VOLUME CONTROL IN NETWORKED AUDIO SYSTEMS

Aki Härmä, Arno van Leest, and Rainer Thaden

aki.harma@philips.com
Philips Research Laboratories, Eindhoven, The Netherlands.

ABSTRACT

If all audio appliances in a building are connected to a network and equipped with microphones it becomes feasible to consider an active volume control system where, e.g., audio played from one device does not disturb other people in the neighborhood. The fundamental problems are the identification of the audio source and the estimation of the acoustic transfer function. For the identification problem we propose a method where a unique imperceivable watermark is embedded in each loudspeaker signal. For the path estimation we compare two algorithms in a simulation and find that a simple method based on subband level differences may be sufficient for many typical cases.

1. INTRODUCTION

The *thoughtless use of* audio appliances is one of the largest causes of noise complaints [1]. The reason for playing audio too loud is not always the lack of thoughtfulness, but rather the absence of practical means of knowing, for example, how loud the music played in a living room is perceived in the bedroom, or in neighbor's bedroom. For the same reason many people also set the volume inconveniently low. In an ideal volume control the audio system would know how loud it really is at different locations in the environment. Naturally one could perform a series of acoustic measurements in the dwelling to find out the acoustic transfer functions $H_{AB}(\omega)$ from each speaker A to each point of interest B. Then the transfer functions could be used for dynamic equalization and level control with the help of a computational loudness model [2]. Systems where a networked audio appliances are calibrated by off-line measurements have been proposed recently by several authors, e.g., [3]. In this paper we want to avoid the off-line measurement and perform the path identification during the normal operation of the system.

A vision that is endorsed by all electronics industries predicts that all appliances at home will be connected to a home network. Moreover, appliances are becoming increasingly *audio-capable* such that many devices in our environment have (built-in) electro-acoustic transducers such as microphones or loudspeakers. A text book example of an UPnP device is a toaster with a built-in mp3-player and a small speaker [4]. The standard home networking protocols such as UPnP makes it possible for the devices to *discover* each other, share information, and transfer media streams. If networked devices are also equipped with microphones then there are various possibilities to measure acoustic characteristics of the environment.

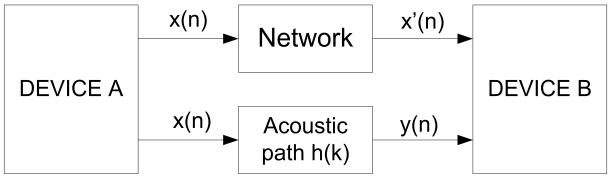

Figure 1: *A networked audio system of two devices.*

2. ONLINE MEASUREMENTS IN A NETWORKED AUDIO SYSTEM

A typical setup considered in this article is the following. User A is listening to music in the living room from home stereos. The aim is to adjust the playback volume such that it is sufficiently high for User A, but at the same time does not disturb User B going to sleep in a bedroom. We assume that User B has, e.g., a network-connected clock radio which has a built-in microphone. Clearly, the acoustic path $H(\omega)$ from the stereos to the clock radio can be estimated by the comparison of the microphone signal to the original music signal available over the network. For volume control it is sufficient that the attenuation is known in a small number of frequency bands. In this article, the attenuation of sound is measured at 42 frequency bands uniformly distributed on the near-logarithmic Equivalent Rectangular Bandwidth (ERB) rate scale. The same frequency representation is also used in current loudness models [2].

The signal paths in a system consisting of two devices are illustrated in Fig. 1. The observed microphone signal is $y(n)$. The original signal $x(n)$ played from the loudspeaker of Device A is also basically available to Device B over the network link. Therefore, it is possible to find an estimate for the acoustic transfer function $H(\omega) = Y(\omega)/X(\omega)$, where $Y(\omega)$ and $X(\omega)$, are the Fourier trans-

forms of the signals. The off-line method of performing this is often called deconvolution, and the online algorithm converging to the same solution is the adaptive filter. In both cases it is necessary to send $x(n)$ over the network for the actual computation, which increases the network load. In this paper we use an adaptive filter as a reference method for path estimation and compare that to a simplified method where the network load is only a fraction of that.

3. THE EXPERIMENT

The experimental setup is based on simulations of the propagation of sound in a building. For this, an algorithm is used which calculates a transfer path from a source in one room to a receiver in an adjoining room. This is performed at the sampling rate of 44.1 kHz. The calculation is based on known properties of building materials and the connections between them. Only transmission paths with no more than one junction are considered. At each junction the sound energy is reduced by app. 10 dB. The reverberation in the receiving room is also included in the simulation. As input data, the building products which the 2 room situation consist of and the geometrical data as well as measured room impulse responses are used. It is, thus, rather simple to collect a large amount of examples for different ranges of sound insulation and to investigate the performance of proposed methods for different levels of reverberation and insulation. The algorithm does not produce a physically exact sound field but it reproduces the correct colouration and loudness, which are sufficient for the current study. A more detailed description can be found in [5].
In the current paper we show one typical example representing the attenuation of sound in propagation from a living room to a bedroom room which has the reverberation time of $T_{60} = 0.3s$. The background noise in the receiving room is modelled by a pink noise sequence added to the target signal.

4. FREQUENCY-DOMAIN ADAPTIVE FILTER

In model-based system identification the acoustic path is approximated as a linear system, for example, an adaptive FIR filter. In this paper we use frequency domain adaptive filter (FDAF), see, e.g., [6] for a review. FDAF is known for a good performance in various applications of acoustic signal processing. The implementation used in the current article is a *constrained* algorithm based on the overlap-save method and computes the error in the time-domain. The gains of the adaptation step size matrix are adapted individually in the control of the variance of the values. This makes the algorithm well suited for a complex case

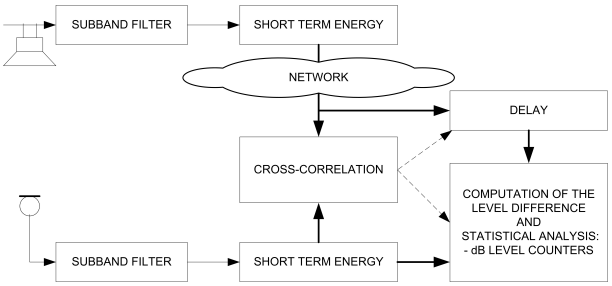

Figure 2: *The subband analysis system based on comparison of subband signal energies.*

of low signal to noise ratio and distracting sources.
For comparison, the obtained frequency domain weights $\mathbf{W}(k)$ can be mapped to the ERB scale to establish the connection to loudness models.

5. COMPARISON OF SUBBAND ENERGIES

The traditional way of measuring the sound propagation in a dwelling is based on (third) octave band energy analysis. For example, van den Eijk introduced a sophisticated system for the characterizing the disturbance of the *neighbour's radio* in 1959 [7]. The measurement system had eight octave band filters each followed by a thyratron circuit and a column of counting devices. Each counter started counting when rectified signal level exceeded 65, 70, $\cdots$, and 90 dB, respectively. After a measuring period the counter values divided by the total count in each column to produce a cumulative amplitude probability density function at each octave band. The derivative of that gives the probability density function (PDF) of level estimates. The van den Eijk's machine can be turned into a device for path identification by replacing the levels by level differences between the original and the observed microphone signal. In this paper this is called subband energy analysis (SBEA) method.
The algorithm tested in this article runs at the sampling rate of 44.1 kHz. The microphone signal and the original signal are split into 42 frequency bands where the bandwidth of each band is one ERB. The processing for one band is shown in Fig. 2. The energy envelope within each band is computed with the temporal resolution of 10 ms and sent over the network to the receiver. There the microphone signal is processed similarly. The time differences between the subband envelopes of the original signal and microphone signal are then compensated by finding the maximum peak of the normalized cross-correlation function between the envelopes. The envelopes corresponding to the original signal are then delayed so that the level differences can be computed by simple the subtraction of the

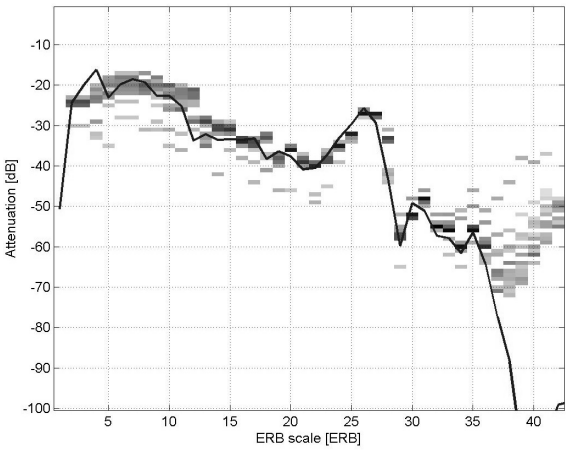

Figure 3: *The PDF of level differences obtained from the subband analysis system in the case of sound propagation from one room to another with a 20 dB SNR. The true transfer function is plotted in the figure.*

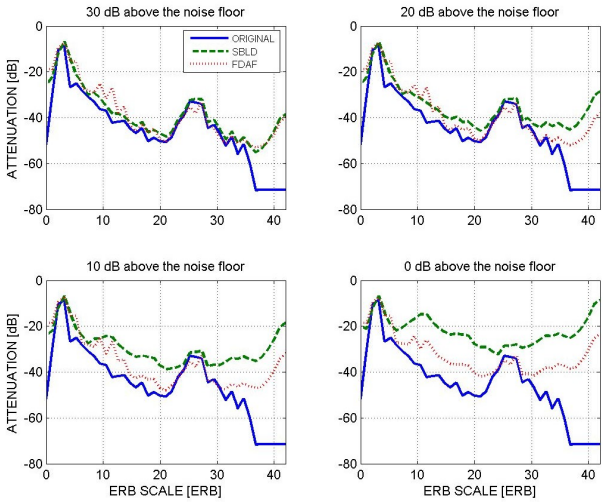

Figure 4: *Attenuation of sound between a living room and a bedroom (solid), and the estimates obtained with SBEA (dashed), and FDAF (dotted) algorithms for four different SNR values. The responses were measured using a 10s fragment of hard rock music.*

decibel-valued short-time envelopes. In the current article the frame size for the envelope processing is one second, that is, only 100 envelope values at each subband.

Each time a certain level difference value is registered, the corresponding counter in the statistics unit (see, Fig. 2) is incremented. The increment step size is adaptive and it is a function of the magnitude of the maximum peak in the normalized cross-correlation function. After an analysis period, the unnormalized PDF of level differences can be read from the counters. The PDFs obtained using a 10 s segment of music are shown (dark gray represents high probability) in Fig. 3. The original response computed from the impulse response is plotted in the same figure. The microphone signal was produced using Setup 1 introduced below with a pink noise background distractor at the dBA level of 20 dB below the music.

In this article the maximum position of the PDF within each subband is selected as the estimate for the attenuation. The attenuation of sound at different ERB bands in the simulated setup is shown in Fig. 4 (solid curve). The panels represent the cases where the level difference between the sound leaking from the other room is 30, 20, 10, and 0 dBA above the level of the pink background noise. The estimates obtained with SBEA (dashed) and FDAF (dotted) method are plotted in the same figure.

At high levels of leaking sound both methods give very similar estimates for the attenuation. Closer to the noise floor, SBEA underestimates the true amount of attenuation while FDAF is capable to follow slightly below the noise floor. However, above the noise floor the two methods are comparable.

6. IDENTIFICATION BY EMBEDDED WATERMARKS

Often there are two or more devices rendering the same audio material, for example, all neighbors watching the same TV program. Therefore, the identity of the device which is heard in the microphone signal remains unclear in both methods described above.

We propose using a very similar technique that is commonly used in audio watermarking. We imperceptibly embed a unique watermark v_i (a periodic random white noise sequence, with a period length of, say, L) into each device signal x_i using a psycho-acoustic model such that the spectrum level of the watermark follows the masked threshold in time and frequency. In the receiver the device is then identified by means of a watermark detector; the received signal y is correlated with all cyclicly shifted watermarks v_i. If the highest correlation, in terms of standard deviation, exceeds a threshold T then it is assumed that the microphone signal y captured the signal x_i.

In more detail, device i plays the signal $\tilde{x}_i = x_i + v_i$, where it is assumed that the audio signals x_i are uncorrelated with the watermark v_i. Moreover it is assumed that the watermarks v_i and its cyclicly shifted versions are uncorrelated as well (at least, the correlation is small). This property makes it possible to distinguish the different devices. The microphone captures signal y, which corresponds, apart from background noise, to the sum of watermarked signals $\tilde{x}_i$ convolved with the corresponding

acoustic paths h_i. Subsequently, this signal is accumulated in a buffer of length L to increase the watermark-to-signal ratio, and is correlated with one period of watermark v_i and its $L-1$ cyclicly shifted versions. This operation results in a correlation buffer b_i of length L. It is not difficult to show, that this correlation buffer b_i contains a noisy version of the acoustic impulse response h_i (provided that the watermark-to-signal ratio is high enough and the length L is large enough to contain the greater part of the impulse response energy). This buffer b_i is normalized by dividing it by its standard deviation. The highest peak (in absolute sense) in this buffer is compared with a threshold T (in our experiment we chose $T = 5$). If this peak exceeds this threshold then it is decided that the microphone signal y contains the audio signal x_i (in fact, we assume here that the elements of the normalized correlation buffer behave as normally distributed uncorrelated random variables with standard deviation equal to one). Moreover, since the the correlation buffer contains the impulse response h_i (although noisy), it is possible to make an estimate on which signal x_i is the most disturbing.

Alternatively, the embedded watermarks v_i can be used as the far-end signal of an adaptive filter (we used the FDAF) with the microphone signal y as the input signal. If the energy of the watermark v_i in the microphone signal y is large enough then the filter weights of the adaptive filter resembles the acoustic path h_i.

The performance of the watermarking method for three different types of music material is illustrated in Fig. 5. The curves represent the value of the maximum of the estimated response as a function of the distance to the pink noise floor. If the threshold is set to 5, the source can be detected reliably in spectrally rich heavy music (Entombed) even at very low levels, but in highly tonal banjo jazz (Bela Fleck) and symphonic music (Bruckner) samples the detectability is low when SNR is below 30 dB. The differences result from the different amplitudes of watermark data in the three signals.

7. CONCLUSIONS

In this article we compared two methods for online estimation of the attenuation of sound in propagating from one room to another in a building. The first method uses adaptive filtering and the second method is based on simple comparison of amplitudes of ERB subband envelopes. It was found that the two method produce similar results for signals above the background noise floor in a simulated system for sound insulation in a dwelling. Secondly, we evaluated the performance of a method for the identification of a sound source from a microphone signal. The method which was based on embedding a watermark signal shaped by the masked threshold curve was found to work well for music signals, although, the performance depends very much on the type of the audio signal.

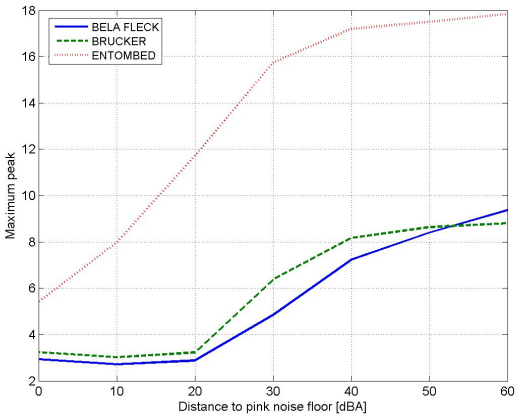

Figure 5: *The value of the maximum peak as a function of the distance of the leaking signal to the pink noise floor.*

8. REFERENCES

[1] B. Berglund, T. Lindvall, and T. H. Schwela, "Guidelines for community noise," Report, World Health Organization (WHO), Geneve, Switzerland, 1999.

[2] B. C. J. Moore and B. R. Glasberg, "A revision of Zwicker's loudness model," *Acustica - Acta Acustica*, vol. 82, pp. 335–345, 1996.

[3] Tom Blank, Bob Atkinson, Michael Isard, James D Johnston, and Kirk Olynyk, "An internet protocol (ip) sound system," in *Proc. 117th AES Convention Paper*, San Francisco, CA, USA, October 2004.

[4] M. Jeronimo and J. Weast, *UPnP design by example - software developers guide to Universal Plug and Play*, Intel Press, 2003.

[5] M. Vorländer and R. Thaden, "Auralisation of airborne sound insulation in buildings," *ACUSTICA united with ACTA ACUSTICA*, vol. 86, no. 2, pp. 76–89, 2000.

[6] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, pp. 14–37, January 1992.

[7] J. van den Eijk, "My neighbour's radio," in *Proc. 3rd Int. Congress Acoustics*, Stuttgart, Germany, 1959, pp. 1041–1044.