Virtualized Listening Tests for Loudspeakers*

TIMO HIEKKANEN,[1] AKI MÄKIVIRTA,[2] *AES Member,* **AND MATTI KARJALAINEN,[1]** *AES Fellow*

[1]*Helsinki University of Technology, Department of Signal Processing and Acoustics, Espoo, Finland*
[2]*Genelec Oy, Iisalmi, Finland*

The precise location of a loudspeaker in a listening room is known to affect loudspeaker preference ratings. When multiple loudspeakers are compared, the evaluation is limited by the poor human auditory memory. To overcome these problems, a method to evaluate and compare loudspeakers using headphones is proposed. The method utilizes personal head-related transfer functions in rendering the sound field recorded in a standard listening room with an artificial head. The equalization of circumaural headphones and the artifical-head responses for individual listeners are investigated. Formal listening tests are conducted to examine differences between the proposed binaural method and real loudspeakers in a standard listening room. Listening tests show that the virtualized loudspeakers can be nearly imperceptible from reality in many but not all cases.

0 INTRODUCTION

Traditionally the subjective evaluation of loudspeakers is done in room acoustics, usually in standardized listening rooms. Listening tests are conducted to assess loudspeaker performance or to establish the preference ranking of several loudspeakers. Initially the performance of the loudspeakers is evaluated by the designer, and the final evaluation is made by the consumer when making a purchase decision. However, there are several aspects that can prevent reliable direct comparisons between loudspeakers.

Human auditory memory is short. We cannot accurately remember complex sound images for longer than a few seconds. Our long-term auditory memory does not yield solid references and our mood of the day can affect the preference ratings severely if we try to compare a current loudspeaker to a loudspeaker that is not presently at hand.

It is well established that the position of a loudspeaker in a room can strongly affect the perceived sound quality. Also, the room itself affects preference ratings even if the loudspeakers to be evaluated are placed in the same position. Bech [1] showed that the listening room will influence the perceived differences between loudspeakers in different positions as well as the perceived differences between loudspeakers in similar positions in a different room. Olive et al. [2] came to similar conclusions. They found that loudspeaker location was the most significant factor in listener preference ratings.

Unfortunately what we see is often what we hear. Visual cues can seriously affect the results of listening tests, and should be prevented by using an acoustically transparent curtain.

To achieve reliable and consistent results that can be compared across tests when evaluating loudspeakers in a listening test, all loudspeakers should be evaluated in the same room placed in exactly the same physical position. The time taken to switch between loudspeakers should be small due to the short time span of the human auditory memory. The listener should remain in exactly the same position all the time. Any visual cue should be eliminated. It is difficult to fulfill all these requirements in real life.

The spatial radiation properties of a loudspeaker are an important part of its fidelity. In anechoic conditions the direct sound radiating from the loudspeaker to the receiving point determines the properties of a loudspeaker. In room conditions sound radiated to directions other than the listening direction can make a significant difference. Depending on the loudspeaker and its position, different room modes are excited and the early reflection pattern received at the listening position changes. To evaluate such spatial properties, loudspeakers must be listened to or measured in room conditions.

Binaural techniques have been used to ease the listening test methods and to ensure that the listening conditions are equal for every test subject [3], [2]. Recently Olive et al. [4] showed that similar loudspeaker preference ratings are achieved with a binaural room scanning method and real loudspeakers. Gilkey and Anderson [5] point out the benefits of binaural technology in the measurement and evaluation of audio signals.

*Presented at the 124th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 2008 May 17–20; revised 2009 January 5.

The performance of binaural recordings and binaural synthesis have been evaluated in numerous studies [6]–[10] (for more, see [11]). Research has mainly focused on the localization performance of measurements and recordings done at the entrance to a closed ear canal, ignoring issues related to sound coloration.

In the present paper measurements are taken at the entrance to an open ear canal. A method for binaural measurement and synthesis using head-related loudspeaker–room responses is proposed, and its use in loudspeaker evaluation is discussed. Spatial and spectral attributes of the method are compared to real loudspeakers in formal listening tests.

0.1 Binaural Recording, Synthesis, and Reproduction

According to Møller [12] the motivation for binaural techniques is that the input to our hearing system consists of only two signals—the sound pressures at the eardrums. If these signals are recreated precisely, including dynamic changes with head and body movements, all auditory aspects of an auditory event are repeated perfectly. Headphones are the most practical reproduction devices for binaural recordings, in spite of their many shortcomings, since they offer almost complete channel separation.

Binaural signals can be recorded either with a head and torso simulator or with a true head using miniature microphones. Different types of head and torso simulators have been built, starting from spheres with two microphones to full-scale replicas of an average human upper body. Møller et al. have shown that in terms of localization, the best results are always achieved with individual recordings [6]. An artificial head is only an approximation and cannot provide good localization and timbre for everyone, and the individual variations in the quality of reproduction are large.

Without compromising the reproduction of spatial information, individualized recordings can be made at any point between the ear drum and a few millimeters outside the ear canal entrance. However, three recording positions are of special interest—at the ear drum, at the entrance to an open ear canal, and at the entrance to a closed ear canal [12]. The position at the entrance to an open ear canal is chosen here for the following reasons.

- Only the measured headphone response needs to be compensated if the microphones used to measure the binaural responses are small enough not to disturb the sound field at the entrance to an open ear canal significantly.
- The measurement of binaural responses as well as headphone responses is straightforward at the entrance, although it is known to be critical to the precise position.
- Measurements at the entrance to an open ear canal give maximum comfort to the test subjects.

The auditory event produced by the loudspeakers can be simulated with headphones if the transfer functions from each loudspeaker to each ear and from each headphone terminal to each ear are known. In a stereophonic listening setup as in Fig. 1(a), the responses Y_l and Y_r of loudspeaker inputs X_l and X_r at the ear canal entrances are

$$Y_l = X_l H_{ll} + X_r H_{rl} \tag{1}$$

$$Y_r = X_l H_{lr} + X_r H_{rr} \tag{2}$$

Here H_{ij}, where i and j are varied for left (l) and right (r), represent the transfer functions from loudspeaker inputs to ears. Fig. 1(b) shows the signals for headphone equalization. If the responses from headphone inputs to ear canal entrances are P_l and P_r, the inverses of them cascaded with loudspeaker–room response transfer functions and headphones in an ideal case duplicate the natural listening condition. Therefore the headphones need to be fed by the signals

$$\bar{Y}_l = Y_l/P_l \quad \text{and} \quad \bar{Y}_r = Y_r/P_r. \tag{3}$$

Notice that the point-to-point transfer functions H_{ij} include all the information needed to be known (measured) for a particular positioning setup. These correspond to long reverberant impulse responses that are not minimum phase, and this feature must be retained, while short

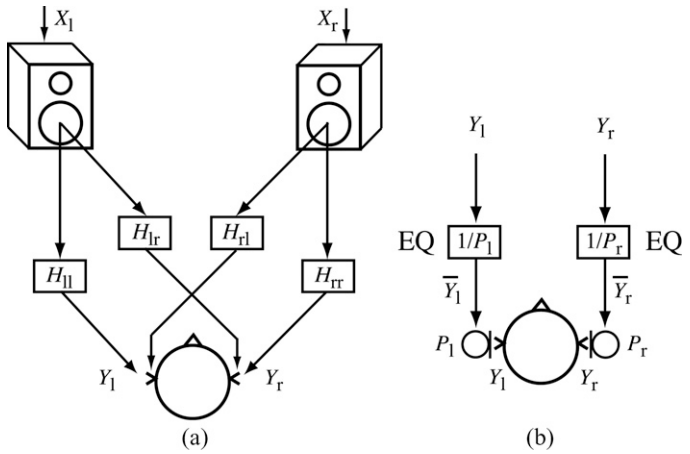


Fig. 1. (a) Transmission paths in stereophonic listening setup. (b) Headphone equalization setup.

equalization filters can be made minimum phase as usual in binaural auralization.

Directional hearing of humans is based on interaural level differences (ILD), interaural time differences (ITD), and spectral cues [13]. If an artificial head provides roughly correct cues, a question arises if the localization properties of artificial-head responses for an individual listener could be improved by an equalizer, that is, a mapping from measured artificial-head responses to the subject's responses. Because the couplings of a subject's ears to the room acoustics are more or less different from those of the artificial head, it is not possible to achieve a perfect mapping even in theory. However, the hypothesis in our study was that this mapping can be made accurate enough in practice.

1 MEASUREMENTS

Transfer functions from loudspeakers to ears are needed for binaural synthesis. A series of measurements were conducted in a standardized listening room to understand how repeatable binaural measurements are in room conditions, and to compare true-head measurements with artificial-head measurements.

All measurements and processing are performed at a 44.1-kHz sampling rate, or if this is not possible, responses are resampled to 44.1 kHz before processing. The head-related spherical coordinate system is used, where φ denotes azimuthal and δ elevation angle. Also a stereophonic listening setup is used if not mentioned otherwise.

1.1 Equipment

Binaural true-head measurements can be made by attaching small microphones to a test subject's ears. Alternatively a head and torso simulator representing an average human upper body can be used. The artificial head has properties that make it superior to true-head measurements. It can be placed accurately and repeatably. Due to the sensitivity of the room responses to placement differences, the exact placement is essential for comparable results. The microphones of the artificial head are mounted permanently, which removes the variance caused by microphone locations.

The artificial head (Manikin MK1 by 01dB-Metravib) used in the measurements is made of polyurethane with Nextel coating. The ear shape complies with the IEC 959 and DIN V 45608 standards. Microphones are 1/2-inch condenser microphones positioned at the end of the 20-mm-long ear canal. The microphone signal is transferred through an AES/EBU connection.

Small electret microphone capsules (Sennheiser KE 4-211-2) were used in the true-head measurements. The diameter of the capsules is 4.75 mm and the height 4.2 mm, the manufacturer promises a flat frequency response from 40 Hz to 20 kHz. The capsules were soldered to cables, and a thin and solid wire was wrapped around the cable to give support and shape.

A two-channel preamplifier (Unides Design UD-MPA 10e) provided polarization voltage for the microphones. The microphones were attached to the test subject's head, as shown in Fig. 2. The wire was twisted to fit behind the ear, and tape was applied to relief strain and to keep the microphones in place.

Circumaural dynamic headphones (Sennheiser HD590) were used in the measurements as well as in the reproduction of binaural synthesis. Measurements were all done in an ITU-RBS.1116–compliant listening room [14] using the logarithmic sine sweep technique [15].

1.2 Artificial-Head Measurements

To test the repeatability of artificial-head measurements and to find the positioning accuracy needed, the following measurements were taken.

The manikin was placed on a chair, its head raised to the level of a true-head listener. Precise and repeatable positioning was confirmed with a plumb line hanging from the ceiling and markings on the floor. The distances from each loudspeaker to the plumb line were measured to be 240 cm.

First the artificial head was moved toward the line between the loudspeakers, and binaural responses were measured for every 2 cm of movement for each loudspeaker. Beyond 10 cm one measurement was made at the 15-cm displacement. Second the artificial head was moved to the left parallel to the line between the loudspeakers, 1 cm at a time. Third the artificial head was rotated horizontally 2.5° at a time from 0° to 10° and measurements were taken as earlier.

The measured impluse responses were convolved with stereophonic commercial rock music (Porcupine Tree: 'Trains' from the record In Absentia) and monophonic pink noise, and processed as in Eqs. (1) and (2). The results were listened to with the headphones. Fast and seamless switching between different convolutions was enabled using the Pure Data programing environment [16].

As expected, moving the artificial head forward was found to cause less perceivable differences than moving

Fig. 2. Microphone attached to subject's head.

it sideways. With music a 15-cm movement in the forward direction provides a difference that is just noticeable. With pink noise a 10-cm movement is noticeable. Displacement to the side direction causes perceivable differences much faster. A 1-cm sideways displacement is noticeable when listening to pink noise, while a displacement of 3 to 4 cm is perceivable with music.

Sensitivity to rotation depends highly on the audio material. With pink noise a rotation of 2.5° made an audible difference, which was expected since earlier studies have shown that human localization blur in the horizontal plane can be less than 2.5° [13]. However, even a change in direction of 10° was found difficult to notice with certain music signals.

To explore the overall repeatability of measurements, the following was done. First the artificial head was placed in the room as described earlier, and the first measurement was made. Then loudspeakers with stands were removed from the room and carried back in and positioned as they had been. After measurements the artificial head was removed and put back, and the final measurements were made.

Similar informal listening as earlier was performed, and it was confirmed that equipment can be located accurately enough to achieve repeatable results. No difference was heard with music or pink noise.

It must be stressed that although these results are based on informal listening by the author, they give an idea of how accurate the placement of the the artificial head must be in order to avoid audible errors due to placement differences. The lateral accuracy should be ±1 cm at least, and the forward direction should be well specified. Inaccuracy of placement in the frontal direction is not as critical as rotation and lateral displacement, but it should not be overlooked. It seems possible to repeat artificial-head measurements without perceivable differences between the measurements. Fig. 3 shows a typical magnitude response of an artificial-head measurement and illustrates the difference between two measurements.

The repeatability of headphone responses of the artificial head was also investigated. Fig. 4 demonstrates the repeatability using circumaural dynamic headphones. Responses were measured five times consecutively. Headphones were taken off and put back on between the measurements. Albeit an effort made to place the headphones in the same way, differences greater than 10 dB can be seen at frequencies above 7 kHz.

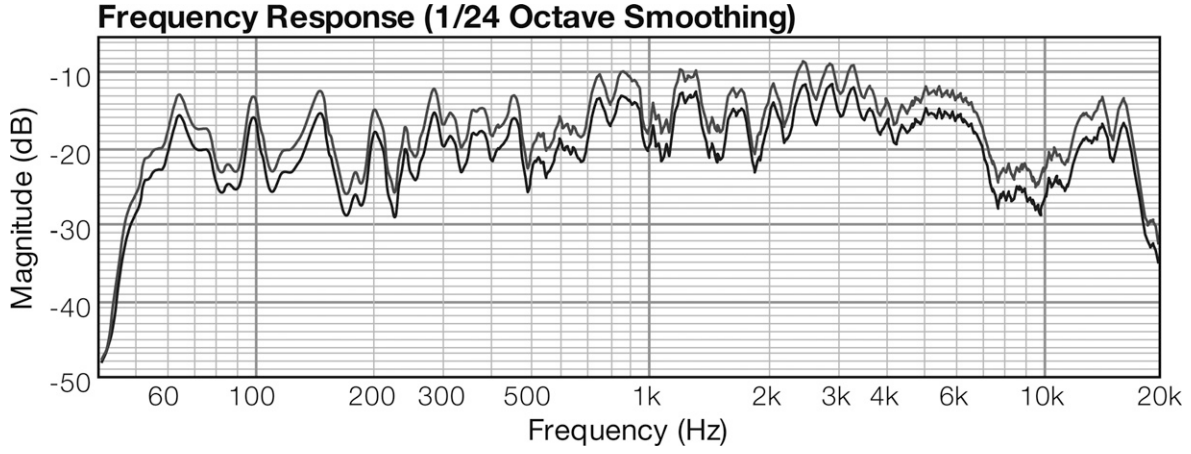


Fig. 3. Responses in listening room to left ear of an artificial head from a loudspeaker at $\varphi = -30°$. Artificial head and loudspeaker were repositioned between measurements. Curves are separated by 3 dB on purpose.

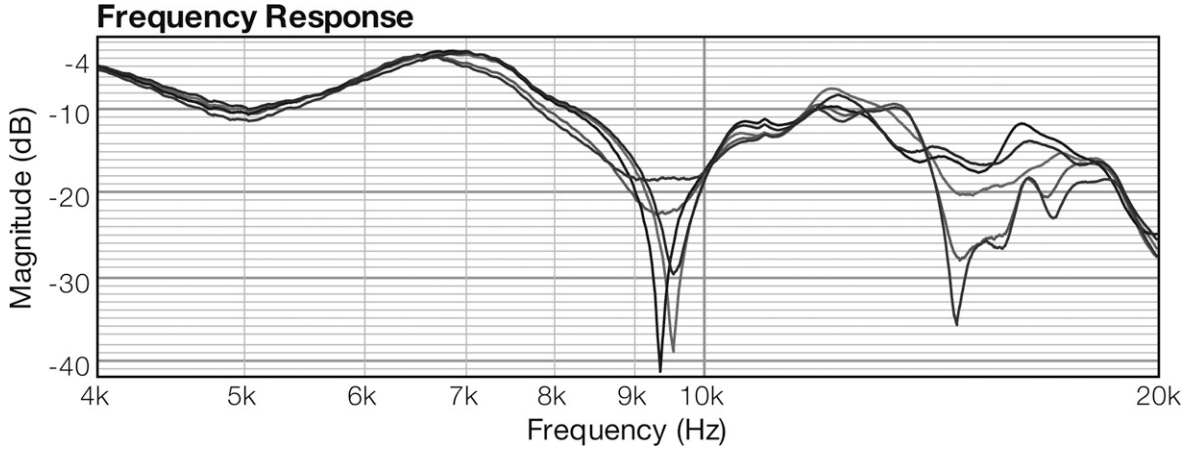


Fig. 4. Five consecutive measurements of headphone transfer functions with artificial head. Zoomed to frequency range 4–20 kHz.

Møller et al. studied headphone responses with human subjects and came to the conclusion that the responses are reliable only up to 7 kHz [17]. Riederer investigated the repeatability of dummy-head responses and noted that below 7 kHz responses agree very well [18]. He achieved ±3 dB repeatability up to 13 kHz with circumaural headphones (Sennheiser HD580).

1.3 True-Head Measurements

To test the repeatability of true-head measurements, three consecutive measurements were made. The microphones were taken off the subject and the subject was allowed to walk for a while between measurements. Photographs were taken of the microphone attachments and special care was taken to place the microphones every time as similarly as possible.

The location and orientation of the subject's head were controlled with a plumb line hanging above the head. The subject was asked to look at a black dot in the front wall and to keep his/her head still.

Measurements agree very well up to 1 kHz, but above that the curves differ. Fig. 5 illustrates the differences above 500 Hz. As could be expected, these differences in measured responses are audible if the responses are compared to each other with headphones in a similar way as for the earlier artificial-head measurements.

Reasons for the differences are not known, but a few guesses can be made. The microphone locations are probably not exact enough, causing variance in the measurements. In another study we found that a displacement of 1 mm at the ear canal entrance of a dummy head can cause several decibels of variation at high frequencies. The test subject's head cannot be located as accurately as the artificial head, and it may move during a measurement. Finally the human body itself is a noise source. For example, blood circulation, breathing, and swallowing cause interferences.

To explore the repeatability of true-head headphone responses, five consecutive measurements were made. Microphones were attached to the subject's head by an experimenter. The headphones were placed by the test subject, since Møller et al. noted that this produces good repeatability [19]. Fig. 6 shows that repeatability appears to be better than with the artificial head. The headphones were taken off and put on a table between measurements. According to Fig. 6, the frequency responses are within 3 dB up to the frequency of 13 kHz. Variation is almost constant with respect to frequency in contrast to the

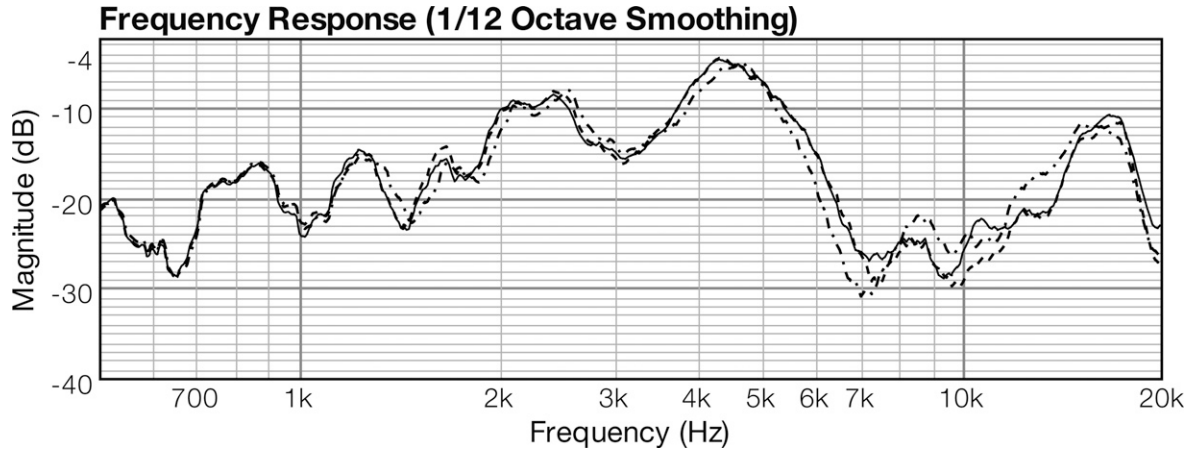


Fig. 5. Comparison of three true-head measurements. Microphones were removed between measurements and test subject was allowed to move.

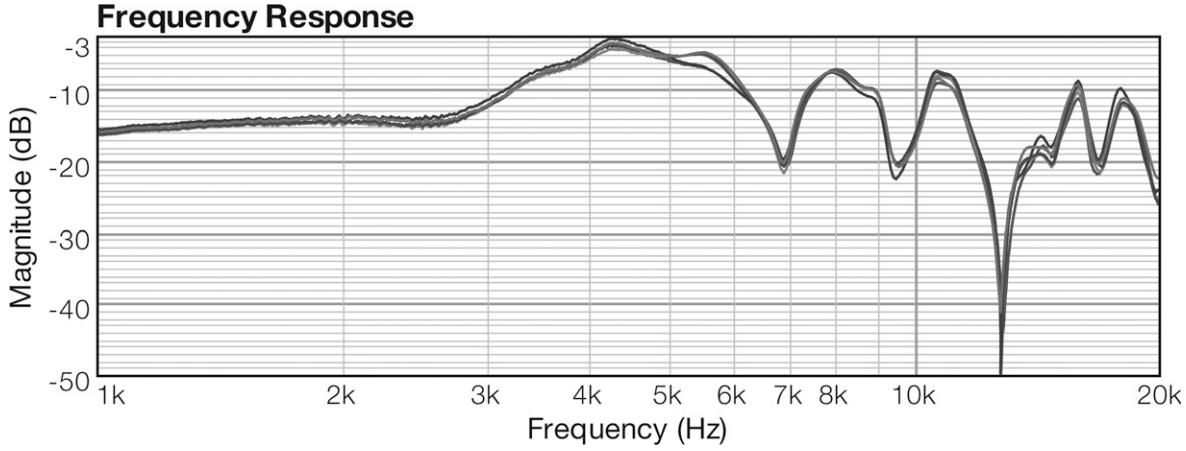


Fig. 6. Repeated true-head headphone response measurements.

artificial-head measurements, where much less variation was present at low frequencies.

The true-head headphone measurements seem to be rather well repeatable. However, the measurements were all made in one session, and an effort was made to place the headphones in a similar manner. Much greater variations are seen if longer pauses are taken, microphones are replaced, or the headphones are put on carelessly.

Measurements of several test subjects show that binaural loudspeaker–room responses as well as headphone responses are highly individual. Also the responses are asymmetrical, meaning that the left- and right-ear responses differ.

1.4 Conclusion from Measurements

The measurements indicate that true-head measurements alone cannot be used to compare loudspeakers in a stereophonic listening setup. Differences between measurements can be greater than differences between loudspeaker responses.

Loudspeaker–room measurements using an artificial head are repeatable but cannot be used since the artificial head cannot offer correct localization and timbre for everyone due to the averaged nature of its responses.

To be able to evaluate and compare loudspeakers in the stereophonic listening setup through headphones, both artificial- and true-head measurements were used. The artificial head gives good measurement accuracy and repeatability, and binaural synthesis using true-head responses gives good timbre and correct localization.

2 METHOD

To use responses from an artificial head instead of individual true-head responses, the artificial-head responses must be equalized to match the individual true-head responses. Fig. 7 shows a true-head response and an artificial-head response from a loudspeaker to the right ear under anechoic conditions. As can be seen, responses agree only below 1 kHz. This could be expected since the artificial head has microphones at the ear drum position.

In theory artificial-head responses can be used together with individual true-head responses as in Eqs. (4) and (5),

$$Y_\mathrm{l} = \left(X_\mathrm{l} \frac{H_\mathrm{ll}^\mathrm{ref} G_\mathrm{ll}}{G_\mathrm{ll}^\mathrm{ref}} + X_\mathrm{r} \frac{H_\mathrm{rl}^\mathrm{ref} G_\mathrm{rl}}{G_\mathrm{rl}^\mathrm{ref}} \right) \frac{1}{P_\mathrm{l}} \tag{4}$$

$$Y_\mathrm{r} = \left(X_\mathrm{l} \frac{H_\mathrm{lr}^\mathrm{ref} G_\mathrm{lr}}{G_\mathrm{lr}^\mathrm{ref}} + X_\mathrm{r} \frac{H_\mathrm{rr}^\mathrm{ref} G_\mathrm{rr}}{G_\mathrm{rr}^\mathrm{ref}} \right) \frac{1}{P_\mathrm{r}} \tag{5}$$

where H^ref and G^ref refer to true-head and artificial-head measurements of a reference loudspeaker, G refers to an artificial-head measurement of a loudspeaker to be evaluated, P refers to headphone responses, and Y, X, and indices are as in Fig. 1.

The problem is to design filters $H^\mathrm{ref}/G^\mathrm{ref}$, which equalize the artificial-head responses to match with individual responses, and individual headphone equalizers $1/P$.

2.1 Equalization of Artificial-Head Responses

There is always some noise in the measured impulse responses. The noise is not part of the loudspeaker–room head transfer function and makes it difficult to determine where the level of impulse response becomes insignificantly low from the auralization point of view. Because of this, binaural responses should be truncated to use them in binaural synthesis. Here all responses are truncated prior to other processing. Fig. 8 demonstrates the signal-to-noise ratio (SNR) achieved with true-head measurements. The starting point of a response was decided based on a fixed threshold. The responses were faded linearly to zero at the point where the signal fell below the estimated noise floor. Truncation of the responses was not considered to be critical since the SNR was good (around 60 dB).

To design the correction filters $H^\mathrm{ref}/G^\mathrm{ref}$ only the magnitude information is used, and minimum-phase impulse responses are created. This enables smoothing of the responses in the frequency domain. It is advantageous since the target was to equalize the general shape of artificial-head responses but not the individual room resonances. The use of Kautz filters [20],[21] was investigated briefly, but the minimum-phase design method was selected for its flexibility and ease. A complex smoothing technique [22] could give one starting point, but it was not investigated here.

In the minimum-phase method 32768-point magnitude responses of the true-head and the artificial-head responses are smoothed in the frequency domain using a

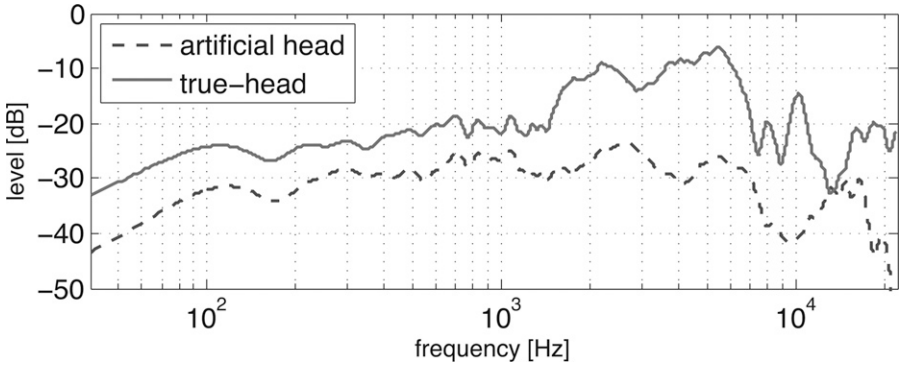


Fig. 7. Magnitude responses of artificial head and true head in anechoic conditions. Measurements are from right ear, loudspeaker being at φ = +30° angle. Different resonances can be seen at high frequencies.

moving Hanning window. The smoothed true-head magnitude response is divided by the smoothed magnitude response of the artificial head. Different smoothing window lengths from 1/24 octave to 1/2 octave were tested and in preliminary listening, 1/4 octave smoothing was found to perform well. A minimum-phase time-domain response is created and the resulting impulse response is truncated after a decay of 60 dB. Fig. 9 shows the magnitude response of a typical correction filter achieved by the minimum-phase method.

The minimum-phase method does not result in imperceptible differences between equalized artificial-head responses and true-head responses, but on the other hand there is no risk of annoying resonances since the magnitude response of the minimum-phase filter is smooth, as shown in Fig. 9.

2.2 Headphone Equalization

The equalization of headphone transfer functions is critical in relation to colorations in binaural reproduction. Headphones are equalized to produce a flat frequency response in the sense of Fig. 1(b) at the physical location of the binaural measurement, in our case at the entrance to an open ear canal.

In theory it is sufficient to design an inverse filter $1/P$. However, direct inversion of the magnitude response does not provide an optimal solution because of the variance in frequency response produced by headphone placement inaccuracy. At high frequencies, magnitudes and frequencies of the resonances vary from one measurement to another depending on the position of the headphones.

According to Bücklein, peaks in the magnitude response should be avoided. A peak is more audible in the reproduction than a corresponding dip [23]. In addition, Toole and Olive state in [24] that wide resonances are detected more readily than narrow peaks. Two guidelines can now be formulated for headphone equalization.

1) Avoid high peaks, especially the wide ones.
2) Do not widen the existing peaks and dips if possible.

The first requirement implies that peaks in the inverted magnitude response of the headphone transfer function should be compressed to ensure that there are no peaks above the average level in the equalized response. The second requirement implies that the inverted magnitude response should not be smoothed excessively since the smoothing widens resonances, and on the other hand it flattens notches, which are needed to compensate for the

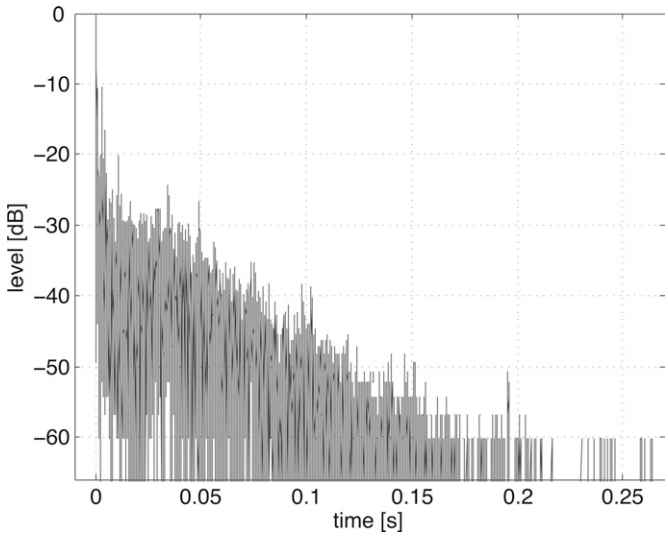


Fig. 8. 12 000 first samples of ipsilateral true-head response squared, plotted on logarithmic scale.

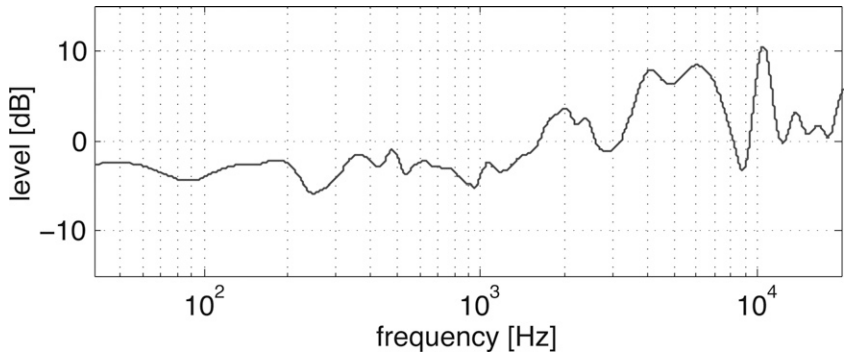


Fig. 9. Magnitude response of a filter achieved by minimum-phase method.

peaks of the headphone transfer function in the reproduction phase.

The proposed method for headphone equalization is as follows. The measured headphone response is truncated to 512 samples. The magnitude of a one-sided, 4096-point spectrum of the headphone response is smoothed to remove small variations caused by noise. The smoothing is done by averaging the magnitude response with a moving Hann window. The width of the window is 1/48 octave.

The smoothed magnitude response is inverted. Fig. 10 shows a typical result of the inversion of the magnitude response of the headphones. The dashed curve represents the smoothed and inverted response. After smoothing, a reference level for peak reduction is decided based on the average level of the inverted response from 40 Hz to the frequency of the first minimum magnitude value below 4000 Hz.

Magnitude values exceeding the peak reduction level are compressed. Compression is only active above the frequency of the minimum-magnitude value below 4000 Hz. Amplitude values exceeding the peak reduction level are multiplied by a compression ratio. For instance, if the peak reduction level is l, the amplitude value of a specific frequency is a, and the compression ratio is r, then the result for a specific frequency would be $(a - l)\, r + l$. In informal listening 1/4 compression ratio was found to give good results. The effect of the peak reduction is shown in Fig. 10. Peaks at high frequencies are taken down from 5 to 10 dB. A slight frequency rolloff starting from 4000 Hz was designed to the inverse filter. The rolloff was used to compensate the sharpness caused by unsuccessfully equalized resonances.

Finally a minimum-phase impulse response is created and truncated after a decay of 60 dB. An individual filter is designed for each ear. It is strongly recommended that the headphone response be measured in the same session where the binaural loudspeaker–room responses are measured. Remounting the measurement microphones can shift the resonance frequencies significantly, resulting in improper headphone equalization for a specific set of binaural loudspeaker–room responses.

2.3 Equalization Method

In this section a new method for subjective loudspeaker evaluation is proposed. The method trades problems of loudspeaker placement and loudspeaker swapping for problems related to measurement and equalization accuracy. To sum up, the method consists of the following steps:

1) Using a reference loudspeaker pair, true-head loudspeaker–room responses are measured at the entrance to an open ear canal. The reference loudspeaker here means the one that is used to obtain the artificial-to true-head mapping, which is then used for any other loudspeaker.
2) Headphone responses are measured with the same microphone placements. The individual inverse headphone filters are calculated.
3) Using the reference loudspeaker pair the artificial head is measured in the same position where the true-head measurements were made.
4) All loudspeakers intended for listening tests are measured in a similar manner using the artificial head. More loudspeakers can be measured later, if the measurement position as well as loudspeaker positions are well documented.
5) Four individual filters are calculated to equalize the artificial-head responses to the true-head responses.
6) Preliminary loudness normalization of the equalized responses is done.
7) Convolutions between the equalized artificial-head responses and test signals are calculated and headphone equalization is done using the precalculated filters.

To allow changing the responses and to improve flexibility, the impulse responses of the equalized artificial-head responses and the impulse responses of the headphone equalization filters are stored separately. A graphical programming environment [16] was used to create a program that performs the real-time convolutions needed for the binaural reproduction of any audio material. Parallel convolutions enable seamless switching

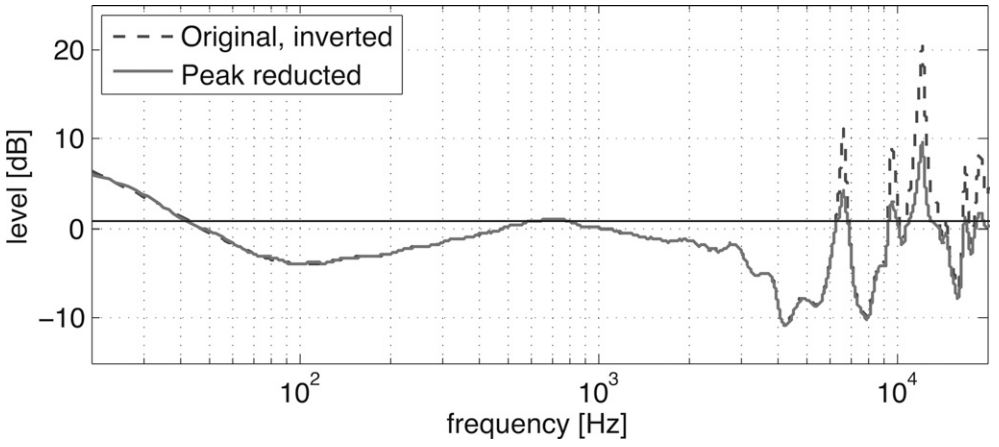


Fig. 10. Typical inverted headphone response (dashed line). Horizontal line indicates peak reduction level. Solid line is inverted headphone response after peak reduction.

between different responses. Four pairs of virtual loudspeakers can be compared without a delay of switching loudspeakers or any physical arrangements. In the demonstration system the program was run on a laptop computer (MacBook) with a 2-GHz dual-core processor (Intel Core 2 Duo) and firewire audio interface (MOTU Traveler). The four parallel binaural convolutions (16 channels of convolution altogether) for 32768-sample impulse responses took about 65% of the processor time.

3 LISTENING TEST: COMPARISON WITH REALITY

To explore the differences between virtual and real loudspeakers, formal listening tests were organized. The aim was to understand how binaural reproduction differs from real loudspeakers in a real room and to evaluate differences between auralization using true-head responses (true-head method) and auralization using individually equalized artificial-head responses (artificial-head method).

The task given to the test subjects was to evaluate differences in reproduction in terms of five attributes: apparent source width, direction of events, distance to events, spaciousness, and tone color. The first three attributes are directly related to localization performance. Apparent source width describes how the width of a sound source or sound sources is perceived. Is the source well defined or is it blurred somehow? Direction of events refers to the direction where the auditory event appears to originate, and distance to events is the distance from the listening position to the point where the auditory event appears to happen. Spaciousness describes the amount of space present in the listening. Tone color describes the spectral content of the sample.

The test subjects were asked to rate the difference between real and virtual loudspeakers using the verbally anchored ITU small-impairment scale from 1 to 5 [14]. The anchor points and the verbal descriptions are shown in Table 1. The test subjects were able to set the difference ratings in increments of 0.1 point.

The listening tests were conducted in an ITU-R BS. 1116–compliant listening room [14]. Two pairs of studio monitoring loudspeakers (Genelec 1030A and Genelec 8030A) were used in the test. The loudspeaker placement was controlled with plumb lines hanging from the ceiling.

3.1 Test Subjects

Seven males and one female subjects participated in the test. Seven of the test subjects reported no hearing damages: one subject reported continuous tinnitus. Although all test subjects cannot be considered experts in loudspeaker evaluation, all had at least some experience in participating in listening tests.

Table 1. ITU small-impairment scale.

Grade	Impairment
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

3.2 Samples and Processing

Three different audio test signals were used in the test. Anechoic male speech, moving slowly from left to right and back to left, gave an easy way to evaluate the directions and discolorations since the human hearing is specialized to analyze speech signals. A 40-s excerpt of a jazz song (Screen play on record Landmark by Mika Pohjola) was used since it has a wide spectrum and simultaneous sound sources located in different positions. Pink noise, meaning wide-band noise, which has equal energy in each octave, was used as the most critical test signal for evaluating sound discolorations.

The test signals were auralized using the truncated true-head responses and the individually equalized artificial-head responses. The responses of loudspeaker A were used to design the equalizers for loudspeaker B and vice versa, resulting in six different test cases for one loudspeaker pair. Table 2 shows the different cases. Each case was repeated once, and all attributes were rated.

3.3 Test Procedure

The test was divided into four sections. First the experimenter attached the microphones on the test subject's head and measured the binaural true-head impulse responses in the stereophonic listening setup for each loudspeaker pair. Headphone responses were measured directly after the loudspeaker measurements. As the validity of the responses was ensured, the microphones were removed. The measurement phase took about 35 min, including microphone positioning and chaning of the loudspeakers.

While the audio files for the listening test were rendered, the test procedure was explained to the test subject. Written descriptions of scale and attributes were given. The test subject was advised not to pay attention to possible loudness differences or background noise, to keep his/her head still, and to look forward when listening to the virtual loudspeakers through the headphones. The listening order of headphones first and real loudspeakers subsequently was recommended but not enforced. The test subject was allowed to familiarize him/herself with the material and to experiment switching between virtual and real loudspeakers. Processing of the measured responses and familiarization took approximately 25–30 min.

Table 2. Different test samples.

	Loudspeaker A Method		Loudspeaker B Method	
Speech	True	Artificial	True	Artificial
Music	True	Artificial	True	Artificial
Noise	True	Artificial	True	Artificial

The evaluation phase was divided into two parts, one for each loudspeaker pair. A short break was taken between the two parts and the loudspeakers were switched.

In the evaluation phase the test subject rated the difference of one virtual loudspeaker pair and one real loudspeaker pair using a computer mouse and the user interface shown in Fig. 11. The test subject could switch between headphones and loudspeakers at any time. Pressing the play button started the audio clip from the beginning, but switching between virtual and real loudspeakers was instantaneous. The test subjects were able to adjust the volumes of the virtual and real loudspeakers to equalize the loudnesses and were advised to do so if perceived loudnesses were not the same. There was no time limit. When one case was rated, the test subject could move on by pressing the next button. The order of the samples was randomized for each test subject. The first case was an additional case for test subject training only, and it was excluded from the analysis.

The average duration of the evaluation phase was 1 hour, including a pause between the two parts. After the second part short verbal comments were obtained.

3.4 RESULTS

The received data were analyzed using the multiway analysis of variances (ANOVA) and multiple-comparision tests. The data were fitted to a normal distribution. The homogeneity of variances was tested using Levene's test, and deviations from normal distribution were visually inspected. Although it was found that the data do not exactly fulfill the assumptions of ANOVA, ANOVA is known to be robust for small violations of the assumptions [25].

Fig. 12 shows the means and the 95% confidence intervals for each attribute, test signal, and processing method. True-head responses and equalized artificial-head responses worked well for the speech signal. Apparent source width, direction, and distance of events are all rated above 4.5, which corresponds to imperceptible on the ITU small-impairment scale. Spaciousness and coloration lie between 4 and 4.5 (perceptible but not annoying). Although the means of the artificial-head method are worse, the differences are small (< 0.1), and the confidence intervals of the true-head and artificial-head results overlap.

All attributes get lower ratings with music and noise signals. With the music signal the difference from reality is rated as perceptible but not annoying. The difference between the auralization methods is greatest in terms of distance to events but the confidence intervals overlap. With the noise signal all attributes except tone color are above 3.5, corresponding to perceptible but not annoying. In terms of coloration, the difference from reality was rated as slightly annoying.

The six most prominent effects in the ANOVA analysis were audio material used (sample), processing method of the binaural measurements (method), attributes used (attrib), repetitions of the ratings (repet), loudspeaker type (speaker), and test subject (subj). All other main effects, except repetitions and loudspeaker type, were found to be significant ($p < 0.01$). There were also a few significant second- and third-order interactions. A full ANOVA table is presented in the Appendix.

The most significant effect was the audio sample. In further investigations with a multiple-comparison test (Tukey's post-hoc test) it was found that the means of all three samples were significantly different. The effect of the test subjects appeared to be significant, which implies

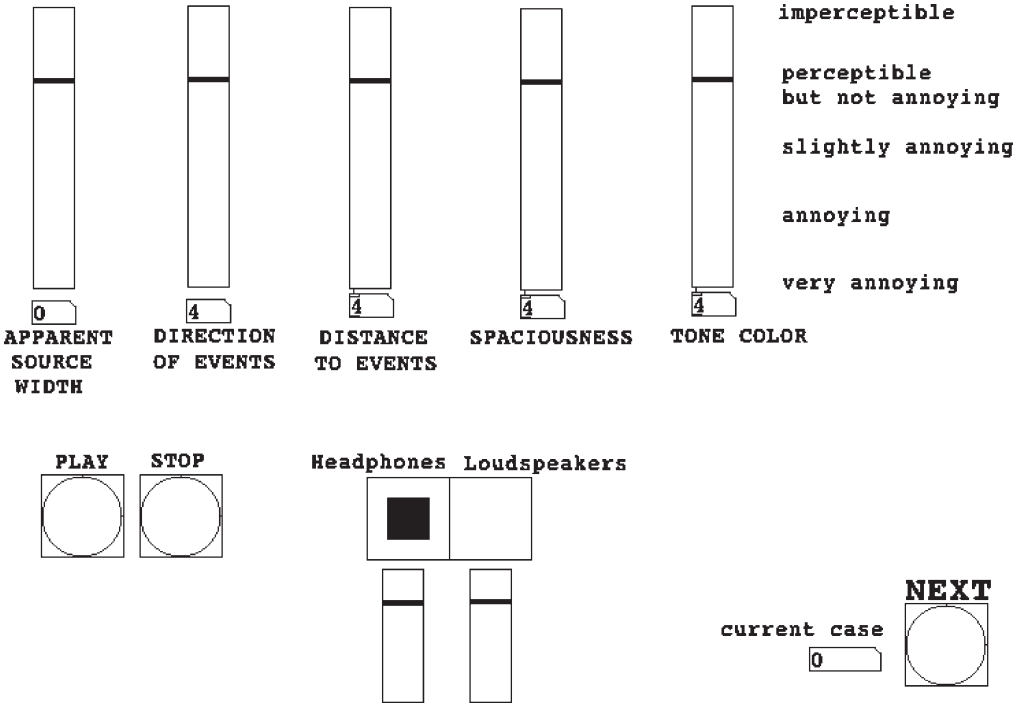


Fig. 11. User interface of listening tests.

that the performance of the binaural method depends on the test subject. The multiple-comparison test showed that one test subject gave significantly lower ratings while one of the eight subjects gave significantly higher ratings than the others.

The effect of the attributes is not interesting per se since it only implies that the attributes were graded differently, which was expected. Also the insignificance of the repetitions and loudspeaker-type effects was expected. The test subjects were experienced, and the method should work similarly regardless of the loudspeakers.

Although the effect of the method was found significant, the F value was low compared to the F values of the significant main effects. By visual inspection of Fig. 12 it was concluded that there is no perceptual difference between the true-head method and the artificial-head method, or the difference is highly insignificant compared to other factors such as intersubject variations. Of course the conclusion is valid only for indirect comparisons such as the test described here.

The significant ($p < 0.01$) second-order interactions in the ANOVA table were sample∗attrib, sample∗subj, attrib∗subj, and repet∗subj. Fig. 12 confirms the sample∗attrib interaction. The three other interactions are related to the test subjects; which confirms that either the performance of the binaural method depends on the test subject or the subjects were not a very homogeneous group. Most of the significant third-order interactions are also related to the test subjects. The sample∗method∗speaker interaction suggests that the loudspeakers might have some effect on the ratings. The conclusion is supported by the low p value of the main effect (0.08). In general the F values of the interactions are low, indicating that the interactions are not as significant as the main effects.

4 SUMMARY, DISCUSSION, AND CONCLUSIONS

In this paper, the repeatability of true-head and artificial-head measurements was investigated under room conditions. The repeatability of headphone transfer functions for an artificial head and a true head was studied. All true-head measurements were taken at the entrance to an open ear canal. Methods for the equalization of

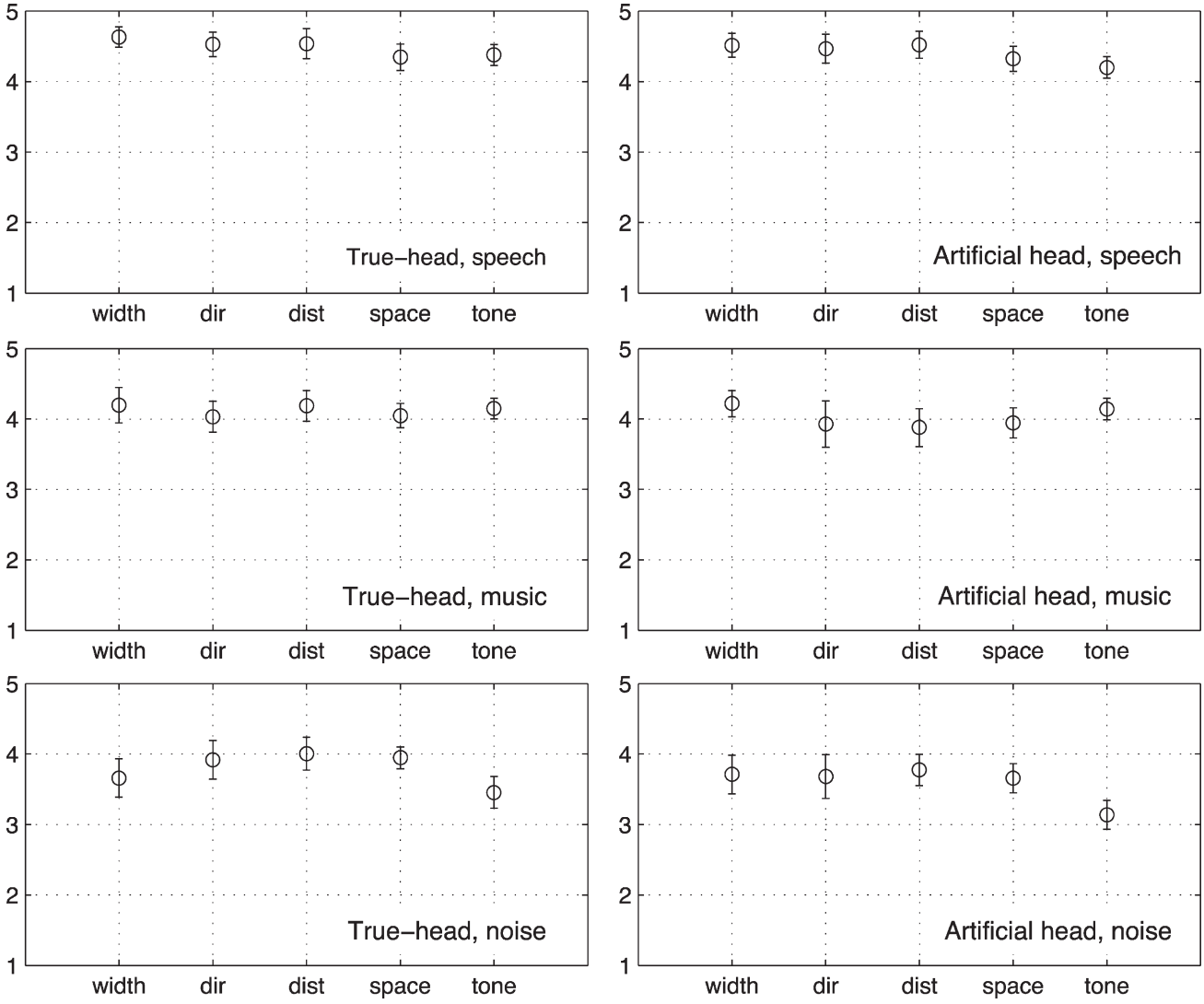


Fig. 12. Means and 95% confidence intervals, combined from both loudspeaker pairs. Y axis—ITU small-impairment scale.

artificial-head responses and headphone equalization of binaural reproduction individually for each subject were proposed. The performance of binaural synthesis using the proposed methods was compared to loudspeakers in a real room in formal listening tests. It was found that the performance depends highly on the test signal used.

The strong dependence of the ratings on the test signal is probably connected with the measurement and equalization inaccuracies at high frequencies. Most of the speech signal energy is below 4 kHz. Above this frequency the headphone equalization is not exact. There is more energy at high frequencies in the music and noise signals, which may lead to the audible differences observed in direct comparison with real loudspeakers. One explanation for the differences can be found from the well-known problems of binaural techniques [3]. The speech signal was moving and the movement started from the direction of a real sound source. The movement gave the feeling of the presence of dynamic localization cues helping the externalization remarkably. The noise signal was stationary and located in front of the listener where the performance of binaural techniques is the worst.

Many of the test subjects were astonished by the quality of externalization of the male speech. All critical comments from the test subjects were related to differences in high frequencies. Either the sound color was too bright or sharp or the high frequencies were not located correctly. This is in agreement with the authors' subjetive findings that the limiting factor in the comparison with reality is coloration. Unnaturalness of the sound color decreases the usability of the binaural method since it draws attention away from other attributes of sound.

Sound coloration could possibly be reduced if it were possible to place the headphones exactly similarly every time. With circumaural headphones it is difficult, but with intra-aural headphones it might be possible. The use of intra-aural headphones would make the measurement procedure more complicated since the measurement microphones should be inserted inside the ear canal. On the other hand, special headphones with integrated in-ear microphones could make careful recalibration of the headphones easy for each listening session.

The listening test indicates that it is possible to use equalized artificial-head responses instead of true-head responses in binaural synthesis. Individual equalization of the artificial-head responses improves spatial properties and reduces sound coloration of the binaural synthesis close to those of binaural synthesis using true-head responses. To assess the audible differences between true-head responses and individually corrected artificial-head responses in more detail, the responses should be compared with each other in a formal experiment, without comparison to reality.

The proposed listening test method trades the problems of traditional loudspeaker listening tests for the problems related to the measurement accuracy and equalization of the virtualized listening setup. The nonlinear properties of loudspeakers cannot be represented by the virtualized listening test method. There are no dynamic localization cues in the method used, and the binaural responses represent only a single position in the listening room. Dynamically varying equalization based on head tracking was excluded in this study because of its high complexity in measurements, equalizer design, and real-time implementation. We noticed that quite small head rotation destroyed the illusion of a realistic sound scene, but keeping the head carefully fixed worked surprisingly well.

Furthermore, virtual loudspeakers should not be compared to real ones in loudspeaker testing. Instead, all loudspeakers to be compared should be virtualized for the comparison task. Virtual loudspeakers are comparable in the sense that the same processing is done to all virtual loudspeaker pairs, and even small differences can be noticed in a way similar to comparing real loudspeakers. We believe that the method is most useful in analytical comparisons of different loudspeakers, such as a tool for a loudspeaker designer. Preference rating should be done with more caution, keeping in mind that rating happens in a specific room environment, in a specific loudspeaker positioning setup, and without dynamic cues related to head movement.

A clear advantage of the method is that loudspeaker comparison by headphones can be done anytime and anywhere, even for loudspeakers that are not available to anyone, as long as their responses are retained in a database. On the other hand, due to the imperfections, the present virtualized method is not recommended as the sole method to evaluate loudspeakers, especially in the most critical testing. The present study should be seen as a basic functionality check of the proposed method, and much further testing and development are needed to more deeply understand its benefits and shortcomings. The results obtained are, however, promising and show directions for future improvements. These include improving especially the headphone equalization for timbre at high frequencies, which was the most problematic feature. Different headphone types should be studied, and the signal processing methods in general could be made more systematic. The effects of different listening rooms, positioning of the sources and receivers, loudspeaker types, head tracking, and so on should be investigated. Due to the high complexity of the problem domain, we had to exclude them to keep the scope and extent of the present study manageable.

5 REFERENCES

[1] S. Bech, "The Influence of the Room and of the Loudspeaker Position on the Timbre of Reproduced Sound in Domestic Rooms," presented at the AES 12th International Conference, Copenhagem, Denmark (1993 June 28–30).

[2] S. E. Olive. P. L. Schuck, S. L. Sally, and M. E. Bonneville, "The Effects of Loudspeaker Placement on Listener Preference Ratings," *J. Audio Eng. Soc.,* vol. 42, pp. 651–669 (1994 Sept.)

[3] F. E. Toole, "Binaural Record/Reproduction Systems and Their Use in Psychoacoustic Investigations," presented at the 91st Convention of the Audio

Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 39, p. 1005 (1991 Dec.), preprint 3179.

[4] S. E. Olive, T. Welti, and W. L. Martens, "Listener Loudspeaker Preference Ratings Obtained in situ Match Those Obtained via a Binaural Room Scanning Measurement and Playback System," presented at the 122nd Convention of the Audio Engineering Society, (Abstracts) www.aes.org/events/122/122ndWrapUp.pdf, (2007 Oct.), convention paper 7034.

[5] R. H. Gilkey and T. R. Anderson, Eds., *Binaural and Spatial Hearing in Real and Virtual Environments* (Lawrence Erlbaum, Mahwah, NJ, 1997), chap. 28.

[6] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings", *J. Audio Eng. Soc.*, vol. 44, pp. 451–469 (1996 June).

[7] P. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "Localization with Binaural Recordings from Artificial and Human Heads," *J. Audio Eng. Soc.*, vol. 49, pp. 323–336 (2001 May).

[8] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Evaluation of Artificial Heads in Listening Tests," *J. Audio Eng. Soc.*, vol. 47, pp. 83–100 (1999 Mar.).

[9] J. Kawaura, Y. Suzuki, F. Asano, and T. Sone, "Sound Localization in Headphone Reproduction by Simulating Transfer Functions from the Sound Source to the External Ear," *J. Acous. Soc. Jpn.*, vol. 12, pp. 203–216 (1991).

[10] E. M. Wenzel. M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 94, pp. 111–123 (1993).

[11] J. Blauert, Ed., *Communication Acoustics* (Springer, New York, 2005).

[12] H. Møller, "Fundamentals of Binaural Technology," *Appl. Acoust.*, vol. 36, pp. 171–218 (1992).

[13] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, rev. ed. (MIT Press, Cambridge, MA, 1997).

[14] ITU-R BS. 1116-1, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," International Telecommunications Union, Geneva, Switzerland (1997).

[15] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept Sine Technique," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 48, p. 350 (2000 Apr.), preprint 5093.

[16] URL http://puredata.info/ (2007).

[17] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc*, vol. 43, pp. 203–217 (1995 Apr.).

[18] K. A. J. Riederer, "HRTF Analysis: Objective and Subjective Evaluation of Measured Head-Related Transfer Functions," Ph.D. thesis, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland (2005).

[19] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Design Criteria for Headphones," *J. Audio Eng. Soc.*, vol. 43, pp. 218–232 (1995 Apr.).

[20] T. Paatero and M. Karjalainen, "Kautz Filters and Generalized Frequency Resolution: Theory and Audio Applications," *J. Audio Eng. Soc.*, vol. 51, pp. 27–44 (2003 Jan./Feb.).

[21] M. Karjalainen and T. Paatero, "Equalization of Loudspeaker and Room Responses Using Kautz Filters: Direct Least Squares Design," *EURASIP J. Adv. Signal Process.* (2007).

[22] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses," *J. Audio Eng. Soc.*, vol. 48, pp. 259–280 (2000 Apr.).

[23] R. Bücklein, "The Audibility of Frequency Response Irregularities," *J. Audio Eng. Soc.*, vol. 29, pp. 126–131 (1981 Mar.).

[24] F. E. Toole and S. E. Olive, "The Modification of Timbre by Resonances: Perception and Measurement," *J. Audio Eng. Soc.*, vol. 36, pp. 122–142 (1988 Mar.).

[25] A. Rutherford, *Introducing Anova and Ancova: A GLM Approach* (Sage Publ., London, UK, 2000), chap. 7.

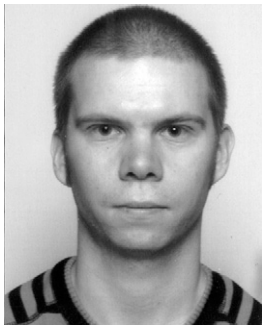
APPENDIX
ANOVA TABLE

Source	Sum sq.	d.f.	Mean Sq.	*F*	Prob > *F*
sample	90.2274	2	45.1137	266.6122	0.0
method	3.9466	1	3.9466	23.3236	1.7366e-06
attrib	7.7543	4	1.9386	11.4566	5.7029e-09
repet	0.081018	1	0.081018	0.4788	0.48923
speaker	0.52384	1	0.53384	3.1549	0.076204
subj	58.9291	7	8.4184	49.7511	0.0
sample∗method	0.71889	2	0.35944	2.1242	0.12041
sample∗attrib	12.7395	8	1.5924	9.4109	2.6985e-12
sample∗repet	0.10874	2	0.05437	0.32131	0.72532
sample∗speaker	0.72308	2	0.36154	2.1366	0.11894

ANOVA TABLE (*continued*)

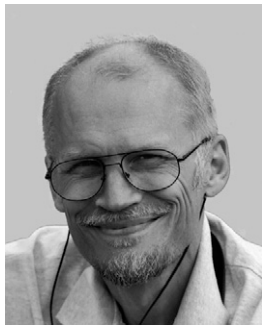
Source	Sum sq.	d.f.	Mean Sq.	F	Prob $> F$
sample∗subj	43.1955	14	3.0854	18.234	0.0
method∗attrib	0.84039	4	0.2101	1.2416	0.29201
method∗repet	0.082316	1	0.082316	0.48647	0.48578
method∗speaker	0.73415	1	0.73415	4.3387	0.037676
method∗subj	2.9358	7	0.4194	2.4786	0.016294
attrib∗repet	0.2901	4	0.072524	0.4286	0.78803
attrib∗speaker	0.45237	4	0.11309	0.66836	0.61413
attrib∗subj	41.7445	28	1.4909	8.8107	0.0
repet∗speaker	0.20431	1	0.20431	1.2074	0.27228
repet∗subj	3.9146	7	0.55922	3.3049	0.0018496
speaker∗subj	2.9021	7	0.41459	2.4501	0.017514
sample∗method∗attrib	1.8972	8	0.23716	1.4015	0.19255
sample∗method∗repet	0.25038	2	0.12519	0.73984	0.47762
sample∗method∗speaker	2.5479	2	1.2739	7.5288	0.00058929
sample∗method∗subj	3.8825	14	0.27732	1.6389	0.064591
sample∗attrib∗repet	1.0023	8	0.12529	0.74041	0.65578
sample∗attrib∗speaker	0.72041	8	0.090052	0.53219	0.83259
sample∗attrib∗subj	33.1641	56	0.59222	3.4999	1.7097e-14
sample∗repet∗speaker	0.16377	2	0.081887	0.48393	0.61659
sample∗repet∗subj	3.0638	14	0.21884	1.2933	0.20603
sample∗speaker∗subj	6.3387	14	0.45276	2.6757	0.00081775
method∗attrib∗repet	0.29973	4	0.074932	0.44283	0.77766
method∗attrib∗speaker	0.46801	4	0.117	0.69146	0.59804
method∗attrib∗subj	3.442	28	0.12293	0.72647	0.84814
method∗repet∗speaker	0.0014278	1	0.0014278	0.0084378	0.92684
method∗repet∗subj	0.85831	7	0.12262	0.72463	0.65116
method∗speaker∗subj	3.6014	7	0.51448	3.0405	0.0037646
attrib∗repet∗speaker	0.25481	4	0.063704	0.37648	0.82549
attrib∗repet∗subj	6.2541	28	0.22336	1.32	0.12703
attrib∗speaker∗subj	6.7536	28	0.2412	1.4254	0.073599
repet∗speaker∗subj	0.77578	7	0.11083	0.65495	0.71031

THE AUTHORS



T. Hiekkanen

A. Mäkivirta



M. Karjalainen

Timo Hiekkanen was born in Finland in 1983. He received an M.Sc. degree in technology from the Helsinki University of Technology, Espoo, Finland, in 2008.

He has worked as a research assistant in the Laboratory of Acoustics and Audio Signal Processing at the Helsinki University of Technology and is presently working in the field of AV programming and usability.

●

Aki Mäkivirta was born in Jyväskylä, Finland, in 1960. He received Diploma Engineer, Licentiate of Science 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.

In 1983 he joined the Medical Engineering Laboratory of the Research Centre of Finland at Tampere and was involved with applications of signal processing in biomedical signals. From 1990 to 1995 he was responsible for digital signal processing applications in television and high-quality audio signal processing at Nokia Corporation Research Center, Tampere. Since 1995 he has been working as an R&D manager at Genelec Oy, Iisalmi, Finland.

Dr. Mäkivirta is a member of the Audio Engineering Society.

Matti Karjalainen was born in Hankasalmi, Finland, in 1946. He received M.Sc. and Dr.Sc. (Tech.) degrees in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 1970 and 1978, respectively.

Since 1980 he has been a professor in acoustics and audio signal processing at the Helsinki University of Technology, Espoo.

Dr. Karjalainen's main interest is in audio signal processing, such as DSP for sound reproduction and auralization, music DSP, and sound synthesis, as well as perceptually based signal processing. In addition his research activities cover speech processing, perceptual auditory modeling, spatial hearing, DSP hardware, software, and programming environments, as well as various branches of acoustics, including musical acoustics and modeling of musical instruments. He has written 370 scientific and engineering papers and contributed to organizing several conferences and workshops, such as serving as the papers chairman of the AES 16th Conference and as technical chairman of the AES 22nd Conference. He is a fellow and silver medalist of the Audio Engineering Society, and a member of the Institute of Electrical and Electronics Engineers, the Acoustical Society of America, the European Acoustics Association, the International Speech Communication Association, and several Finnish scientific and engineering societies.