Perception of Temporal Decay of Low-frequency Room Modes

Matti Karjalainen[1], Poju Antsalo[1], Aki Mäkivirta[2], and Vesa Välimäki[1]

[1] *Helsinki Univ. of Tech., Lab. of Acoustics and Audio Signal Processing, Espoo, FIN-02015 HUT, Finland*

[2] *Genelec Oy, Olvitie 5, FIN-74100 Iisalmi, Finland*

Correspondence should be addressed to Matti Karjalainen (`matti.karjalainen@hut.fi`)

ABSTRACT
Modal equalization has recently been of research interest in order to improve sound reproduction in rooms that have excessively strong modes at low frequencies. Instead of acoustic treatment by expensive and space-reserving absorbing structures, modal equalization is based on DSP affecting the electric-to-acoustic reproduction chain. Several DSP-based techniques for modal equalization have been proposed recently and tested in performance. From a perceptual point of view, however, no clear picture on the importance of controlled temporal decay has been shown, although it is known that towards the lowest frequencies the human hearing becomes increasingly insensitive to temporal details. In the present study we conducted listening tests where only a single synthetic mode with increased decay time but magnitude-equalized response was used to find the JND threshold of excessive decay time. The main conclusion is that at typical listening levels and down to 100 Hz the modal decay time T_{60} is allowed to increase from about 0.3 seconds by 0.1 to 0.4 seconds, while at 50 Hz even decay times of up to two seconds do not make a noticeable difference.

1. INTRODUCTION

Improvement of sound reproduction in a room at low frequencies has recently drawn research interest from a new point of view that is based on signal processing rather than on acoustic treatment of the room. Careful control of low-frequency modal behavior in a room can become impractical or very expensive to implement because the physical size and cost of necessary absorbers increases rapidly with decreasing frequency. In contrast to this, active control by DSP becomes feasible as the wavelengths become long and the sound field develops less diffuse [1]-[5]. Methods to optimize the response at a listening position by selecting suitable loudspeaker locations have been proposed [6] but cannot fully solve the problem. Because of these reasons, there has been an increasing interest in active methods of sound field control at low frequencies.

Modal resonances in a room can become audible as timbre changes because they modify the magnitude response of the primary sound or, when the primary sound ends, because they are no longer masked [7, 8]. The ability to detect a modal resonance appears to be very dependent on the signal content. Olive et al. report that for mid and high frequencies low-Q resonances are more readily audible with continuous signals containing a broad frequency spectrum while high-Q resonances become more audible with transient discontinuous signals [8]. They also report detection thresholds for resonances for both continuous broadband sound and transient discontinuous sound.

At low Q values, antiresonances (notches) are as audible as resonances. Audibility of antiresonances reduces dramatically for wideband continuous signals as the Q value becomes high [8]. Detectability of resonances reduces by approximately 3 dB for each doubling of the Q value [7, 8] and low Q resonances are more readily heard with zero or minimal time delay relative to the direct sound [7]. Duration of the reverberant decay in itself appears an unreliable indicator of the audibility of the resonance [7] as audibility seems to be more determined by frequency domain characteristics of the resonance. While this fact is the basis of classical studies, the ideas of modal equalization emphasize the combined time-frequency viewpoint in considering responses of audio reproduction systems.

Several methods have been proposed for DSP-based modal equalization and for the estimation of modal parameters for the modification of the rate of modal decay [9]-[16]. Much less is done to compare their performance from objective or subjective viewpoints. In a recent study we tried it [18] and found some objective differences in modal parameter estimation techniques which might favor AR and ARMA estimation techniques (called ARMA in [18]) over non-parametric time-frequency techniques (called AMK in [18]), particularly for cases with closely-in-frequency spaced modes that need correction. In subjective listening comparison, however, we failed to show clear improvement by any modal equalization against magnitude-only equalization. Recently, a new modal equalization technique was proposed in [16], with improved ability to control the temporal decay details of system responses, but it has not so far been compared against any other method.

Based on these experiences and lack of understanding of the underlying perceptual phenomena at low frequences, we decided to approach the problem from a very elementary experimental setting. If there is only a single mode with excessively long decay time but magnitude being equalized, what is the just noticeable increase of the decay time over otherwise uniform decay/reverberation time that can be noticed? Before describing the experiment and presenting the results, a short overview of modal equalization methodology is given.

2. OVERVIEW OF MODAL EQUALIZATION

We define the modal equalization as a process that modifies the rate of modal decay. A modal resonance is represented in the z-domain transfer function as a pole pair with pole radius r and pole angle θ

$$H_{\mathrm{m}}(z) = \frac{1}{(1 - r\, e^{j\theta} z^{-1})(1 - r\, e^{-j\theta} z^{-1})} \qquad (1)$$

To shorten the decay time the Q value of the resonance must be decreased by moving the pole pair closer to the origin. Each mode is modeled by an exponential decay function

$$h_{\mathrm{m}}(t) = A_{\mathrm{m}} e^{-\tau_{\mathrm{m}} t} \sin(\omega_{\mathrm{m}} t + \phi_{\mathrm{m}}) \qquad (2)$$

where A_{m} is the initial envelope amplitude of the decaying sinusoid, τ_{m} is a coefficient that denotes

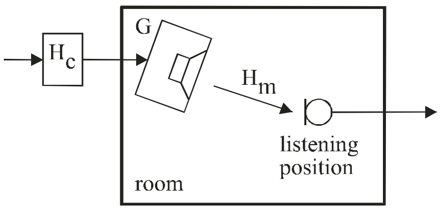

Fig. 1: A basic set-up for modal equalization.

the decay rate, ω_m is the angular frequency of the mode, and ϕ_m is the initial phase of the oscillation.

One method of implementing modal equalization is to modify the audio signal fed into a loudspeaker (Fig. 1). The total transfer function from the sound radiator to the listening position represented in z-domain is

$$H(z) = G(z)\,H_\mathrm{m}(z) \tag{3}$$

where $G(z)$ is the transfer function of the sound radiator from the electrical input to acoustical output and $H_\mathrm{m}(z)$ is the transfer function of the path from the sound radiator to the listening position.

The task in modal equalization is to design a correction filter $H_\mathrm{c}(z)$ in Fig. 1 so that the decay rate τ_m of a given mode is changed to yield a desired (shortened) decay time. If we assume that the loudspeaker response $G(z)$ does not contribute essentially to the decay time, the modal correction filter is a cascaded filter

$$H_\mathrm{c,m}(z) = \frac{(1 - r\,e^{j\theta} z^{-1})(1 - r\,e^{-j\theta} z^{-1})}{(1 - r_\mathrm{d} e^{j\theta} z^{-1})(1 - r_\mathrm{d} e^{-j\theta} z^{-1})} \tag{4}$$

where the numerator cancels out the original mode and the denominator creates a replacing mode with desired decay rate, controlled by pole radius r_d.

Modal equalization is applicable in practice at frequencies below 200 Hz and within a restricted listening area. The main techniques to estimate the modal parameters for correction are discussed in detail in [10]. A short characterization is as follows.

2.1. Decay estimation by time-frequency analysis

Short-time spectral analysis is often used to find the decay rate of isolated modes in a way resembling the analysis of reverberation time [13]. Modal frequencies are found as peaks in the power spectrum of a measured impulse response. Figure 2 shows a waterfall plot of a response with five synthetic modes. The decay time can be obtained by fitting a line to the temporal evolution of a modal spectrum peak as represented on a dB scale.

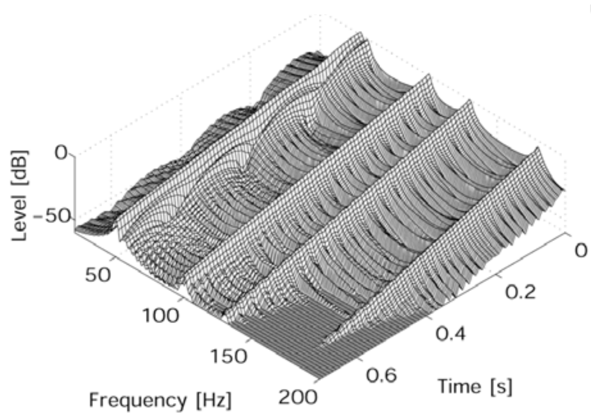

Fig. 2: Waterfall plot of five synthetic modes.

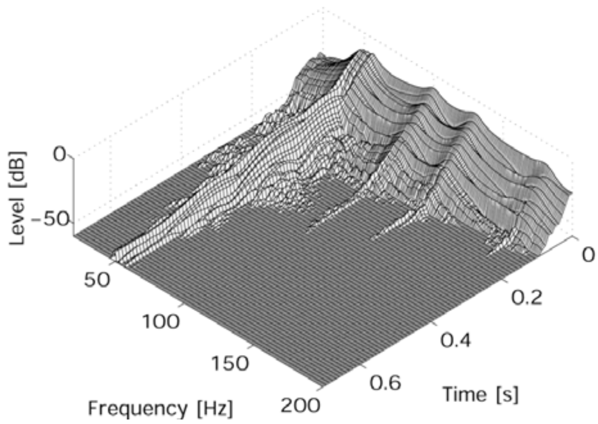

Fig. 3: Waterfall plot of five synthetic modes after modal equalization.

Problems with this approach (called AMK in [18]) appear when two or more modes are close in frequency, as can be seen around 50 Hz in Fig. 2. Figure 3 plots a case where this response is equalized to yield shorter decay times of the modes. Due to errors in modal parameter estimation the correction around 50 Hz works only for the early decay.

2.2. ARMA analysis of modal parameters

The second method (called ARMA in [18]) identifies the pole pairs describing modal resonances by

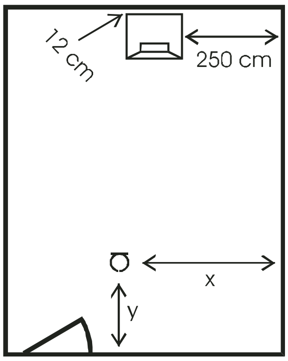

Fig. 4: Experiment setup in the HUT listening test room.

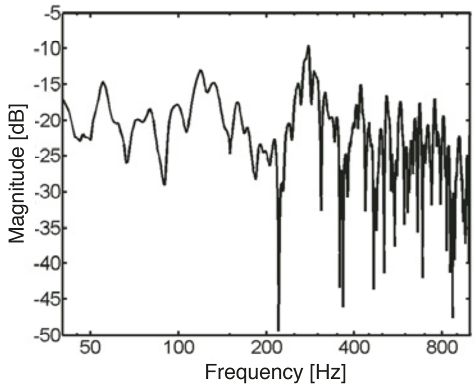

Fig. 5: Magnitude response in the listening position.

fitting a pole-zero least-squares model directly to a measured room impulse response. Pole pairs can be searched for iteratively, by finding the pair with the longest decay time and removing it from the measured response, and then repeating this iteratively.

A variant of ARMA estimation is studied in [11] and [12] whereby the frequency range of interest is scanned through in subbands in order to improve frequency resolution and robustness of finding poles closely spaced in frequency.

2.3. Non-parametric modal equalization

In [16] we have proposed a new method to modify in detail the temporal properties of a given impulse response. While the two methods mentioned above were based on the estimation of isolated modes and their parameters, the new method is a technique to change the time-domain response more directly. It is an advanced windowing technique where the temporal shaping of a given impulse response can be carried out in a frequency-dependent manner.

3. LISTENING TEST FOR PERCEPTION OF SINGLE EXCESSIVE MODE DECAY

Subjective tests carried out in our previous study [18] did not provide results that could be used as guidelines for human perception of low-frequency modal decay at low frequencies. Thus we decided to try a more analytic experiment with simple and well controlled responses. The simplest case, which actually can appear also in practice, is a single mode with excessively long decay time but the response otherwise having a relatively uniform decay/reverberation time.

Such a case can be created for a listening experiment by adding a synthetic mode to a real room response. There is a perfect mode-corrected reference available then: the room response itself! When a synthetic mode is added, the result can be equalized with different methods for A/B-comparison.

3.1. Experiment setup

The test requires a room with good modal behavior, such as a carefully realized listening room. We decided to use the listening test room of the HUT Laboratory of Acoustics and Audio Signal Processing, which is designed for multichannel reproduction (ITU-R BS.1116) with dimensions 6.25m x 5.66m x 2.95m [19]. The magnitude response of the room was measured with the setup of Fig. 4 to several points, and the microphone point shown was selected as the listening point (x = 413 cm, y = 135 cm) having a relatively flat magnitude response below 200 Hz, as shown in Fig. 5. Loudspeaker (Genelec 1032A) was elevated 41 cm from the floor and the listener's ear was assumed to be 92 cm from the floor.

The reverberation/decay time of the room was estimated at low frequencies from a measured impulse response by the Lundeby method [20] obtaining values shown in Table 1. (Shorter times were given at 100 and 200 Hz in the original documentation of the listening room.)

Table 1: Reverberation/decay time of the listening room as a function of frequency.

F/Hz	50	100	200	400	800
T_{60}/s	0.5	0.43	0.43	0.26	0.28

3.2. Listening test

The listening test was organized so that the test subject was sitting on a chair in the listening room as pointed by Fig. 4. Test samples were played through the GuineaPig2 (GP2) listening test software [21, 22]. Instructions and information from GP2 to guide the test subject were projected to the room wall by a video projector, and the subject responded by a computer mouse.

The task given to the subjects was to compare a reference sound sample which did not include any synthetic mode against a test sound sample that may or may not include a synthetic mode. Samples were magnitude equalized so that the difference, if any, was primarily due to the temporal decay in the cases of added synthetic mode.

Sound samples were preprocessed by Matlab scripts. The synthetic mode, which was a cascaded biquad filter (pole pair plus zero pair), was varied in resonance frequency and related decay time. Discrete frequencies 50, 100, 200, 400, and 800 Hz were selected as the center frequencies of the synthetic mode, thus covering also mid frequencies. The decay time was varied for the 50 Hz mode between 0.5 and 1.9 seconds by steps of 0.2 seconds. For the other modal frequencies the range of decay variation was scaled according to the mean value of the reverberation/decay time of the room for the modal frequency. So for example in the case of the 800 Hz mode the decay time was varied between 0.28 and 1.06 seconds by steps of 0.11 seconds.

Both the original response and the one with a synthetic mode were magnitude equalized in the following way. First the magnitude responses were 1/3-octave smoothed. Then an FIR equalizer with lowest order to fit within ± 3 dB to the smoothed response was designed by the Remez algorithm in Matlab. The equalizer was applied to obtain the magnitude-equalized response. This process has only a minor effect on long decay times.

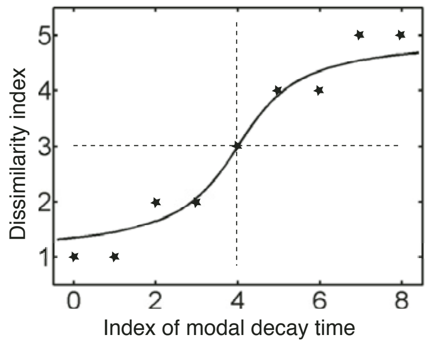

Fig. 6: Example of fitting the S-curve function to test response data ($a = 4.0$).

Finally, the magnitude-equalized responses were used to convolve with test signals to yield the samples to be playd to the subject. The test signals contained the following cases:

1. Short sample of male speech in German
2. Burst of bandpass noise (20–200 Hz), 0.5 s
3. Drum hit sound repeated twice
4. Short excerpt of rock music (medium heavy)

The subjects were listening to randomized pairs of sample set 1–4. They had a possibility to listen to the samples of a pair as many times they wished. Then they had to give a rating about the perceived difference on a 5-point scale:

1. = "no" (no difference)
2. = "maybe no"
3. = "can't say"
4. = "maybe yes"
5. = "yes" (clear difference)

There were total of 50 pairs in randomized order. Eleven subjects with normal hearing attended the test. The subjects had a little training period and they were told that some pairs consist of identical samples, encouraging them to respond "no" when they really did not hear any difference. Sound level was set to approximately 75–80 dB.

4. LISTENING TEST RESULTS

It can be expected that the ratings in noticing a difference will make an S-type curve as a function of decay time index. Such a curve can be approximated by a properly scaled and shifted function $y = \arctan(x)$ as characterized in Fig. 6. More specifically, we fitted the function

$$f(x) = \frac{4}{\pi}\arctan(b(x-a)) + 3 \qquad (5)$$

to the data so that the mean square error between response data and the function was minimized for parameters a and b. Here a is the shift and b is the scaling in the x-axis direction for an optimal fit. For the mid-point of the transition function, assumed to be symmetric, it holds

$$f(a) = 3 \qquad (6)$$

We defined this value as the threshold of noticeable difference. The steepness parameter b was of less interest here.

Figure 7 plots the results of data analysis. The curve fitting was first applied to each listener and each modal frequency data, yielding individual thresholds of difference noticeability in the form of decay time indices for each modal frequency. The indices were then mapped to decay times. The JND thresholds of modal decay time of the synthetic mode vs. no synthetic mode are plotted by dots in Fig. 7, each dot being a single listener data point in the test. The solid line denotes the average over subjects and the dashed line the corresponding median value. The lower dotted line shows the estimate of decay/reverberation time of the room itself and the upper dotted line shows the maximum synthetic decay time used in the test.

From comparing the average or the median curve against the lower dotted curve in Fig. 7 we find that the JND threshold of excess value for single-mode decay-time remains quite constant (0.2–0.3 seconds) down to 100 Hz, maybe showing a slight increase down in frequency. Below 100 Hz the value explodes so that at 50 Hz in these listening conditions the subjects could not notice systematically any difference within the given samples even for a synthetic mode decay time of almost two seconds[1]. Recall that the responses to be compared were magnitude equalized as described above. Without such equalization the difference can be noticed much more easily, which indicates the obvious fact that the frequency domain properties of responses are prominent over the time domain properties.

[1] The reason why curves at 50 Hz explode beyond the range of decay time used in stimuli is that the curve fitting extrapolates to very large values when subjects respond 'no' to almost all decay cases.

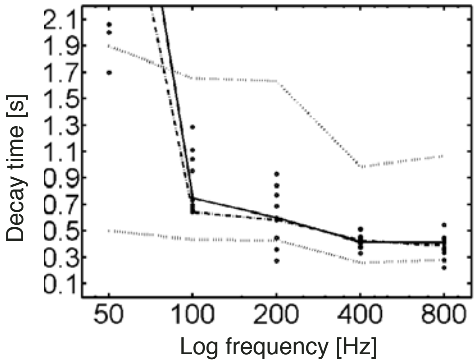

Fig. 7: Listening test results for the noticeability of single mode excess decay time. X-axis: frequency of the synthetic mode; Y-axis: modal decay time. Dots are data points for 11 test subjects. Solid line: average over subject data points. Dashed line: median over subject data points. Lower dotted line: decay/reverberation time of the room. Upper dotted line: maximum decay time of synthetic mode used in the test.

5. DISCUSSION AND CONCLUSIONS

The experiment shows clearly that the human auditory perception of temporal details in the given conditions degrades rapidly below 100 Hz. Such tendency may be slightly noticeable already from mid frequencies to 100 Hz. It can be concluded that down to about 100 Hz the temporal properties of modal decay are somewhat critical. This means that while it is of primary importance to perform magnitude equalization, also excessively long decay times should be corrected acoustically or by modal equalization techniques. Below 100 Hz, particularly at or below 50 Hz even very long decays, up to 2 seconds and above, may not be noticeable as far as the magnitude response is equalized well enough. This makes it questionable if modal equalization is worth of applying at the very lowest frequencies except in extreme cases.

An explanation to this frequency-dependent temporal resolution of the human auditory system can be searched for from several known phenomena. First, if we assume that the frequency resolution follows the ERB (equivalent rectangular bandwidth) scale [23], i.e., auditory analysis bandwith narrows for decreasing frequencies down to and below 100 Hz, then the corresponding auditory temporal resolution must get worse due to the time-frequency uncertainty principle. Secondly, the equal loudness curves [24] get denser (in dB) at low frequencies, which makes a slow decay correspond to a faster decay at higher frequencies. One may also consider that the auditory system learns to be less critical at low frequencies where the modal behavior in different positions of a room can vary more than at mid-to-high frequencies.

The experiment carried out in this study has to be taken as a tentative one, calling for more thorough investigations. Among remaining questions are for example how the basis level of reverberation/decay time of a room affects the JND of excessive decay time for a single mode, how a group of such modes is perceived, and how the listening level affects the perception (one may guess that increased level could make long decays more easily noticeable). A further questions is how binaural listening conditions with interaural difference may effect the perception. A set of carefully conducted listening experiments could throw light upon these questions. Finally, if the results of such experiments are consistent, a computational auditory model could be devised that is able to predict the noticeability of time-frequency degradations in low-frequency audio reproduction.

6. ACKNOWLEDGMENTS

This work is related to the VÄRE/TAKU project (Control of Closed Space Acoustics) that was funded by TEKES (the National Technology Agency of Finland.) The work was also funded by the Academy of Finland projects 201050 (SARA) and 53537.

7. REFERENCES

[1] S. J. Elliott and P. A. Nelson, "Multiple-Point Equalization in a Room Using Adaptive Digital Filters," *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899-907 (1989 Nov.).

[2] R. P. Genereux, "Adaptive Loudspeaker Systems: Correcting for the Acoustic Environment," in *Proc. AES 8th Int. Conf.*, (Washington D.C., 1990 May), pp. 245-256.

[3] J. Mourjopoulos and M. A. Paraskevas, "Pole and Zero Modelling of Room Transfer Functions," *J. Sound and Vibration*, vol. 146, no. 2, pp. 281-302 (1991).

[4] J. Mourjopoulos, "Digital Equalization of Room Acoustics," preprint 3288, *AES 92nd Conv.*, (Vienna, Austria), 1992 March.

[5] S. J. Elliott, L. P. Bhatia, F. S. Deghan, A. H. Fu, M. S. Stewart, and D. W. Wilson, "Practical Implementation of Low-Frequency Equalization Using Adaptive Digital Filters," *J. Audio Eng. Soc.*, vol. 42, no. 12, pp. 988-998 (1994 Dec.).

[6] A. G. Groh, "High-Fidelity Sound System Equalization by Analysis of Standing Waves," *J. Audio Eng. Soc.*, vol. 22, no. 10, pp. 795-799 (1974 Oct.).

[7] F. E. Toole and S. E. Olive, "The Modification of Timbre by Resonances: Perception and Measurement," *J. Audio Eng. Soc.*, vol. 36, no. 3, pp. 122-141 (1988 March).

[8] S. E. Olive, P. L. Schuck, J. G. Ryan, S. L. Sally, and M. E. Bonneville, "The Detection Thresholds of Resonances at Low Frequencies," *J. Audio Eng. Soc.*, vol. 45, no. 3, pp. 116-127 (1997 March).

[9] A. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Välimäki, "Low-Frequency Modal Equalization of Loudspeaker-Room Responses," Preprint 5480, *AES 111th Conv.*, (New York, USA), 2001 Nov./Dec.

[10] A. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Välimäki, "Low-Frequency Modal Equalization of Loudspeaker-Room Responses," *J. Audio Eng. Soc.*, vol. 51, no. 5, pp. 324–343 (2003 May).

[11] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "AR/ARMA Analysis and Modeling of Modes in Resonant

and Reverberant Systems," Preprint 5590, *AES 112th Conv.*, (Munich, Germany), 2002 May.

[12] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Frequency-Zooming ARMA Modeling of Resonant and Reverberant Systems," *J. Audio Eng. Soc.*, vol. 50, no. 12, pp. 1012-1029 (2002 Dec.).

[13] M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of Modal Decay Parameters from Noisy Response Measurements," *J. Audio Eng. Soc.*, vol. 50, no. 11, pp. 867-878 (2002 Nov.).

[14] M. R. Avis, "Q-factor Modification for Low-frequency Room Modes," *Proc. AES 21st Int. Conf.*, (St Petersburg, Russia), 2002 June.

[15] M. R. Avis, "IIR Biquad Controllers for Low Frequency Acoustic Resonance," Preprint 5474, *AES 111th Conv.*, (New York, USA), 2001 Nov/Dec.

[16] M. Karjalainen, P. Antsalo, and A. Mäkivirta, "Modal Equalization by Temporal Shaping of Room Response," *Proc. AES 21st Int. Conf.*, (St Petersburg, Russia), 2002 June.

[17] M. D. Capp, J. R. Stuart, and R. Wilson, "The Loudspeaker-Room Interface — Controlling Excitation of Room Modes," *Proc. AES 23rd Int. Conf.*, (Copenhagen, Denmark), 2003 May.

[18] P. Antsalo, M. Karjalainen, A. Mäkivirta, and V. Välimäki, "Comparison of Modal Equalizer Design Methods," Preprint 5844, *AES 114th Conv.*, (Amsterdam, The Netherlands), 2003 March.

[19] A. Järvinen, *Kuunteluhuoneen suunnittelu ja toteutus*, (Design and realization of a listening test room). Master's Thesis (in Finnish), Helsinki University of Technology, Espoo, Finland (1999). Available on-line at http://www. acoustics.hut.fi/publications/.

[20] A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, "Uncertainties of Measurements in Room Acoustics," *Acustica*, vol. 81, pp. 344–355 (1995).

[21] J. Hynninen and N. Zacharov, "GuineaPig - a Generic Subjective Test System for Multichannel Audio," *AES 106th Conv.*, Munich, (preprint 4871), 1999 May. See also http://www.acoustics.hut.fi/projects/GuineaPig2.

[22] J. Hynninen. *A Software-based System for Listening Tests*. Master's Thesis, Helsinki University of Technology, Espoo, Finland (2001). Available on-line at http://www. acoustics.hut.fi/publications/.

[23] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "Auditory Filter Shapes at Low Center Frequences," *J. Acoust. Soc. Am.* vol. 88, pp. 132–140, (1990 July).

[24] ISO standard 226:1987, "Acoustics – Normal Equal-Loudness Level Contours".