Requirements for Low-Frequency Sound Reproduction, Part II:

Generation of Stimuli and Listening System Equalization*

JAN ABILDGAARD PEDERSEN, *AES Member*

Research & Development, Bang & Olufsen a/s, DK-7600, Struer, Denmark

AND

AKI MÄKIVIRTA, *AES Member*

Genelec Oy, FIN-74100 lisalmi, Finland

In part I of two papers the requirements for low-frequency sound reproduction were investigated by the variation of lower cutoff frequency and slope and by the introduction of different levels of amplitude ripple and group delay ripple in the passband of a high-performance sound reproduction system. Listening tests were performed at three different sound pressure levels using both loudspeakers in an anechoic chamber and headphones in an audiometric booth. Two reproduction setups were used to confirm that equal results of the listening tests could be obtained in the two cases when proper equalization was implemented.

It is described how DSP was used to generate stimuli and perform equalization of the two reproduction setups. The shape and magnitude of amplitude and group delay ripple were derived from room simulations of an IEC 268-13 sized room with varying reverberation time. Proper equalization included the introduction of head-related transfer functions in the signal path to the headphones. This ensured that the sound pressures at the ear drums were very similar in two cases—a person sitting in front of the loudspeakers in the anechoic chamber and a person wearing headphones in the experimental booth. Level calibration was performed on both setups using pink noise. The nonlinearities measured in the physical loudspeakers were introduced into the signal path to the headphones using a nonlinearity simulator program.

0 INTRODUCTION

The aim of this work was to study the psychoacoustically determined performance requirements for the design of a loudspeaker system for high-quality reproduction of low-frequency sound in small rooms.

This is part II in a series of papers, and it describes how digital signal processing was used for the generation of stimuli and the equalization of the listening system. Part I described the actual listening experiments, including experimental strategy, procedure, and results [1].

The increasing use of digital surround sound systems, such as 5.1 systems, increases the attention to low-frequency reproduction. Considering 5.1 systems, especially the .1 channel and modern program material, loudspeaker manufacturers have the desire to quantify the requirements for low-frequency sound reproduction.

* Manuscript received 2001 April 27; revised 2002 March 22.

In order not to limit the applicability of the conclusions it was decided to assume that the system variables were those available in a system employing digital signal processing (DSP) for all the filtering and control tasks.

The essential requirements for high-quality sound reproduction at low frequencies are the lower cutoff frequency and slope because these are linked to maximum displacement of drivers, system tuning, and power requirements. In this context it was also decided to focus on two additional parameters—amplitude ripple and group delay ripple, both in the passband. Amplitude ripple sets the requirements for how much the amplitude response can be allowed to deviate from an ideal response. Requirements regarding maximum group delay ripple impose a limit on how rapidly the phase is allowed to change. Both amplitude and delay ripple can arise due to a number of physical phenomena such as variations in the parameters of the loudspeaker drive units, resonances in the cabinet, diffraction around the cabinet, reflections, and modes in the listening room. The experiments were divided into two separate groups—lower cutoff frequency and slope in the first group and amplitude and group delay ripples in the second group. All parameters in each group were required to be varied independently, that is, orthogonally.

The work reported both here and in paper I formed part of the EUREKA Project 1653 Medusa (Multichannel Enhancement of Domestic User Stereo Applications). The Medusa project was a 3.5-year joint research project with the following partners: British Broadcasting Corporation, Institute of Sound Recording at University of Surrey, Nokia Research Centre, Genelec Oy, and Bang & Olufsen a/s.

1 GENERATION OF STIMULI

Stimuli for the listening tests were produced by different filtering of seven selected program pieces. First the lower cutoff frequency and slope were changed and then both amplitude and group delay ripples in the passband were added.

1.1 Program Material

Seven items of program material were selected from a wide set of materials [1]. Two items were used during the training of subjects, whereas four other items were used in the actual experiment. The last item was used for calibration purposes only. The spectral distribution, especially below 200 Hz, was of special interest when selecting the seven items of program material. Table 1 lists the specifications for the selected seven pieces.

Total loudness at the listening position was used to calibrate the playback level. The calculation of total loudness required a stationary signal, which was chosen to be pink noise, as seen in Table 1. The reproduction level of pink noise was adjusted from loudness measurements, which are described in Sections 3.1 and 3.3, and the levels of the other six pieces were individually adjusted relative to pink noise, based on calculations of the total power content of the seven pieces. Eq. (1) gives the relative gain G_A of program item A relative to pink noise,

$$G_A = \sqrt{\frac{\sum_{n=0}^{N-1} \left|\mathrm{pink}[n]\right|^2}{\sum_{n=0}^{N-1} \left|A[n]\right|^2}} = \sqrt{\frac{P_{\mathrm{pink}}}{P_A}} \tag{1}$$

where pink $[n]$ is the time signal of pink noise and $A[n]$ the time signal of program item A, N is the number of samples in the program pieces, P_{pink} is the total power content of pink noise, and P_A is the total power content of program item A. The total power content is given in Table 1 for each of the seven pieces.

Power spectra of all the program items were calculated using the 1 min of each program item. Fig. 1 gives the power spectra of the two training items and Fig. 2 those of the four items that were used in the actual experiment. The original program items were turned into Microsoft wave files with two channels, 16 bits, and 44 100-Hz sampling frequency. All the programs for processing had Microsoft wave files as both input and output files.

1.2 Variable Lower Cutoff Frequency and Slope

The lower cutoff frequency was varied from 20 to 50 Hz by prefiltering the program material using a digital high-pass filter with variable cutoff frequency. The range from 20 to 50 Hz was considered to be a typical range of lower cutoff frequencies. An intermediate value of 35 Hz was chosen as a third cutoff frequency. Additional information about the experimental plan was presented in part I [1]. Fig. 3 illustrates the variation of the lower cutoff frequency.

It is not only the cutoff frequency that is of importance when considering the lower rolloff of a loudspeaker system, but also the order, that is, slope. A closed-box system is of second order and a vented box is of fourth order. Considering these very commonly used systems, it was chosen to vary the order of the high-pass filter between second, fourth, and sixth order. Fig. 3 shows this variation of the slope from second order up to sixth order.

Table 1. Specifications of selected program material.*

Title	Main Artist	Record Company	Number	Track	Time	Power† (dB)
Rush	Eric Clapton	Reprise	926794-2	2	0:00 1:00	−18.8
Fourplay	Fourplay	Warner Bros.	7599-266565-2	10	0:25 1:25	−12.9
Ray of Light	Madonna	Maverik/Warner Bros.	9362-46847-2	5	0:40 1:40	−8.9
Bladerunner	Vangelis	East West	4509-96574-2	1	1:00 2:00	−16.2
The Hunter	Jennifer Warnes	BMG	261974	9	3:40 4:40	−13.9
Amused to Death	Roger Waters	Columbia	468761-2	3	0:05 1:05	−25.3
Pink noise					0:00 1:00	−9.7

* The first two programs were used during training and the following four during the actual experiment. Pink noise was used for calibration.
† Power averaged over the whole duration of each program item. Only relative levels were necessary for level alignment of program pieces.

1.3 Amplitude and Group Delay Ripple

Ripples both in amplitude response and in the group delay are found in practical usage of loudspeaker systems. Ripples can arise for a number of reasons, such as tolerances of components, limited possibilities of equalization, poor tuning of the system, diffraction of the cabinet, reflections, and modes in the listening room. When designing and installing a loudspeaker system, it would be useful to have knowledge about how large the ripples are allowed to be. Ripples of different magnitudes were added to the test items before playback by introducing a digital filter com-

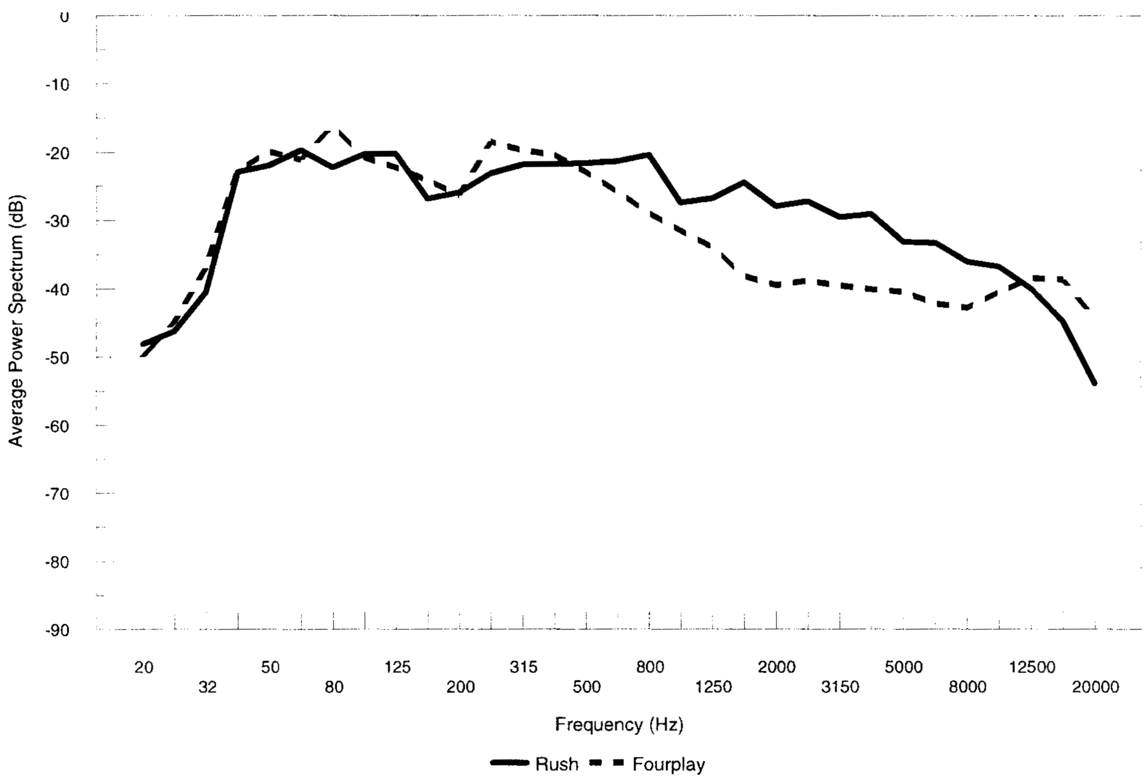

Fig. 1. Power spectrum (one-third) of first two music segments for training (aligned).

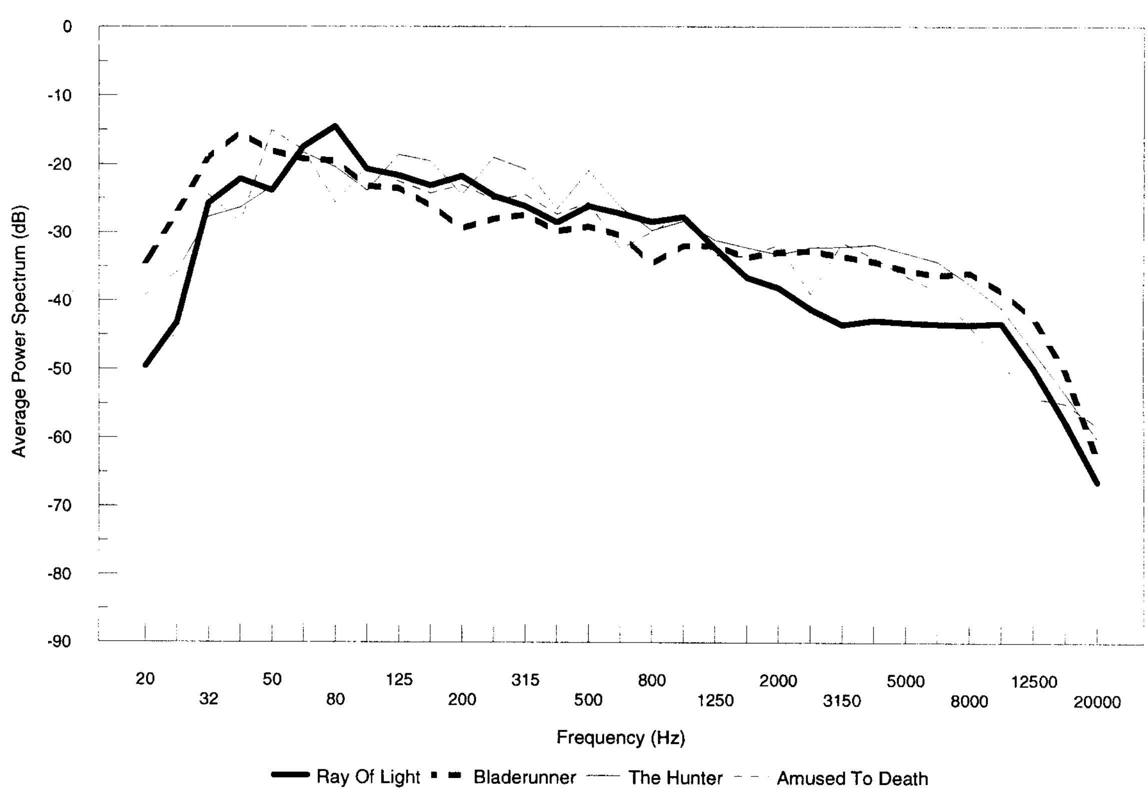

Fig. 2. Power spectrum (one-third) of four music segments for actual test (aligned).

prising a nonflat amplitude response or a nonlinear phase response. A nonflat amplitude response in the passband of the loudspeaker yielded amplitude ripple and a nonlinear phase response yielded a nonconstant group delay, that is, group delay ripple.

The shape and magnitude of both amplitude and group delay ripple were calculated from room simulations of an IEC 268-13 sized room [2]. The room simulations were based on the principle of mirror images and carried out using a developed program [3]. The room simulator program was used to calculate the first 1000 ms of the impulse response in a specified receiver position, given a specified source position. Both source and receiver were assumed to be omnidirectional. The impulse response was sampled at a rate of 4 kHz, which was appropriate considering the bass region of a loudspeaker system.

The simulated room was a rectangular room with dimensions of 5.03 m × 6.02 m × 2.50 m (width × length × height). The source position was chosen to be $(X, Y, Z) = (0.30\ \mathrm{m}, 0.40\ \mathrm{m}, 0.35\ \mathrm{m})$, which was considered to be a realistic corner position that would excite all room modes, which in turn would yield large ripples. The receiver position was chosen to be at a typical listening position near the center of the room, but offset about 0.30 m to avoid abnormal effects due to perfect symmetry, such as perfect cancellation: $(X, Y, Z) = (2.20\ \mathrm{m}, 2.70\ \mathrm{m}, 1.10\ \mathrm{m})$.

These parameters of the room simulation were fixed in all simulations, but the reverberation time T_{60} was varied from an anechoic environment to 0.8 s in order to control the magnitude of the ripples, that is, a hard room yielded large ripples compared to a well damped room. The reverberation times of the simulated rooms were estimated by time-reversed integration of the squared impulse response. Eq. (2) was used to calculate a Schroeder integral plot $rp[n]$, based on the impulse response $h[n]$, which was of length N,

$$rp[n] = \sum_{i=n}^{N} h[i]^2 . \tag{2}$$

Further information about time-reversed integration can be found in [4]. The reverberation time T_{60} was found through inspection of a plot of the Schroeder integral by determining the time where the level was decreased 60 dB relative to the initial level. Note that the length of the simulated impulse response was longer than the true reverberation time in order to avoid errors in the Schroeder integration.

Four levels of ripples were calculated by simulating four rooms with different reverbation times T_{60}: 0 s (anechoic), 0.2 s, 0.4 s, and 0.8 s. The reverberation time was adjusted by changing the reflection coefficients of the walls, floor, and ceiling. Table 2 lists the reflection coefficients for the six surfaces in each of the four rooms. The amplitude and phase responses of each room were normalized according to the anechoic room, that is, the anechoic case yielded zero amplitude ripple and zero group delay ripple. The group delay was calculated from the unwrapped phase response by differentiation. Figs. 4–6 show all four levels of amplitude ripple and group delay ripple. Please note that ripples were only implemented from these target curves up to a certain upper frequency, that is, to emulate the upper frequency limit of a sub-

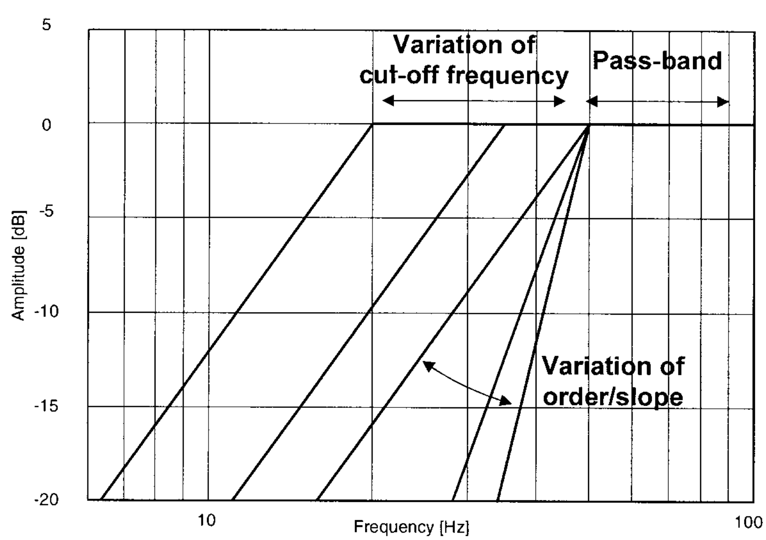

Fig. 3. Variation of cutoff frequency and slope.

Table 2. Reflection coefficients of walls, floor, and ceiling

T_{60}	Walls	Floor	Ceiling
Anechoic	0.000	0.000	0.000
0.2 s	0.650	0.440	0.290
0.4 s	0.805	0.540	0.360
0.8 s	0.895	0.600	0.400

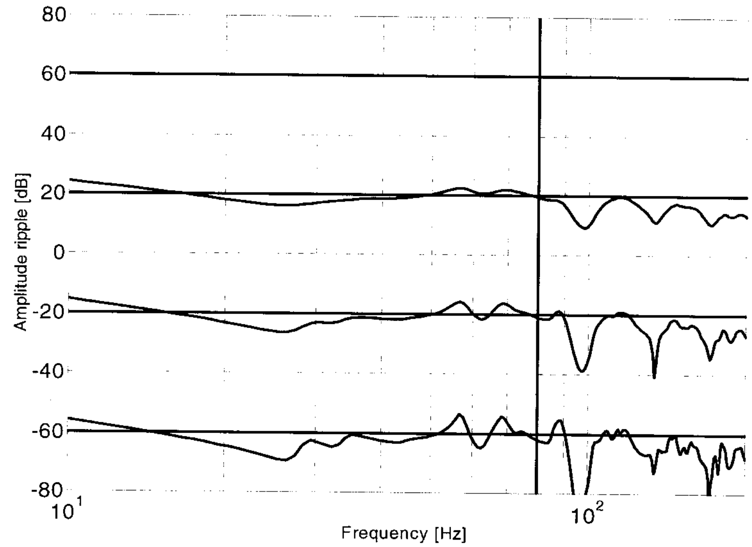

Fig. 4. Amplitude ripple implemented up to 80 Hz in four cases. From top—anechoic, T_{60} = 0.2 s, T_{60} = 0.4 s, T_{60} = 0.8 s. For clarity, 0-dB level has been shifted to −60 dB, −20 dB, 20 dB, and 60 dB, respectively, as indicated by bold horizontal lines.

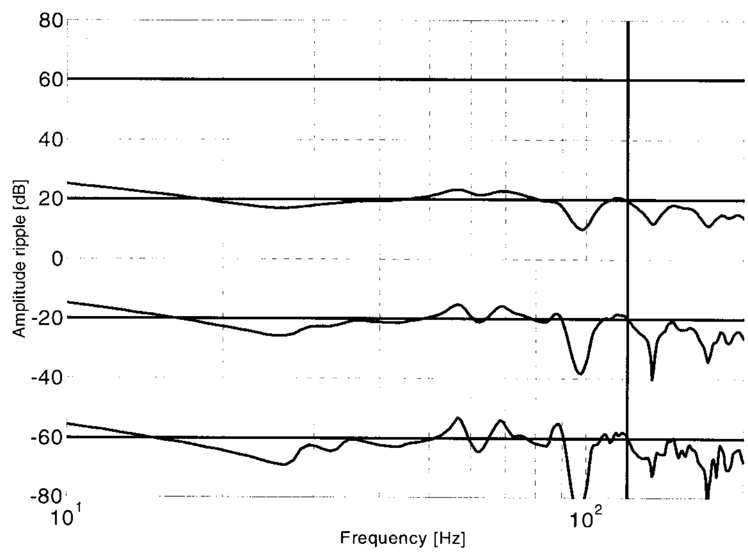

Fig. 5. Amplitude ripple implemented up to 120 Hz in four cases. From top—anechoic, T_{60} = 0.2 s, T_{60} = 0.4 s, T_{60} = 0.8 s. For clarity, 0-dB level has been shifted to −60 dB, −20 dB, 20 dB, and 60 dB, respectively, as indicated by bold horizontal lines.

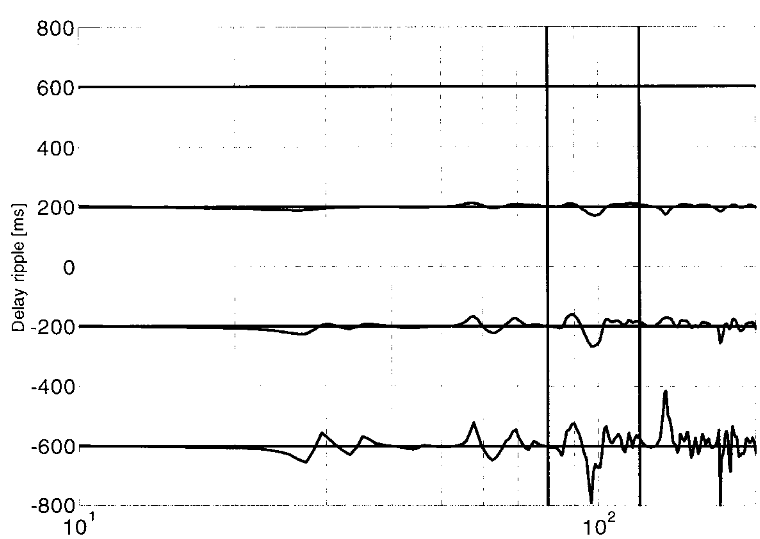

Fig. 6. Group delay ripple in four cases. From top—anechoic, T_{60} = 0.2 s, T_{60} = 0.4 s, T_{60} = 0.8 s. Both 80 and 120 Hz are marked as upper limits for implementing the ripples. For clarity, 0-s level has been shifted to −600 s, −200 s, 200 s, and 600 s, respectively, as indicated by bold horizontal lines.

woofer. Two upper frequency limits were implemented to investigate the influence of the range in which ripples were implemented: 80 Hz as indicated in Fig. 4, and 120 Hz as indicated in Fig. 5. Both 80 Hz and 120 Hz are indicated for the group delay ripple in Fig. 6. 80 Hz was considered to be a typical crossover frequency for a subwoofer and 120 Hz was the specified crossover frequency for the low-frequency enhancement (LFE) channel of a 5.1-channel system.

In order to minimize the influence of loudness differences between the four different amplitude ripple filters, it was decided to normalize the total power relative to the anechoic filter, that is, 0-dB filter. The power of each amplitude ripple filter was calculated in the frequency range from 20 to 80 Hz in Fig. 4 and from 20 to 120 Hz in Fig. 5. Then a pure gain was applied to the individual filters in order to obtain equal total power in the specified frequency ranges.

1.4 Implementation

Variation of the lower cutoff frequency and slope was implemented using digital high-pass filters with variable cutoff frequency of second, fourth, and sixth order according to the experimental design [1].

The amplitude and group delay ripple filters were designed based on the impulse responses produced by the simulated room response calculations found in Section 1.3. Only ripples in the bass range were included, meaning that ripples in amplitude and delay were only included up to a certain frequency, and no ripples existed above this frequency.

The experimental design [1] had three variables for the ripples: the limiting frequency below which the ripples exist (two values), the amount of amplitude ripple (four values), and the amount of delay ripple (four values).

Two filter designs for each value of the reverberation time were then produced—a linear phase filter, which contained the amplitude ripples but no delay ripples, and an all-pass filter, which contained just the delay ripples and no amplitude ripples. The delay filters were made causal by adding a constant delay.

The experimental design [1] used the amplitude ripple data from one reverberation time simulation, combined with the delay ripple data from any of the four reverberation times. This was implemented by convolving an amplitude filter with a delay filter, producing one filter for each of the amplitude/delay deviation combinations. The filters had a sampling rate of 4410 Hz. The filter length was 2.04 s (8996 samples) when it contained both amplitude and delay ripple.

The audio files containing the desired deviations were produced using a multirate convolution process. The audio data were crossover filtered at one of the two desired frequencies, the low-pass section was convolved with the ripple filters described, and the final signal was produced by summing the two signals with the necessary gain and delay alignment. For delay ripple filtering, the delay alignment could not be exact over the whole crossover frequency band, and this produced changes in the amplitude and delay properties at the crossover point (maximum 1.5 dB). The introduction of delay ripple also introduced some unwanted amplitude ripple. But the opposite was not the case, that is, the introduction of amplitude ripple did not introduce any unwanted delay ripple. It follows from this that amplitude ripple and delay ripple were not completely orthogonal, and as a consequence it was decided not to use the test results obtained from the stimuli where any delay ripple was introduced [1].

The convolution was performed in floating-point arithmetic, and the source and destination files were 16-bit fixed point, sampling frequency 44.1 kHz. The low-pass section was processed at a sampling frequency of 4410 Hz (decimation by 10). Distortion and noise levels were studied at each step of the process, and proper dithering was applied at quantizations to maintain linearity throughout the processing chain and to minimize noise.

2 EXPERIMENTAL SETUPS

The experiments were carried out using two different setups—loudspeakers in an anechoic chamber and headphones in quiet surroundings.

2.1 Setup in Anechoic Chamber

The processed audio samples were played back using a pair of reference loudspeakers built especially for this purpose. These reference loudspeakers extended down to 20 Hz ($\pm$2.5 dB) and kept the harmonic distortion below 2% (90 dB SPL at 1 m) at low frequencies, in the range from 20 to 50 Hz. Above 50 Hz, in the range from 50 to 250 Hz, the harmonic distortion was below 0.5% (90 dB SPL at 1 m). Above 50 Hz the group delay was kept below 10 ms, in the frequency range from 50 Hz to 20 kHz. The group delay was kept below 20 ms in the full frequency range, from 20 Hz to 20 kHz. It was an active three-way closed-box system. Crossover frequencies were 160 Hz and 1.6 kHz (fourth order). The separate power amplifiers for the three bands were 400 W, 160 W, and 120 W.

The setup comprised two reference loudspeakers and a listening position placed in a 3-m equilateral triangle in a large anechoic chamber, as shown in Fig. 7. The subjects were seated in a chair. Both loudspeakers and the chair were placed on grids that were held in place by poles

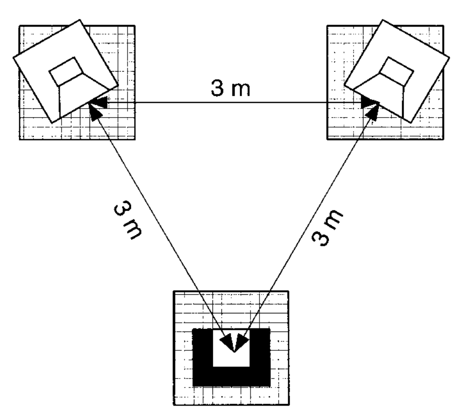

Fig. 7. Setup in anechoic chamber, comprised of two reference loudspeakers and a chair for listening. Both loudspeakers and chair were placed on grids held in place by poles extending to the bottom of the anechoic chamber.

extending to the bottom of the anechoic chamber. The anechoic chamber is located at the Technical University of Denmark, Department of Acoustic Technology. The free space in the anechoic chamber is about 1000 m^3. The three dimensions are of similar magnitude, but not equal, and the lower limiting frequency is 43.7 Hz. A detailed description can be found in [5].

A computer-based system for automated control of the listening tests was utilized. The system is known as GuineaPig [6], [7]. This software was run in a workstation (Silicon Graphics Octane) that included eight channels of 24-bit (ADAT) input and output. Graphical user interface and feedback from the subject were handled by a laptop PC connected to the workstation.

Control of analog level and conversion from digital to analog were performed using a digital mixing console (YAMAHA O3D). This mixer utilized 20-bit linear 8× oversampling digital-to-analog converters. Fig. 8 shows an overview of the entire signal path starting at the original program pieces located on the original Compact Discs (CDs). The program material from the original CDs was processed prior to the listening experiments and put onto the hard disk of the workstation. During the actual listening experiments the program material was taken from the hard disk of the workstation.

2.2 Headphone Setup

As an alternative to using an anechoic chamber, headphones were utilized in an attempt to simulate the setup in the anechoic chamber. A number of different headphones were tested and an electrostatic type was chosen (Sennheiser HE 60, electrostatic, amplifier Sennheiser HEV 70). This choice was motivated by a wide frequency response and low distortion.

Fig. 9 gives measurements of the distortion of the headphone at a maximum level of 100 dB SPL. The headphones were placed on a Brüel & Kjær head and torso simulator (B&K 4128), and measurements were taken using a Brüel & Kjær audio analyzer (B&K 2012). The head and torso simulator was mounted with hard ears (optional). Measurements took place in an anechoic chamber at Bang & Olufsen. The levels of both second and third harmonics were generally more than 75 dB (0.02%) below the fundamental, except for a number of frequencies where the level was 60 dB (0.1%) below the fundamental. The levels of both second and third harmonics rose to about 50 dB (0.3%) below the fundamental at 20 Hz.

During listening experiments, subjects were seated in quiet surroundings, an audiometric booth. The audiometric booth was located next-door to the anechoic chamber described in Section 2.1.

In an attempt to simulate the reference loudspeakers in the anechoic chamber, it was decided to include a nonlinear simulation of the reference loudspeakers, that is, distortion due to nonlinearities of the reference loudspeaker was added to the signals before playback using the headphones. Section 3.2 describes how a nonlinear simulator program was used to perform this simulation. The sound

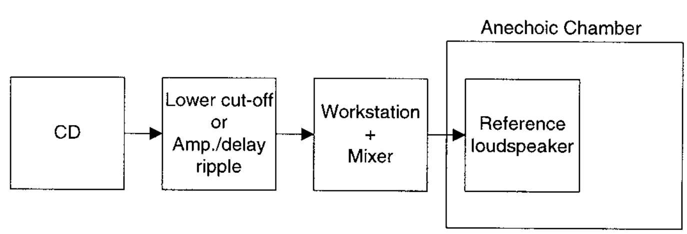

Fig. 8. Signal path from original CD to playback in anechoic chamber using reference loudspeakers. Program material from CD was processed prior to listening experiments and put onto hard disk of workstation.

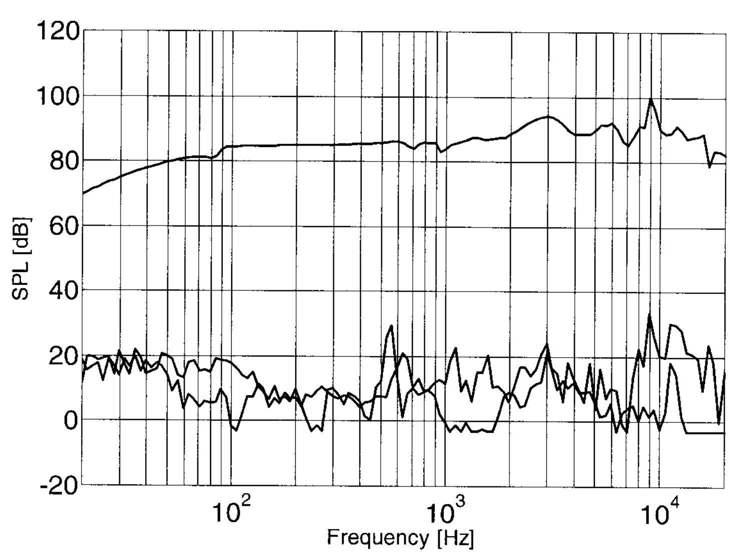

Fig. 9. Fundamental, second, and third harmonics measured with head and torso simulator and headphones used in the experiment. Maximum SPL at 9 kHz was 100 dB.

pressures at the ear drums in the two setups were made similar by applying advanced equalization to the signals before playback via the headphones. Section 3.3 describes this equalization in detail. Fig. 10 shows the signal path when headphones were used. The graphic user interface and feedback from the subject were handled directly by the workstation.

3 LISTENING SYSTEM EQUALIZATION AND CALIBRATION

The setup in the anechoic chamber was measured using pink noise, and calibration was performed. Both the nonlinearities of the loudspeaker and the difference in linear transfer functions of the headphones compared to listening in the anechoic chamber were taken into account in the equalization process.

3.1 Anechoic Chamber

The middle curve in Fig. 11 shows a measurement of the left reference loudspeaker placed in the anechoic chamber according to Fig. 7. The measurement was taken using a free-field microphone (B&K 4133), measurement amplifier (B&K 2690: NEXUS), and an audio analyzer (B&K 2012). The free-field microphone was placed at the listening position pointing directly toward the left reference loudspeaker. The first mode of the anechoic chamber is clearly seen in Fig. 11 at approximately 35 Hz, which was a frequency so low that the damping in the room was insufficient to be anechoic. This mode was unavoidable, so it was decided to leave the system like it was and take the response in Fig. 11 as part of the circumstances for the experiment.

During the listening experiment three different playback levels, spaced 10 dB between each level, were used. The highest level was found by applying full-scale pink noise, and adjusting the level to 2 dB below the level where clipping is indicated on the rear of the loudspeaker. Adjusted to this level, a sound pressure level of 69 dB (lin, slow) was measured using the free-field microphone at the listening position, when pink noise was applied to the left reference loudspeaker. In an IEC 268-13 standard listening room [2], this adjustment corresponded to approximately 80 dB (lin) sound pressure level.

Still using the free-field microphone and playing the pink noise, the loudness level was measured to 22 sones via a loudness meter (B&K 2144). This loudness meter was based on ISO 532B "Zwicker" loudness [8]. When a

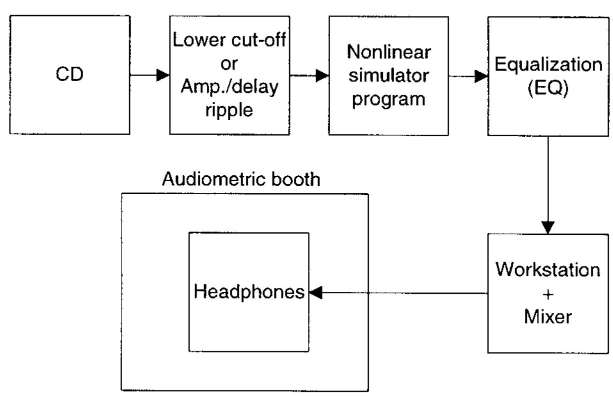

Fig. 10. Signal path from CD to playback in audiometric booth using headphones. Program material from CD was processed prior to listening experiments and put onto hard disk of workstation.

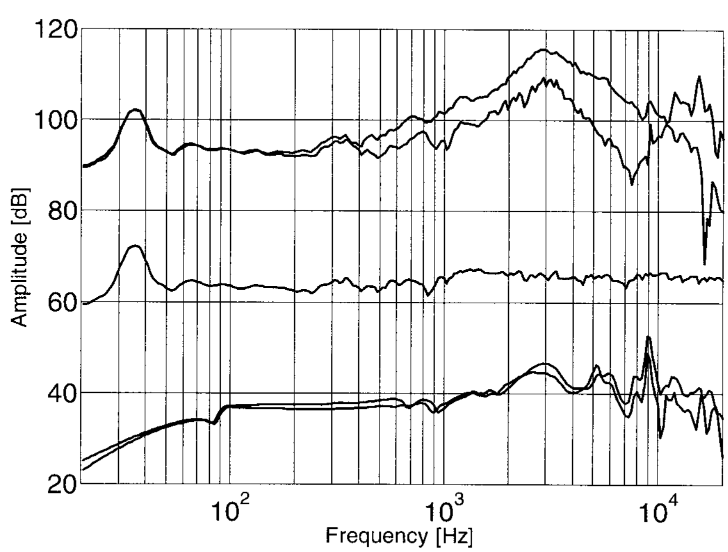

Fig. 11. Transfer functions. From top—left loudspeaker to left ear, left loudspeaker to right ear, left loudspeaker to free-field microphone, left headphone and right headphone both mounted on head and torso simulator. Curves were shifted apart to increase clarity. Free-field microphone measurement was lowered 30 dB relative to HRTFs, e.g., left loudspeaker to left ear.

1-kHz sine wave, full scale on a CD, was applied using this adjustment, 78.4 dB (lin) was measured via the free-field microphone. The two lower playback levels were simply 10 and 20 dB below this adjusted level.

3.2 Nonlinear Simulation

A nonlinear simulation of the reference loudspeaker was utilized when playback was performed using headphones. This ensured that distortion due to nonlinearities of the reference loudspeaker was added to the signals before playback using the headphones. This operation was performed using a nonlinear loudspeaker driver simulator program (LoDist) [9].

The central parameter in this nonlinear simulator program was the diaphragm position, that is, the traditional loudspeaker driver parameters were dependent on the diaphragm position. All relevant parameters were determined as a function of absolute displacement from the neutral position. Measurements were performed on a woofer driver identical to the ones used in the reference loudspeakers. Based on these parameters, the nonlinear simulator program solved numerically the differential equations that described the loudspeaker driver and its environment, such as its cabinet. More information about measuring position-dependent parameters, and the operation and principle of the nonlinear simulator program can be found in [9].

Care was taken to ensure that the electrical voltage applied to the simulated driver was identical to the voltage applied to the physical driver in the reference loudspeaker. This was achieved by applying appropriate scaling when relating the digital signal feed to the nonlinear simulator program and the corresponding analog voltage.

3.3 Headphones

The sound pressures at the ear drums in the two setups were made similar by applying advanced equalization to the signals before playback via the headphones. Head-related transfer functions (HRTFs) were measured in the anechoic chamber using the setup described in Section 2.1. HRTFs are described in [10].

The HRTFs were measured using the left reference loudspeaker and a head and torso simulator (B&K 4128) placed in the listening position in the anechoic chamber, facing a point just halfway between the loudspeakers. The head and torso simulator was used because it was assumed that individual HRTFs for each subject in the listening test were not necessary. This assumption was made because timbre was judged to be the important parameter, whereas exact location was judged to be a secondary issue. Especially at low frequencies, which was the important frequency range in this context, the individual HRTFs are very similar, which lead to the choice of using a standard head and torso simulator. Fig. 11 includes two HRTF measurements at the top of the figure—left loudspeaker to left ear (the uppermost curve) and left loudspeaker to right ear. The shadow effect of the head is obvious above 400 Hz, where left loudspeaker to left ear is much higher than left loudspeaker to right ear. Below 150 Hz both HRTFs coincide with the middle curve, the free-field microphone measurement, but the curves are shifted 30 dB apart for clarity. Perfect symmetry of the acoustic setup was assumed when right loudspeaker to right ear and right loudspeaker to left ear were obtained, but differences in the sensitivities of the two ears of the head and torso simulator were taken into account.

The selected headphones were placed on the head and torso simulator, and the transfer functions to each ear were measured. These amplitude responses are shown at the bottom of Fig. 11. As described in Section 3.2, a nonlinear simulator program was used to simulate the nonlinear behavior of a loudspeaker driver, but to do so accurately, it had to have the correct linear response as well, that is, the linear response of the reference loudspeaker. But the linear response of the reference loudspeaker was also included in the measured HRTFs—transfer functions from the input of the loudspeakers to the output of the head and torso simulator. For this reason an inverse filter of the linear transfer function of the nonlinear simulator program was included in the equalization in order to cancel one of the two linear transfer functions of the reference loudspeakers found in the signal chain.

The linear transfer function of the nonlinear simulator program was found by feeding an audio file, containing a digital Dirac impulse, through the nonlinear simulator program. This was done at a signal level that did not give rise to displacements of an order where nonlinearities come into play, such as a small fraction of 1 mm. Fig. 12 shows the determined amplitude response along with a bold curve, which has limited attenuation toward lower frequencies. This limitation below 15 Hz was used to avoid dynamic/numerical problems when the transfer function was inverted.

The placement of the digital filter can be seen in Fig. 10, where the filter is marked with EQ. The filter was designed from the measured HRTFs, the transfer function of headphones on the head and torso simulator, and the linear transfer function of the nonlinear simulator program. The digital filter was used to process the signals, which were meant for playback via the reference loudspeakers, in such a way that the output signals from the digital filter were ready for playback via headphones. This was performed so that the sound pressures at the ear drums of the head and torso simulator were identical in the two setups—headphones mounted on the head and torso simulator and the head and torso simulator placed in the anechoic chamber in front of the reference loudspeakers.

During calibration with pink noise, the function of the digital filter EQ was somewhat different. Section 3.1 described how the pink noise signal was played back via the reference loudspeakers and how the sound pressure was measured by a free-field microphone. In this situation the digital filter was used to process the pink noise signal so that the electrical output of the head and torso simulator wearing the headphones was identical to the electrical output of the free-field microphone in the anechoic chamber, when the loudspeakers were playing the unprocessed pink noise. This was necessary because the loudness meter used required an input signal from a free-field microphone in order to calculate the correct loudness level.

Fig. 13 summarizes all the considerations regarding the calculation of the digital equalizer filter EQ in the two cases—playback of program pieces, namely, using the measured HRTFs, and calibration with pink noise, namely, using the free-field microphone measurement.

When a human is listening to two loudspeakers, each of the two ears receives sound coming from both loudspeakers—crosstalk from both loudspeakers to both ears. Fig. 14 shows a schematic of how this was achieved using four individual filters for each equalizer EQ when the program pieces were processed, that is, the equalizer for the program pieces consisted of the four filters, H_{RR}, H_{RL}, H_{LR}, and H_{LL}.

H_{RR} was calculated from the ratio of the HRTF from the right loudspeaker to the right ear divided by the product of the transfer function of the right headphone and the linear transfer function of the nonlinear simulator program . H_{RL} was calculated using the HRTF from the right loudspeaker to the left ear, the left headphone, and the linear transfer function of the nonlinear simulator program. H_{LR} and H_{LL} were calculated accordingly. The equalizer for the pink noise did not have crosstalk, that is, only two equalizer filters, because the HRTFs were replaced by the free-field microphone measurements. Due to the assumption of perfect acoustic symmetry, the numerator in Fig. 13 was exactly the same for both right and left loudspeakers. The only difference was between the two headphones, that is, in the denominator in Fig. 13.

Fig. 15 shows all four filters that formed the equalizer for the program pieces, H_{RR} and H_{LL} being the two upper curves in the frequency range from 400 to 8000 Hz. Fig. 16 shows the two filters that formed the equalizer for the pink noise. All measurements were taken using 1/24 octave logarithmic frequency resolution, and the filters were designed using an FFT size of 65 536. This called for interpolation, which was performed in the complex plane after removal of the pure time delays of all the relevant transfer functions. After interpolation, all pure time delays were reintroduced to ensure the correct time responses. The final impulse responses of the equalizer filters were truncated after 10 000 samples. Filtering was then performed using a convolution program developed for this

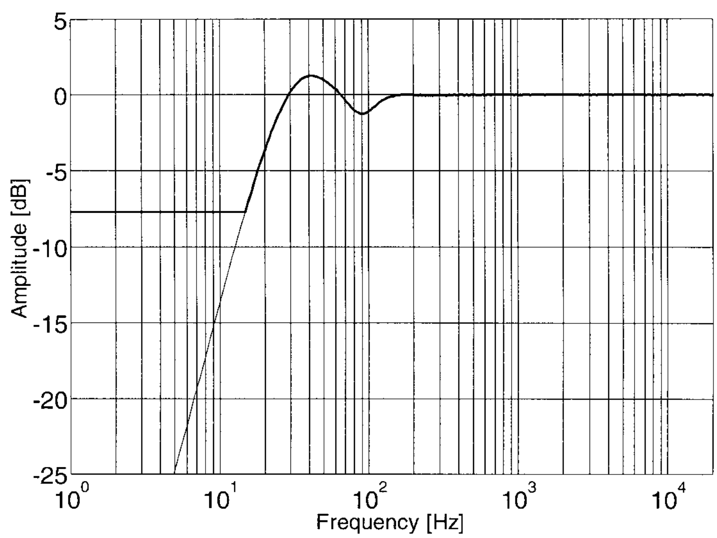

Fig. 12. Measured linear transfer function of nonlinear simulator program. Transfer function was limited below 15 Hz to avoid high gains when implementing inverse filter.

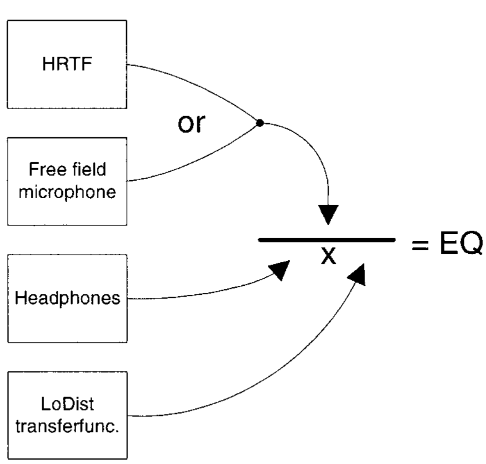

Fig. 13. Calculation of equalization filters EQ for either program material using HRTF measurements or pink noise using free-field microphone measurements.

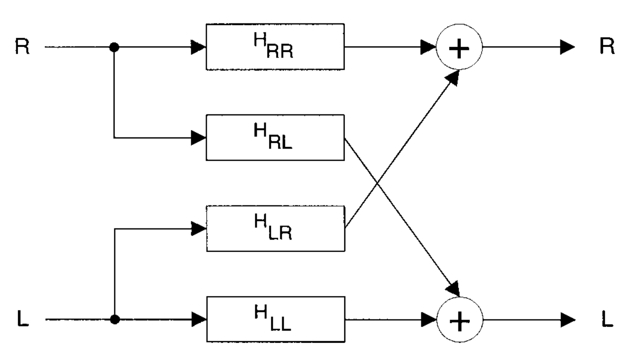

Fig. 14. Four different filters used to generate signals for headphone playback based on two signals for loudspeaker playback.

purpose, based on floating-point arithmetic.

The fixed-point nature of the audio files, that is, 16 bits, was taken into account by down-scaling the equalizer filters appropriately and reintroducing the removed gains in the conversion process to analog signals. The level calibration was performed by adjusting the gain of the digital mixing console, so that the resulting loudness level was 22 sones. This corresponded to a sound pressure level of 69 dB (lin) for one channel playing the processed pink noise.

This calibration and the equalization ensured that the transfer function for the headphone setup was adjusted such that for the reproduction of the same signal, the sound pressure level at the ear drum in each ear of a dummy head was identical to that measured if the dummy head was placed at the listening position in the anechoic chamber.

4 CONCLUSION

Seven program pieces were processed using DSP in order to prepare the program pieces of the listening experiments, described in part I [1]. Pink noise was used to calibrate the loudness level of the two setups—reference loudspeakers in an anechoic chamber and headphones in an audiometric booth. Three different reproduction levels were used, all 10 dB apart. The loudest level was calibrated to 22 sones, which corresponded to 69 dB SPL for one channel playing in the anechoic chamber.

Three different lower cutoff frequencies were implemented—20 Hz, 35 Hz, and 50 Hz—combined with three different orders/slopes—second, fourth, and sixth order. Based on room simulations, four different levels of amplitude ripples and group delay ripples in the passband were implemented. Amplitude ripples varied between 0 and ± 20 dB, and group delay varied between 0 and ± 200 ms. These variations were found by varying the reverberation time T_{60} from anechoic to 0.8 s in room simulations of an IEC 268-13 sized room. When delay ripple was introduced, some unwanted amplitude ripple was also introduced (maximum 1, 5 dB). No unwanted delay ripple was introduced when amplitude ripple was introduced.

In an attempt to have similar conditions in the two

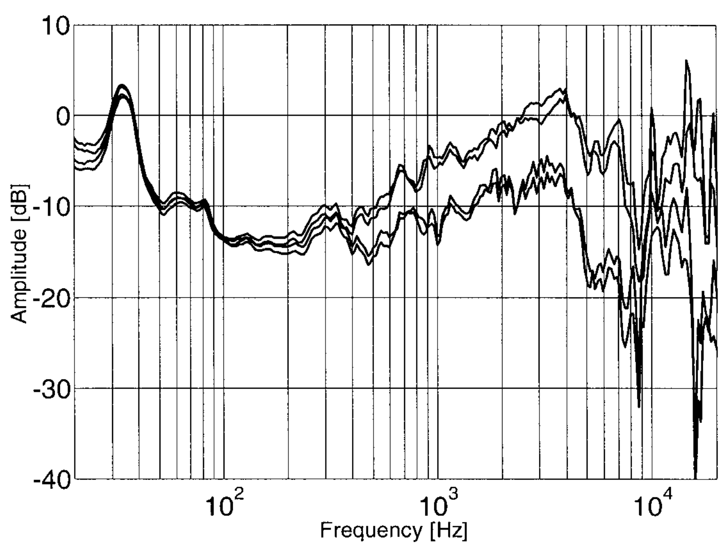

Fig. 15. Amplitude response of four filters used to generate signals for headphone playback based on signals for loudspeaker playback.

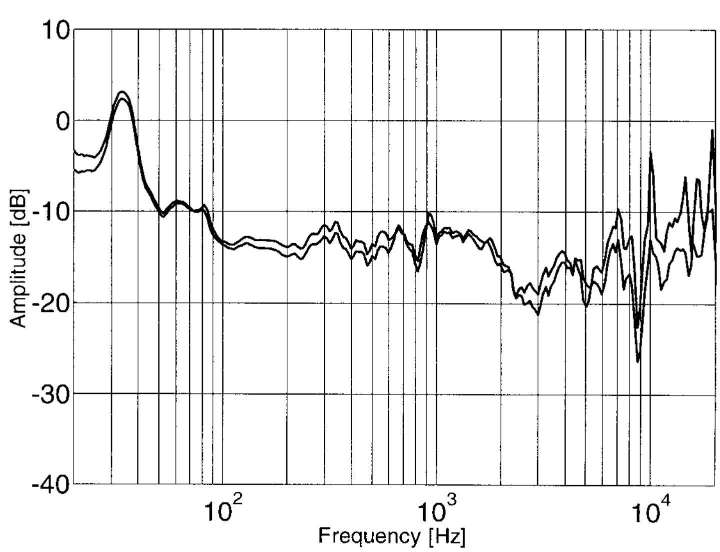

Fig. 16. Amplitude response of two filters used to process pink noise before calibration of headphone setup.

setups, advanced equalization of the signals and simulation of the nonlinear behavior of the woofer drivers were performed before the signals were fed to the headphones. The equalization was based on measurements of HRTFs in the anechoic chamber, the transfer function of the reference loudspeakers using a free-field microphone, the transfer function of headphones mounted on a head and torso simulator, and the linear transfer function of the simulator program for the nonlinear simulation of loudspeaker drivers. All signal processing was performed directly on audio files.

5 ACKNOWLEDGMENT

The authors would like to thank their colleagues in EUREKA Project 1653, Medusa, for countless fruitful discussions and valuable ideas. Great effort was shown, and large tasks were carried out by our colleagues, whom the authors would like to thank: Gert Munch, Peter Chapman, and David Ward at Bang & Olufsen and Ari Varla at Genelec. The authors would especially like to thank Søren Bech from Bang & Olufsen for outstanding enthusiasm and general view of the whole process. In addition the authors would like to thank the Department of Acoustic Technology at the Technical University of Denmark for the use of their large anechoic chamber and for practical support with the measurement setup.

6 REFERENCES

[1] S. Bech, “Requirements for Low-Frequency Sound Reproduction, Part I: The Audibility of Changes in Passband Amplitude Ripple and Lower System Cutoff Frequency and Slope,” *J. Audio Eng. Soc.* vol. 50, pp. 564 – 580 (this issue).

[2] IEC 268-13, “Sound System Equipment,” Part 13: “Listening Tests on Loudspeakers,” International Electrotechnical Commission, Geneva, Switzerland (1985).

[3] J. A. Pedersen, K. Hermansen, and P. Rubak, “The Distribution of the Low Frequency Sound Field and Its Relation to Room Equalization,” presented at the 96th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 42, p. 408 (1994 May), preprint 3852.

[4] F. Jacobsen and J. H. Rindel, “Time Reversed Decay Measurements,” *J. Sound Vibration*, vol. 117, pp. 187 – 190 (1987).

[5] F. Ingerslev, O. J. Pedersen, P. K. Møller, and J. Kristensen, “New Rooms for Acoustic Measurements at the Danish Technical University,” *Acustica*, vol. 19, pp. 185 – 199 (1967/1968).

[6] N. Zacharov, J. Huopaniemi, and M. Hämäläinen, “Round Robin Subjective Evaluation of Virtual Home Theater Sound Systems at the AES 16th International Conference,” in *Spatial Sound Reproduction*, *Proc. AES 16th Int. Conf.* (1999), pp. 544 – 556.

[7] J. Hynninen and N. Zacharov, “GuineaPig — A Generic Subjective Test System for Multichannel Audio,” presented at the 106th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 47, p. 513 (1999 June), preprint 4871.

[8] ISO 532, “Method for Calculating Loudness Level,” International Electrotechnical Commission, Geneva, Switzerland (1975).

[9] E. Sandermann Olsen, “Nonlinear Modeling of Low-Frequency Loudspeakers — A Practical Implementation,” presented at the 102nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 45, p. 414 (1997 May), preprint 4469.

[10] K. Inanaga, Y. Yamada, and H. Koizumi, “Headphone System with Out-of-Head Localization Applying Dynamic HRTF (Head-Related Transfer Function),” presented at the 98th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 43, pp. 401, 402 (1995 May), preprint 4011.

THE AUTHORS

J. A. Pedersen

A. Mäkivirta

Jan Abildgaard Pedersen received a master's degree in electrical engineering from the University of Aalborg in Denmark in 1993, where he specialized in acoustics.

After graduating, Mr. Pedersen began working at Bang & Olufsen as a DSP research engineer in the Electroacoustics Research & Development department, and since 1998 he has been a technology specialist for acoustics and DSP. In this capacity, he is not only responsible for Bang & Olufsen following this technology area, but also contributes to the development of it. He works in many other research areas including low-frequency sound field, room equalization, near-field acoustic measurements, subwoofer requirements, and especially what DSP has to offer in acoustic applications. He is a member of the AES.

●

Aki Mäkivirta was born 1960 in Jyväskylä, Finland. He received Diploma Engineer, Licentiate in Technology, and Doctor of Science in Technology degrees in electrical engineering from Tampere University of Technology, Tampere, Finland, in 1985, 1989 and 1992, respectively. His doctor's thesis described applications of nonlinear signal processing methods for haemodynamic monitoring in medical intensive care.

In 1983, Dr. Mäkivirta joined the Medical Engineering Laboratory of the Research Center of Finland at Tampere, where he worked in various research positions in the field of biomedical signal analysis. In 1990 he joined Nokia Corporation Research Center in Tampere, Finland, where he served as a project manager, and after 1992 was the research manager responsible for building a research group for DSP applications in television audio and high-quality loudspeaker reproduction. In 1995 he joined Genelec Oy, Iisalmi, Finland, where he is a R&D manager responsible for digital system and signal processing applications.

Dr. Mäkivirta is a member of the AES and IEEE. He holds 13 patents, and he is the author of more than 40 journal and conference papers.