussed. Then the proposed analysis method is presented in section 3 and directional analysis is discussed in section 4. In section 5 two room impulse responses are analyzed with the proposed method. Finally, conclusions are drawn with a discussion on future guidelines of research.

2. FREQUENCY AND TIME RESOLUTION OF HUMAN HEARING

The frequency resolution of human hearing is a complex phenomenon which depends on many factors, such as frequency, signal bandwidth, and signal level. Despite of the fact that our ear is

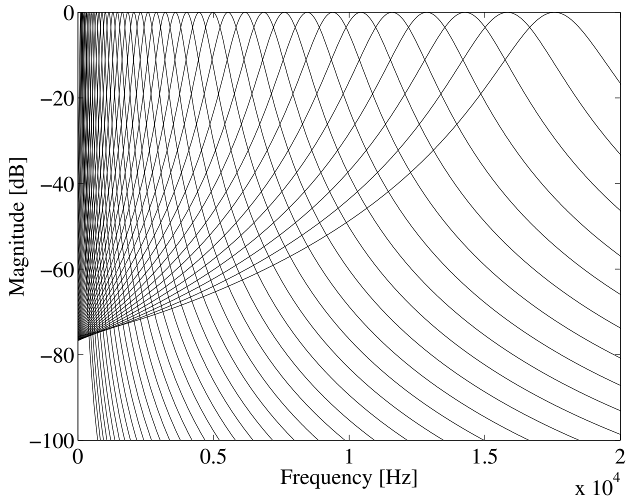

Figure 1: *Magnitude responses of a gammatone filterbank (40 channels, 100-20000 Hz).*

very accurate in single frequency analysis, broadband signals are analyzed using quite sparse frequency resolution. Critical bandwidth theory (see, e.g., [1]) and Bark scale is a classical way to explain the frequency resolution of human hearing with broadband signals. Another scale, considered more accurate for auditory research, is the Equivalent Rectangular Bandwidth (ERB) scale [2, 3]. It has logarithmic behavior in a wider frequency band than the Bark scale. The width of an ERB band (in Hz) is typically 11-17 % of center frequency. One ERB band, as a function of center frequency f_c, can be calculated with equation [2]

$$ERB = 24.7 + 0.108 * f_c \quad (1)$$

where f_c is the center frequency (in Hz) of the band.

The ERB band is a psychoacoustic measure of width of the auditory filter bandwidth at each point of the cochlea. A practical implementation of ERB filters as a filterbank was presented by, e.g., Slaney [4]. The filters are based on gammatone functions, one of which is defined by

$$g(t) = at^{n-1}e^{-2\pi b(f_c)t}\cos(2\pi f_c t + \theta) \quad (2)$$

where $at^{(n-1)}$ defines the start of the response, $b(f_c)$ is the bandwidth of the ERB band (in Hz), f_c is center frequency and θ is

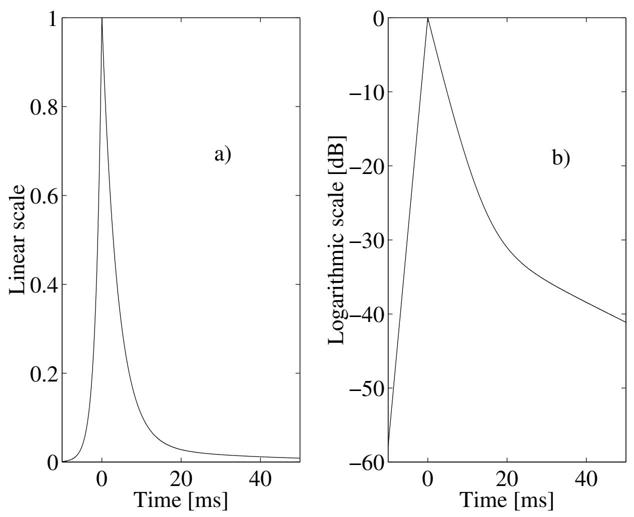

Figure 2: *Integrating window used in the analysis, using a) linear and b) logarithmic amplitude scale.*

phase. In Fig. 1 magnitude responses of a gammatone filterbank, which contains 40 ERB filters, are presented.

The time resolution of human hearing is even more complex phenomenon than the frequency resolution. In some cases monaural time resolution of our hearing is 1-2 ms at high frequencies and a little bit worse at lower frequencies. On the other hand the temporal integration time constant and the postmasking effect after a noise masker (when masker is longer than 200 ms) are over 100 ms, even 200 ms.

A complete model for time resolution is not known. In this study we have tried to find an integrating window which simulates the temporal integration phenomenon of human ear. After applying several windows we ended up using a slightly modified version of the window presented by Plack and Oxenham [5]. It is claimed to be sufficiently good for various situations. The shape of the temporal window is described by a combination of two exponential functions:

$$W(t) = 0.975e^{(t/4)} + 0.025e^{(t/29)}, \quad \text{when } t < 0 \tag{3}$$

and

$$W(t) = e^{(-t/3.5)}, \quad \text{when } t \geq 0 \tag{4}$$

where $W(t)$ is a temporal weighting function and t is time (in ms) measured relative to the maximum of the weighting function. A picture of the temporal window applied is depicted in Fig. 2.

3. AN AUDITORILY MOTIVATED ANALYSIS METHOD

A block diagram of the proposed analysis method is presented in Fig. 3. The input signal is fed to a gammatone filterbank which divides the signal into 40 ERB bands, similar frequency bands than the human ear does.

After the division to the ERB bands the signals are squared which resembles the half-wave rectification done by the hair cells in the human hearing. Then there is a sliding window which simulates the time resolution of the ear. The implementation of the temporal window used is discussed in more detail in section 3.1.

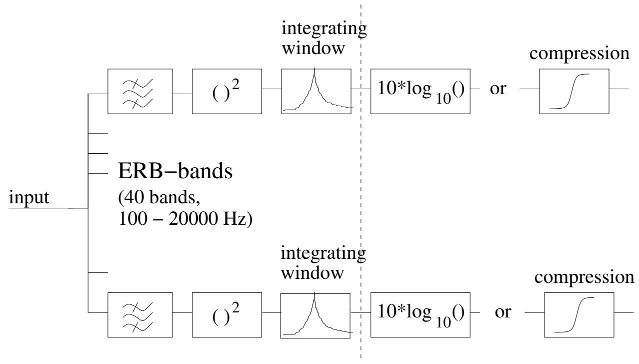

Figure 3: *A block diagram of the analysis method.*

The human auditory system exhibits varying sensitivity as a function of frequency. This can be modeled as a frequency weighting filter, such as the inverse of 60 dB equal loudness curve. For the purpose of this study we did not add such processing since in auditory perception such permanent emphasis is at least partly compensated for and thus it can be dropped in the visualization of analysis results.

The final step in the analysis is to use some mathematical operation for visualization purposes. By taking the logarithm of the rectified and temporally processed signal in each frequency band we can depict the decibel values in a time-frequency plot. Another useful tool for visualization is to apply compression to get a desired part of the whole dynamic range emphasized.

3.1. Implementation issues of the proposed method

Implementation details of designing the gammatone filterbank are out of the scope of this article, for more information see, e.g., [4]. Another implementation and a free Matlab code is available in the HUTear toolbox [6].

The effective duration of the temporal window (see Fig. 2) is several thousand signal samples (at 44.1 kHz sampling frequency). An FIR implementation of this response leads to a computationally expensive implementation. Härmä [7] has proposed an efficient implementation by dividing the filter into causal and non-causal parts. First the causal part is implemented with a second order IIR filter (Z-transform of the IIR implementation of equation (4), at sampling rate = 44.1 kHz), the transfer function of which is

$$H_c(z) = \frac{1 - 0.999097z^{-1}}{1 - 1.993565z^{-1} + 0.993569z^{-2}} \tag{5}$$

The non-causal part of the window function is a time-reversed exponential function. There is no causal IIR implementation for this kind of impulse response but it is possible to implement by using a time-reversed signal with the following filter

$$H_n(z) = \frac{1}{1 - 0.985z^{-1}} \tag{6}$$

As a summary the filtering algorithm is (for the input signal $s(n)$)

1. Filter $s(n)$ using $H_c(z)$ to produce signal $y_1(n)$
2. Reverse $s(n)$ in time and filter with $H_n(z)$ to produce signal $y_2(n)$

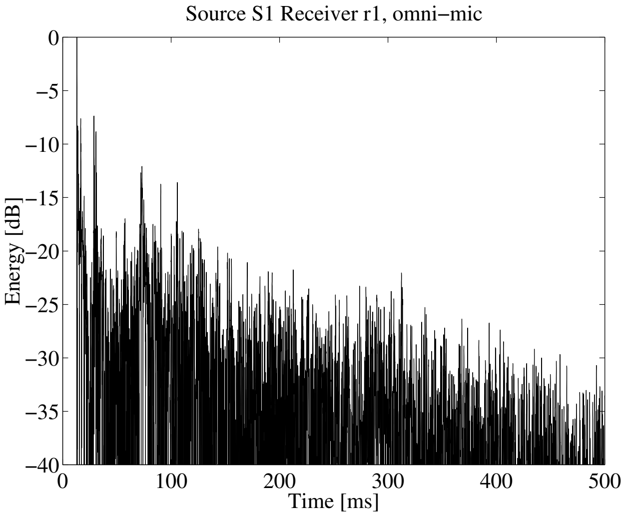

Figure 4: *An example energy-time curve of analyzed impulse response. The response is measured on the top of the second seating row of a 500-seat concert hall. Both the source and the microphone had omnidirectional directivity patterns.*

3. Reverse $y_2(n)$ in time again and shift it backwards by one sample period
4. Final output is given by $y(n) = y_1(n) + y_2(n)$.

In this way the implementation is easy and efficient.

A final implementation problem of the proposed method relates to the visualization of results. The amount of analyzed data from one impulse response is quite extensive and the result is a function of both time and frequency. If colors can be used, the best plots can be obtained with a 2-D plot (see Fig. 5) where the magnitude is indicated with different colors. The other way to present results is to use a 3-D waterfall plot, which is useful in detecting decaying properties of each channel (see Fig. 6).

4. DIRECTIONAL ANALYSIS OF ROOM RESPONSES

A proper way to include directional and spatial properties of auditory analysis would be to develop a binaural auditory model [8]. Perception of source direction, based on direct sound but discarding the influence of early reflections (precedence effect), perceiving spatial attributes due to reflections and reverberation at different time moments, etc., are generally known phenomena. However, there exist no detailed binaural models for room acoustics analysis that include these effects beyond interaural cross-correlation [9] or similar simplified methods.

Instead of hypothesizing new advanced binaural models we combined monaural auditory analysis and signals captured by directional microphones. In this way the physics of the arriving sound wavefronts is also easily interpretable. For example, cardioid microphones can capture the component of a sound field that is arriving from the main axis frontal direction. If this first order directional accuracy is not enough, microphones with higher directivity can be applied as well.

Based on this kind of directional selectivity it is possible to study the spatiotemporal formation of the sound field in a room, and yet apply monaural auditory analysis for proper time-frequency resolution. For example discrete echoes can be analyzed using this approach. Two concert hall cases will be discussed below where the arrival of sound energy at different time spans is analyzed.

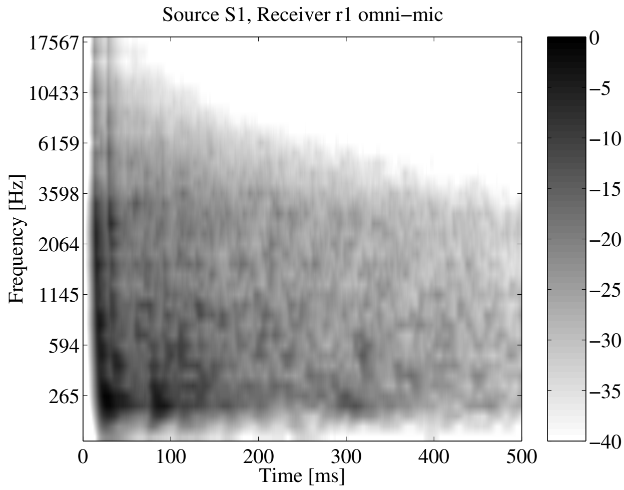

Figure 5: *An example of auditorily motivated analysis of an impulse response.*

5. EXAMPLE ANALYSIS OF TWO IMPULSE RESPONSES

To illustrate the analysis method, two example room impulse responses are analyzed. First one is measured in a 500-seat concert hall while the other is from a 2000-seat concert hall.

5.1. Small concert hall

The broadband energy-time curve (ETC), which is the squared impulse response, of a small concert hall is plotted in Fig. 4.

The same impulse response is analyzed with the proposed method and the result is depicted in Figs. 5 and 6. The analysis is done on the frequency range of 100-20000 Hz, regardless of the fact that the source used in the measurement does not radiate much energy above 10 kHz. This can be seen in Figs. 5 and 6, as well as the rapid attenuation of high frequencies over time. An interesting detail in Fig. 5 is the dark areas around 300 ms. From the ETC curve (Fig. 4) it can be seen that there is a group of reflections around 300 ms. Again from Fig. 5 it is seen that the energy of this reflection group is at low frequencies around 250 Hz and around 600 Hz a dozen milliseconds later. It would be interesting to know from which directions these sound components come from.

The proposed method allows us also study the directional characteristics of the impulse responses. For this study we have done the same impulse response measurement with two cardioid microphones which were pointed to the stage and to the audience. With these microphones positioned between the stage and the audience area we obtained two impulse responses that tell us some facts about the directional characteristics of the sound field at the measurement point. If the two responses are analyzed with the proposed method and subtracted from each other, an estimation of the direction of sound energy flow at each time moment is acquired.

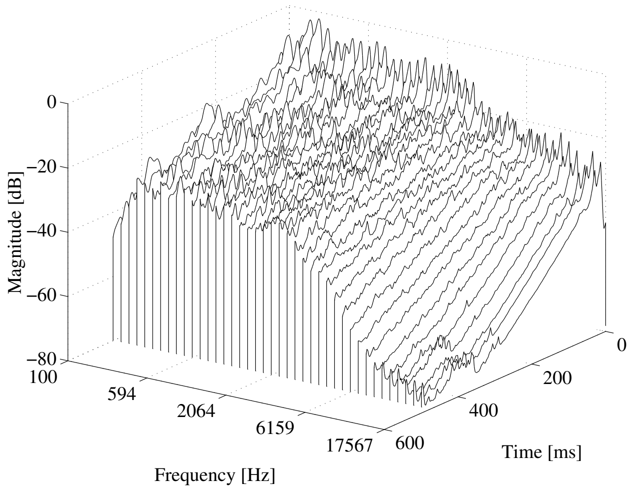

Figure 6: *The same result as in Fig. 5, but presented as a waterfall plot.*

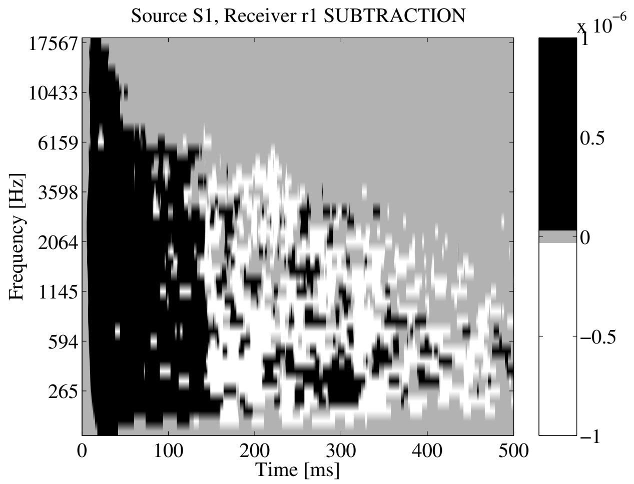

Figure 7: *An example of the analysis of directional aspects of the sound flow.*

Because of temporal integration of the analysis method this subtraction is more reliable than a subtraction of two ETC curves.

The above described directional analysis was done and the result is shown in Fig. 7. The black areas are obtained when there is more energy propagating from the stage area to the audience area than the other way. In other words when the result of subtraction has positive values, sound flows from stage to the audience. It is seen in Fig. 7 that in this case the energy before 150 ms is flowing to the audience area and then back during the next 100 ms. This is an expected result, since 150 ms corresponds to about 50 meters distance, which in this hall is the distance from sound source to the back wall and then to the measuring point. After 250 ms the sound field is more or less diffuse because no black neither white areas are dominating. An interesting finding can be made around 300 ms. The reflections around 250 Hz are coming from the stage area (black color in Fig. 7) while the other group of reflections around 600 Hz is coming from the audience direction (white area in Fig. 7).

5.2. Large concert hall

The broadband ETC curve of a large concert hall is plotted in Fig. 8. From this curve we can see that there is one distinct reflection at about 20 ms after the direct sound and later after about 50 ms there is a group of strong reflections. The auditorily motivated analysis (see Figs. 9 and 10) tells us the frequency contents of these reflections. For example, there is a possible group of reflections at low frequencies after 100 ms time stamp, because at this time the magnitude is even higher than the magnitude of direct sound at low frequencies.

In this case two cardioid microphones were also used, but this time they were pointing to the side walls of the hall. By this way we could have information on the direction of the lateral energy flow at the measuring point. The auditorily motivated analyses were done for both impulse responses and a subtraction of them is plotted in Fig. 11.

It can be seen that the above-mentioned distinct reflection is coming from the right side of the measuring point while the group of reflections after 100 ms time stamp is coming from the left side. (At least major part of reflections is coming from left side because the energy at measuring point at this particular time moment is flowing from left to right.)

6. CONCLUSIONS

A new way to analyze room impulse responses is presented. The analysis method resembles the traditional one-third octave band spectrogram analysis. It filters the impulse response to several subbands and then applies a temporal smoothing to the energy envelope of each band.

Although the proposed method is not based on a full-scale auditory model, it better respects the frequency and time resolution of human hearing than a one-third octave band spectrogram. Also the integrating temporal window is a simplified model of the time resolution of human hearing and it might not be an ideal one for the analysis of impulse responses for small rooms. Nevertheless, the features, such as frequency or time analysis parameters of the model, can be adjusted according to desired results.

The model is monaural but it can be used to study directional aspects of sound fields by applying two or more directional microphones. An interesting application of this feature is search for

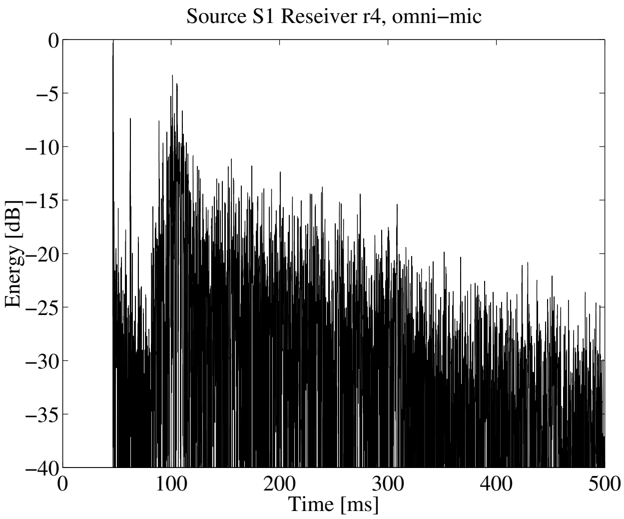

Figure 8: *The ETC curve is measured in the middle of main floor of a 2000-seat concert hall. Both the source and the microphone had omnidirectional directivity patterns.*

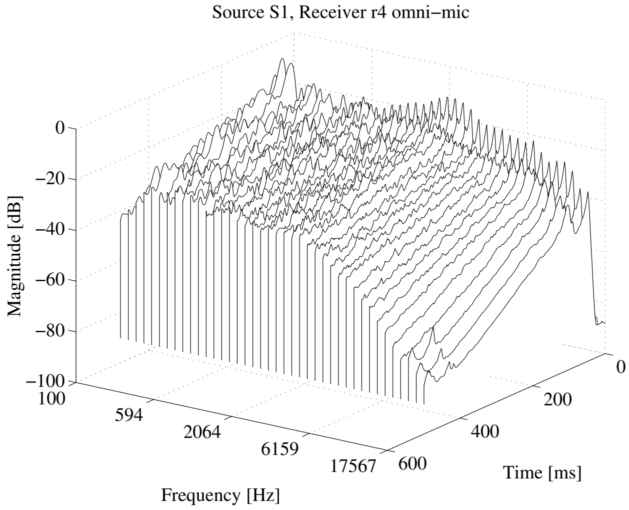

Figure 10: *An auditorily motivated analysis, presented as a waterfall plot, of the ETC curve shown in Fig. 8.*

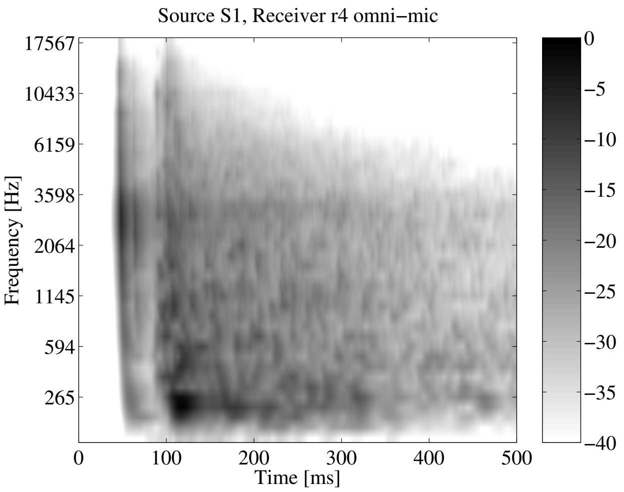

Figure 9: *An auditorily motivated analysis of the ETC curve shown in Fig. 8.*

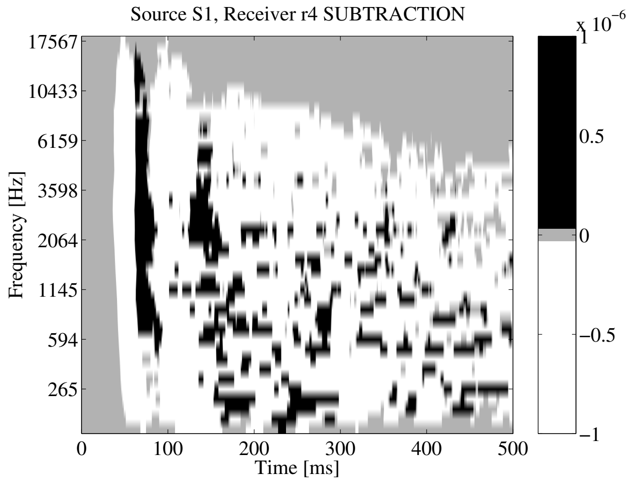

Figure 11: *An example analysis of lateral energy flow. White areas are obtained when to the left-pointing cardioid microphone is dominating and black areas when to the right-pointing cardioid microphone is dominating.*

disturbing discrete echoes and their possible sources by directional analysis.

The proposed method is only a framework for more accurate and auditorily motivated analysis of room acoustics, even if it is already proven to be an applicable tool, as presented with two examples above. Future work should include adding auditory modeling details, particularly binaural features, in order to see if they contribute to the analysis and design for better room acoustics, virtual acoustics applications, or evaluation of spatial audio effects.

7. ACKNOWLEDGMENTS

This work has been financed by the Technology Development Centre of Finland (TEKES) and the Helsinki Graduate School in Computer Science and Engineering.

8. REFERENCES

[1] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, Heidelberg, Germany, 1990.

[2] B.C.J. Moore, R.W. Peters, and B.R. Glasberg, "Auditory filter shapes at low center frequencies," *J. Acoust. Soc. Am.*, vol. 88, pp. 132–140, 1990.

[3] B.C.J. Moore and B.R. Glasberg, "A revision of Zwicker's loudness model," *ACUSTICA united with acta acustica*, vol. 82, pp. 335–345, 1996.

[4] M. Slaney, "An efficient implementation of the Patterson – Holdsworth auditory filter bank," Tech. Rep. 35, Apple Computer, Inc., 1993, Available at: http://www.slaney.org/malcolm/apple/tr35/PattersonsEar.pdf.

[5] C.J. Plack and A.J. Oxenham, "Basilar-membrane nonlinearity and the growth of forward masking," *J. Acoust. Soc. Am.*, vol. 103, no. 3, pp. 1598–1608, Mar. 1998.

[6] A. Härmä and K. Palomäki, "HUTear – a free Matlab toolbox for modeling of auditory system," in *Proc. 1999 Matlab DSP Conference*, Espoo, Finland, Nov. 1999, pp. 96–99, Available at http://www.acoustics.hut.fi/software/HUTear/.

[7] A. Härmä, "Temporal masking effects: single incidents," Tech. Rep., Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 1999, Available at: http://www.acoustics.hut.fi/˜aqi/papers/Time.ps.gz.

[8] J. Blauert, *Spatial Hearing. The psychophysics of human sound localization*, MIT Press, Cambridge, MA, 2nd edition, 1997.

[9] Y. Ando, *Concert Hall Acoustics*, Springer Series in Electrophysics 17. Springer-Verlag, Berlin, 1985.