Audio Engineering Society
Convention Paper
Presented at the 113th Convention
2002 October 5–8 Los Angeles, CA, USA

Three-dimensional headphone sound reproduction based on active noise cancellation

Daniël Schobben[1], and Ronald Aarts[1]

[1] *Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands*

Correspondence should be addressed to Daniel Schobben (Daniel.Schobben@Philips.com)

ABSTRACT

Headphone signal processing systems that are commercially available today are not optimized for the individual listener. This results in large localization errors for most listeners. In the present work, a system is introduced that requires a one time calibration procedure, which can be carried out conveniently by the listener. This system consists of conventional headphones into which small microphones have been mounted. An active noise cancellation method is used to achieve a sound reproduction via headphones, which is as close as possible to a reference loudspeaker setup. The active noise cancellation system is based on adaptive filters that are implemented in the frequency domain.

INTRODUCTION

Headphone virtualizers are systems that aim at giving the user the illusion that the sound is coming from loudspeakers rather then from the headphones themselves [1, 2, 3, 4]. Systems that are commercially available today are not optimized for the individual listener. This results in large localization errors for most listeners [5]. The system at hand is personalized in that it requires a calibration procedure, which can be carried out conveniently by the listener. This system consists of conventional headphones into which miniature microphones have been mounted. The sound reproduction using headphones gives the same listening experience to the user as the reference (multichannel) loudspeaker system. This is achieved by taking all contributions into account: the room impulse responses, the loudspeaker characteristics, the headphone characteristics and the properties of the listener's head and torso. Besides the usual computational requirement for a headphone virtualizer, this system needs in addition two low-cost microphones and two analog to digital converters to convert the microphone signals.

Technology background

The way in which sound propagates from the loudspeaker towards the ear-drums of the listener depends on the loudspeaker, the room and the physical properties of the listener (e.g. the shape of the head, ears, torso). If loudspeaker reproduction is emulated using headphones, these sound charateristics have to be taken into account and compensation for the sound reproduction characteristics of the headphones is required.

The physical properties of the head and outer ears of the listener modify the sound as it travels from the source to the ear-drums. The transfer functions describing this sound propagation from multiple sound sources to both ears are known as Head-Related Transfer Functions (HRTFs). Multichannel audio can be filtered with the HRTFs of the listener and the inverses of the headphone to ear transfer functions prior to headphone sound reproduction. In this way the multichannel loudspeaker system can be emulated very accurately [6]. Note that only one loudspeaker driver is required at each side of the head in order to make multichannel virtual sound as a person has only one ear per side. Sounds add in a linear way in the air so that headphone signals of the virtual left loudspeaker can be added to those of the virtual right loudspeaker to obtain virtual stereo for example.

When audio is filtered with HRTFs that are measured from another person, there are large errors in the vertical and front/back localization [7]. Therefore the sound reproduction system should be personalized.

Configuration

Fig. 1 shows a 5 channel loudspeaker setup. A person who is wearing headphones resides inside the sweet spot, i.e. the region in which the sound reproduction of the loudspeaker setup is optimal. The headphones are

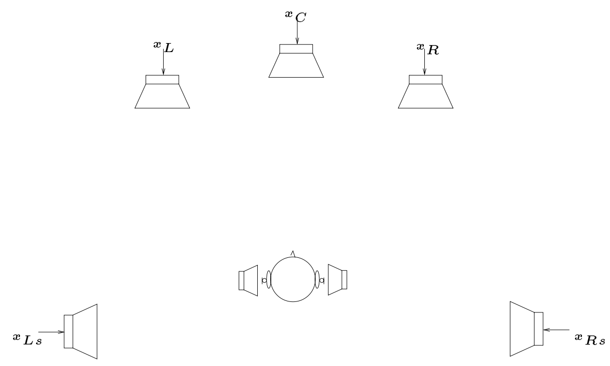


Fig. 1: A 5 channel sound reproduction setup with headphones with integrated microphones.

equipped with integrated microphones [8] and are connected to a digital signal processing unit (DSP). During the calibration, the DSP is connected to the multichannel loudspeaker setup. A noise signal is played through each of the loudspeakers consecutively and is picked up by the microphones. The DSP then computes how the sounds should be processed prior to headphone reproduction, such that exactly the same sound is generated at the position of the microphones, which are very close to the ears. The algorithm that is used is described in the next section. When the calibration is completed, the listener can manually choose between loudspeaker or headphone sound reproduction, showing the capabilities of the system. Note that the calibration can be carried out using only one loudspeaker. The subject needs to change his/her orientation after each measurement such that this loudspeaker corresponds to the left front, right front, left rear, right rear and center loudspeaker position.

ACTIVE NOISE CANCELLATION

The algorithm that is used during the calibration is essentially an active noise cancellation algorithm. An introduction to active noise cancellation is given next, a noise cancellation primer can be found in [9]. Its application to headphone listening will be explained below.

In sound reproduction systems, sound signals can be filtered prior to reproduction by loudspeakers to ideally obtain perfect sound reproduction at a finite number of positions in space. The filters can be found by first placing microphones at the relevant positions and using the difference between the ideal sound and the reproduced sounds at these positions as errors signals for an adaptive algorithm. In classic adaptive-filter theory these error signals are obtained by comparing the desired signals with the adaptive filter outputs. In active noise cancellation the error signals are obtained by comparing the desired signals with adaptive filter outputs that are filtered by acoustic transfer functions. The classical adaptive filter is depicted in Fig. 2 (*top*) where it is used to equalize the acoustic transfer function from the loudspeaker to the microphone $H(z)$. Here, the update uses the input signal of the adaptive filter $W(z)$ and the difference between the reference signal $d[n]$ which resembles $x[n]$ and the adaptive-filter output. The reference signal $d[n]$ can be a delayed version of $x[n]$ for example, so that $W(z)$ can converge to a stable solution with $W(z)H(z)$ equal to this delay.

Instead of equalizing a signal $x[n]$ that is filtered by an acoustic transfer function $H(z)$, sound reproduction systems need to filter this signal prior to playback as depicted in Fig. 2 (*bottom*). Both systems are equivalent if the adaptive filter $W(z)$ is constant. In practical applications it suffices to demand that the adaptive filter is slowly varying. The latter system is termed the filtered-

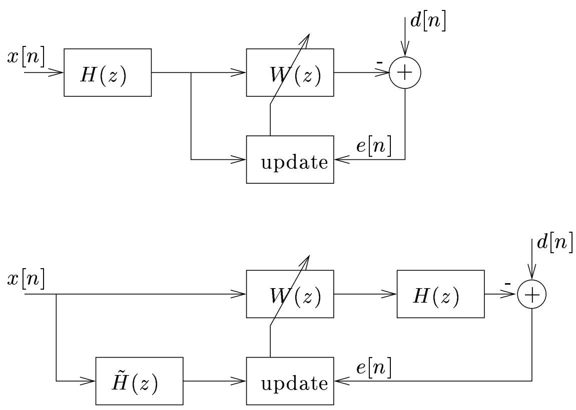


Fig. 2: Conventional adaptive filter (*top*) and filtered-x equivalence (*bottom*).

x algorithm in [10] which indicates that a filtered version of the signal $x[n]$ is used in the update. This filter $\tilde{H}(z)$, which corresponds to the acoustic transfer function $H(z)$, is not exactly known in practical situations. An estimate of it can be used however, and the filtered-x algorithm is known to be robust to estimation errors herein. In [11] it was found that supperior perfomance is obtained by replacing $\tilde{H}(z)$ by an all-pass filter with the same phase. Application areas include noise cancellation in ducts and phantom sound source generation [9].

Phantom sound source generation using noise cancellation

In Fig. 3 the filtered-x algorithm is applied in an HRTF measurement system. The system is shown for only one side of the headphones. The processing for the other side is identical and works independently. A noise signal is fed to the loudspeaker. This noise signal is also filtered before it is fed to the headphones. The filtering is done in such a way that the microphone signal is minimized. In this way the adaptive filter $h_{L,L}$ will become approximately equal to the transfer function of the loudspeaker to the microphone $g_{L,L}$ times the inverse of the transfer function of the headphone to the microphone gx_L

$$h_{L,L} = -g_{L,L}(gx_L)^{-1}. \tag{1}$$

A noise suppression of about 20 dB can be achieved at the position of the microphone. This means that the sound of the headphones is almost identical to that of the loudspeaker. When the loudspeaker is switched off the listener will have the illusion to listen to the loudspeaker[1].

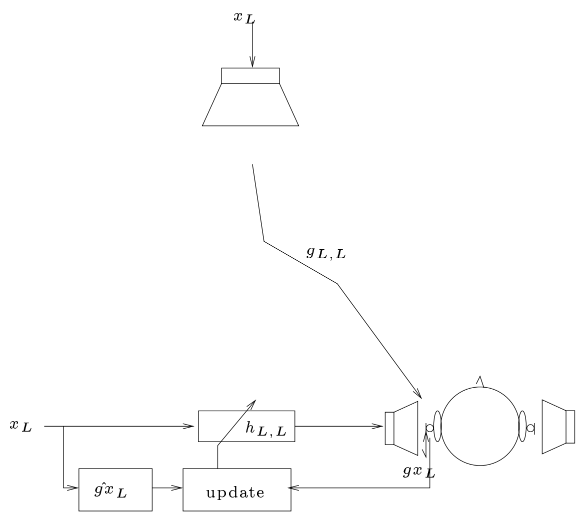


Fig. 3: HRTF measurement using noise cancellation.

INFLUENCE OF THE HEADPHONES

Many prototypes of headphones with integrated microphones have been realized based on both in-ear head-

[1] Note that theoretically the headphones signals should be reversed in sign first. However, perceptually this has only a small impact.

phones and full-size headphones. Extensive tests have shown that both the in-ear headphone prototypes as well as the full-size headphone prototypes are suitable to create a convincing 3-D sound image. The tests also show that the positioning of the in-ear headphones was critical; once a calibration was done removing and remounting the headphones was sufficient to destroy the 3-D sound image. The full-size headphones are much more robust to repositioning.

The headphones with integrated microphones that is examined further in this paper is based on the Philips SBCHP890. The headphone prototype is shown in Fig. 4. The microphones are electrets which are 6 mm in diameter. The microphones are glued onto the grill of the headphones. A tube is mounted onto them so that the microphones can measure more closely to the ears. Some foam is entered into the tubes' ends in order to damp tube resonances[2]. The headphone prototype is shown in Fig. 4.

Fig. 4: Headphones with integrated microphones. The silicon tube runs from the microphone to the ear-canal entrance.

A drawback of most ordinairy full-size headphones is that they obstruct the sound path from the loudspeaker to the ears. There are two main reasons for this. One is that the sound reflects from the headphone surface which can be localized by the listener as a secondary sound source. This could be remedied by using a different (smaller) headphone design and by changing the acoustic absorption characteristics of its surface. An absorbing coating could have a positive effect for the high frequencies. The second reason is that the headphones are not acoustically transparent. The foam in the cushions will absorb sounds in a frequency dependent manner. In the headphone prototype an of-the-shelf SBCHP890 is used and the ear-cushions are replaced with specially prepared cushions that have foam and cloth which are acoustically as transparent as possible.

It is not possible to fully reflect the influence of wearing headphones on the perceived sound localization of sound around a listener in a measured curve. Psycho-acoustic experiments would be required to assess this accurately. It is possible however to show the sound coloration that is introduced by the headphones which is likely to be the most significant effect. Fig. 5 shows the spectrum of the sound at the position of the 'ear-drums'[3] of an artificial head[4] without headphones, with the original HP890 and with the HP890 with modified cushions. The first plot of Fig. 5 shows that the original HP890 attenuates the sound by roughly 20 dB above 1 kHz. The second plot shows that the headphone prototype attenuates sounds above 3 kHz with no more than 10 dB. The third plot shows the reproduceability by remeasuring the headphone curve after reinstalling the headphones. Once the calibration is done the tubes are no longer needed. They may be removed by the user. It can be seen from the fourth plot that this has no effect on the sound coloration; the curves with and without tubes are as closely related to each other as the curves in the reproduceability experiment of the third plot. This means that although they are required for the measurement, the tubes do not change the way the sound propagates towards the ear-drums and hence they may be removed after the calibration phase.

Similar measurements have been done for two in-ear headphones, neckband headphones and full-size headphones which have been specifically designed to be acoustically transparent. The measurements are shown in Fig. 7. These headphones are shown in Fig. 6. It can be concluded that in-ear headphones change the sounds much less. The HP890 headphone prototype is quite good however, even when compared to the very open Sony full-size headphones.

[2]The filtered-x algorithm equalizes the transfer function from the headphones to the microphones including all tube resonances. Strong tube resonances require a large dynamic range in the system however. Also, if the headphones are repositioned the filtered-x algorithm can become instable more easily for strong tube resonances.

[3]The 'ear-drums' are the measurement microphones inside the dummy head.

[4]The B&K 4128 dummy head was used

Fig. 6: Headphone prototypes based on Philips in-ear headphones (*top-left*), Sony full-size headphones (*top-right*), Sony in-ear headphones (*lower-left*), close-up of Sony in-ear headphones (*lower-right*).

INFLUENCE OF THE POSITIONING OF THE ERROR MICROPHONE

The influence of the positioning of the error microphone is assessed next. The sound propagation in the ear canal is assumed to be one dimensional[5]. White noise is used during the calibration to identify the acoustic transfer functions. Noise cancellation is achieved at the position of the microphone. The noise cancellation is quite accurate in the close proximity of the microphone (roughly within one tenth of a wavelength). Further away from the microphone, the anti-noise can even add constructively to the noise. It is assumed that unreliable results are obtained when the distance between the microphone and the entrance of the ear canal are in the order of the wavelength or more.

[5]If the wavelength is in the order of the diameter of the ear canal or less, the sound propagation in the ear canal is no longer one dimensional. This will happen at approximately 17 kHz so that half a wavelength is 1 cm which is roughly equal to the diameter of the ear canal.

Zone-of-quiet

The active noise control configuration that is used here is quite a special one. The noise source (i.e. the loudspeaker) is relatively far away from the listener (e.g. 2 m). The secondary sound source (i.e. the headphone) that should cancel the noise is relatively near by the ear canal entrance (e.g. 3 cm) and the error microphone that is used is somewhere in between the ear canal entrance and the secondary sound source. The zone-of-quiet is defined as the area around the error microphone where the noise suppression is still quite accurate. It is interesting to see what the size and shape of the zone-of-quiet is for this particular configuration.

Fig. 8 on page 10 shows a computer simulation of noise cancellation for this configuration. The ear canal entrance is at the origin. The noise source is 2 meters away from the origin. When the headphone is silent, the noise level is about 35 dB at the position of the origin. The distance of the headphone to the ear canal is fixed for a given pair of headphones (27 mm in this case). It is assumed for simplicity in this simulation that the sources

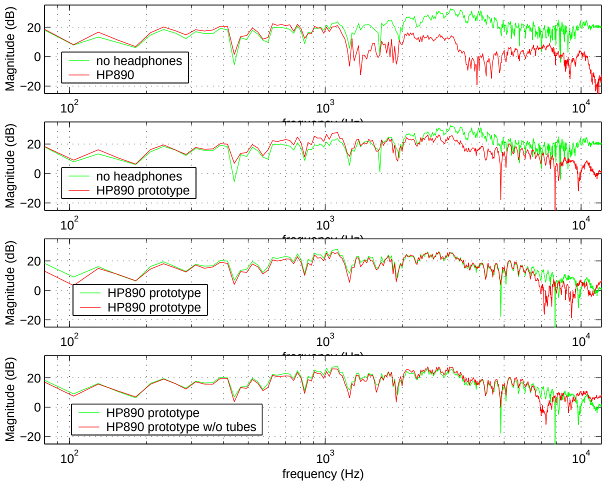


Fig. 5: The four graphs show how the sound is affected when: Top graph: an original SBCHP890 is put on (1), the HP890 headphone prototype is put on (2), the HP890 headphone prototype is put on two times (3), Bottom graph: the HP890 headphone prototype is put on with the tubes installed or removed (4).

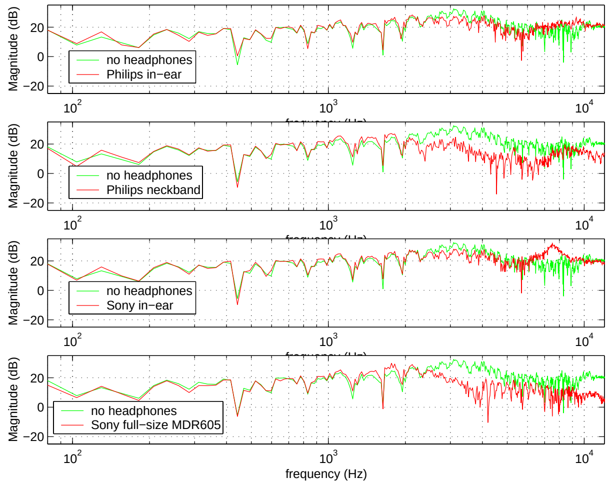


Fig. 7: Coloration of other headphone types.

are point sources. If the headphone driver would coincide with the grill on which the microphone is mounted, the distance d corresponds to the tube length. The graphs on the left side of the figure are made for $d = 10$ mm so that the noise is cancelled at the position 10 mm to the left of the headphone. For the frequencies that are considered there is not much noise cancellation at the position of the ear canal entrance.

There are two fundamental reasons for this. The obvious one is that the error microphone is too far from the ear canal entrance. The second one is by far more important however and can be explained from the graphs on the right side of the figure which are made for $d = 23$ mm. The zone of quiet is substantially larger here, showing that it is favorable to put the noise cancellation loudspeaker driver (the headphone in this case) further away from the error microphone. An observation that may be non-intuitive is that the width of the zone of quiet is more or less constant over a wide frequency range (100-10000 Hz). This can be explained as follows. The amplitude of sound decays with the inverse of the distance to the sound source. Far away from the sound source the sound is therefore not attenuated much if one moves over a certain distance. In the simulation the error microphone is very close to the headphone however. Small displacements give huge differences in the noise power that originates from the headphone. In the graphs on the left of Fig. 8 for example the noise level at the position half way in between the microphone and the headphone is the same as when the headphone would be absent (35 dB). This is true no matter how low in the frequency is. For very low frequencies the noise signal at the position of the microphone is for example r. The headphone should therefore make $-r$ at this position to achieve noise cancellation. This means that the headphone makes approximately $-2r$ half way in between the microphone and the headphone, so that the total signal will be roughly equal to $-r$ at this position. Although these simulations give some indication about the zone-of-quiet, real measurements are necessary as the headphone is not a point source and the sound does not propagate in free space.

Dummy head measurements

The influence of the tube length is measured next. In the measurement, a thin tube of 39 mm length is used in the HP890 prototype headphones and no foam is applied. The headphones are placed on the dummy head[6]. Once the calibration is completed, a noise signal is played back through the loudspeaker and later through the simulated loudspeaker using headphones. The noise is picked up using the left microphone inside the headphones. The power spectrum is computed for the microphone signal corresponding to both the loudspeaker and the virtual

[6]For these measurements, the KEMAR dummy head was used as it has an 'ear-canal' (Zwislocki coupler).

loudspeaker (headphones). The difference between these two represents the error power and is shown in the top graph of Fig. 9. The vertical axis runs from -3 dB to 3 dB. If the curve goes outside these limits, this means that the noise cancellation actually increases the noise for those frequencies. It can be seen that the noise suppression is excellent at the position of the error microphone. This experiment is repeated using the left microphone inside the dummy head. The graph shows that the noise cancellation is still quite accurate up to 10 kHz on the position of the dummy head's 'ears' (measurement microphones).

This measurement is a repeated for various tube lengths using the dummy head's 'ears'. The results are shown in Fig. 10. It can be seen from these graphs is that increasing the tube length beyond 24 mm does not improve the performance. When the tube is made shorter than 24 mm the performance rapidly decreases. A tube length of 5 mm for example only gives good results up to 3 kHz. Based on these results the tube length is fixed

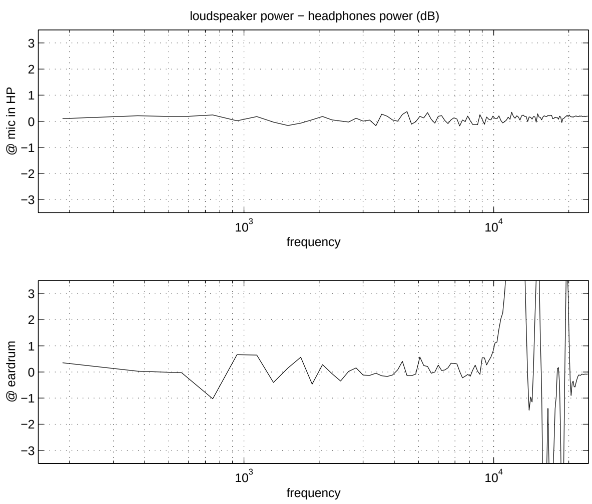


Fig. 9: Difference between the power of noise played through loudspeaker and the power of noise played through virtual loudspeaker (dB) a tube length of 39mm measured at the headhones' microphone (*upper*) and measured at the dummy head's microphones (*lower*). The vertical axis runs from -3.5 dB to 3.5 dB.

at 22 mm. At this length, the tubes do not touch the concha for most people when the headphones are put on properly. Note that the tube does not go into the ear canal at this length. The calibration is repeated while

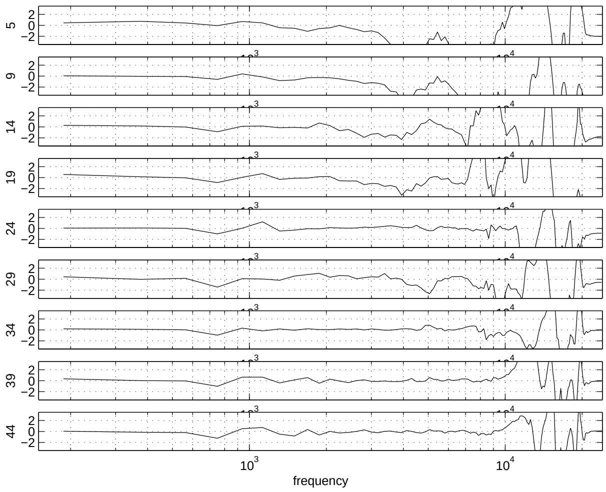


Fig. 10: Difference between the power of noise played through loudspeaker and the power of noise played through virtual loudspeaker (dB) for various tube lengths measured at the dummy head's microphones. The graphs are shown for tube lengths of 5,9,14,19,24,29,34,39, and 44 mm. The vertical axis runs from -3 dB to 3 dB.

recording both the left microphone of the the dummy head and the left microphone of the headphones. Spectrograms of these signals are shown in Fig. 11 on page 11. The top-left one corresponds to the microphone in the headphones and the lower-left one corresponds to the microphone inside the dummy head. The horizontal axis corresponds to time (s) and the vertical axis corresponds to frequency (Hz). The gray scale indicates the power of the signals (dB). The adaptation is started at about 7 s. It is apparent from this figure that noise is canceled over a wide frequency range at the position of the microphone inside the headphones. At the position of the dummy head's 'ear-drum', noise cancellation is only achieved upto 7 kHz (lower left spectogram). From the dummy head measurements it appears that the tube should be somewhat longer. The dummy head is quite different from the human head however. Also, the headphones fit better (i.e. closer) on a human head. The two spectrograms on the left side of Fig. 11 are also normalized on their initial energy and they are shown on the right side of the figure. From these graphs the accuracy of the noise cancellation can be seen. At the position of the microphone inside the headphones this accuracy is more than 20 dB for almost all frequencies. There is almost no noise to be canceled to start with for the frequencies where less cancellation is achieved. This can be seen from the left graphs. At the position of the 'ear-drum' of the dummy head the noise cancellation is somewhat less but still close to 20 dB. The frequency range is limited to 7 kHz however. The figure also shows that the convergence time that is needed to achieve about 20 dB accuracy is about 5 s for most frequencies. Note that in the adaptive filter there is a trade off between the convergence speed and the final noise suppression. For the very low frequencies the convergence time is higher however due to the poor signal-to-noise ratio. Below roughly 100 Hz the room noise is stronger than the measurement noise.

INFLUENCE OF THE FILTER LENGTH

The algorithm consists of a calibration stage and a sound reproduction stage. The sound reproduction stage is computationally more intensive than calibration stage in which the filters corresponding to one loudspeaker are computed.

The filter length should be chosen such that most of the acoustic impulse response is captured at a given sample frequency. The first part of the impulse response consists of propagation delay and the direct path from the loudspeaker to the headphones and some early room reflections. This part is typically about 20 ms long. The second part of the impulse response contains the reverberation of the room. The reverberation time T_{60} is defined as the time that is required for sound presure level to drop 60 dB after switching off the sound source. This reverberation time depends on the size of the room and the furniture that is inside in a frequency dependent manner. The reverberation time of a living room is typically 500 ms [12]. The preferred reverberation time values for various enclosure types are shown in Table 1 [13]. The question that arises now is how long the filter should be in order to the capture most of the reverberation energy for a particular room with a known reverberation time.

Table 1: Preferred reverberation time values

enclosure type	T_{60} (s)
living room	0.5
cinema	0.7–1.0
theatre	0.9–1.3
chambre music hall	1.2–1.5
opera house	1.2–1.6
concert hall	1.7–2.3
church	1.5–2.5

The sound pressure level decays roughly exponentially. The relative amount of reverbation energy that is not captured using a filter of length T_{60} is -60 dB. The decay is linear on a logarithmic scale so that the relative amount of reverbation energy that is not captured using a filter of length αT_{60} is[7]

$$\alpha T_{60} = -\alpha 60\text{dB}, \tag{2}$$

with $\alpha > 0$. Small rooms that have carpet on the floor have a T_{60} of about 0.25 s. Filters of length 50 ms correspond to 2048 coefficients at 44100 Hz. Only 30 ms is used for modeling the reverberation. From (2) it follows that the reverberation is modeled with an accuracy of 7.2 dB which suffices in practice. The same filter length gives only 3.6 dB accuracy in a typical living room ($T_{60} = 0.5$ s) which is insufficient in practice. Filters of length 100 ms will do in most small to medium sized rooms. The headphone simulation will approach the loudspeaker playback more and more when the filter can capture more of the tail of the impulse response (i.e. the reverberation). Listening experiments are required to assess the amount of reverberation that needs to be captured in order to provide transparency.

CONCLUSIONS

The performance of a system that delivers multichannel sound using headphones is analyzed. The system is calibrated using active noise cancellation techniques. It is found that the error microphone should be very close to or even inside the ear. Its relative position with respect to the ear and the headphones determines the upper bound on the frequency for which the system gives convincing results. Frequencies up to 7 kHz can be processed with a high accuracy for practical configurations. Higher frequencies, up to 10 kHz, can be dealt with too if a thin tube is used that runs somewhat into the outer ear.

REFERENCES

[1] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization.* 1996.

[2] Henrik Moller. Fundamentals of binaural technology. *Applied Acoustics*, 36:171–218, 1992.

[3] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. i: Stimulus

[7] The normalized energy in the reverberant tail of an exponentially decaying transfer function with time constant τ from t_1 to t_2 is
$-\frac{\tau}{2}\int_{t_1}^{t_2} e^{\frac{-2t}{\tau}}\,dt = e^{\frac{-2t_1}{\tau}} - e^{\frac{-2t_2}{\tau}}$
with $T_{60} = 3\tau\ln(10)$. For $t_1 = \alpha T_{60}$ and $t_2 = \infty$ an energy of $-\alpha 60$ dB is obtained.

synthesis. *J. Acoust. Soc. Am.*, 85(2):858–867, Feb. 1989.

[4] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. ii: Psychophysical validation. *J. Acoust. Soc. Am.*, 85(2):868–878, Feb. 1989.

[5] Henrik Moller et al. Binaural technique: Do we need individual recordings? *J. Audio Eng. Soc.*, 44(6):451–469, June 1996.

[6] A.W. Bronkhorst. Localization of real and virtual sound sources. *J. Acoust. Soc. Am.*, 98:2542–2553, 1995.

[7] M. Morimoto and Y. Ando. On the simulation of sound localization. *J. Acoust. Soc. Jpn. (E) 1*, pages 167–174, 1980.

[8] R.M. Aarts. Headphones with integrated microphones. *patent app. WO01/49066*, 1999.

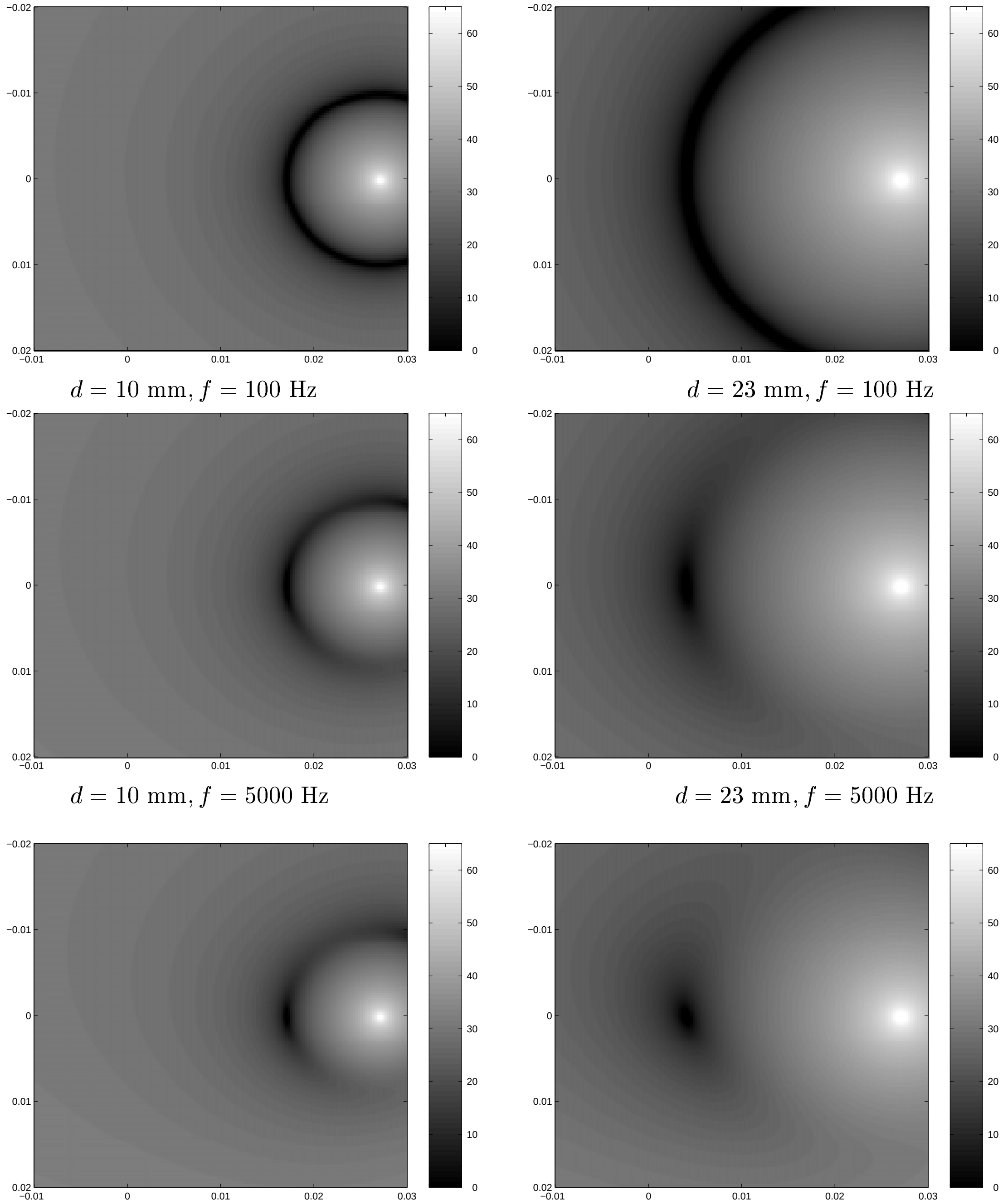
[9] Scott D. Snyder. *Active Noise Control Primer*. Springer-Verlag, 2000. ISBN 0-387-98951-X.

[10] P.A. Nelson and S.J. Elliott. *Active control of sound*. Academic Press, 1992.

[11] J. Garas and P.C.W. Sommen. The all-pass filtered-x algorithm. In *Proc. EUSIPCO, Rhodos, Greece, pp. 101-104*, September 1998.

[12] M.A. Burgess and W.A. Utley. Reverberation times in British living rooms. *Applied Acoustics*, 18:369–380, 1985.

[13] M.M. Boone, D. de Vries, and A.J. Berkhout. *Sound control*, volume 1. Delft university of technology, 1995. College diktaat, vakgroep akoestiek.



$d = 10$ mm, $f = 100$ Hz

$d = 23$ mm, $f = 100$ Hz

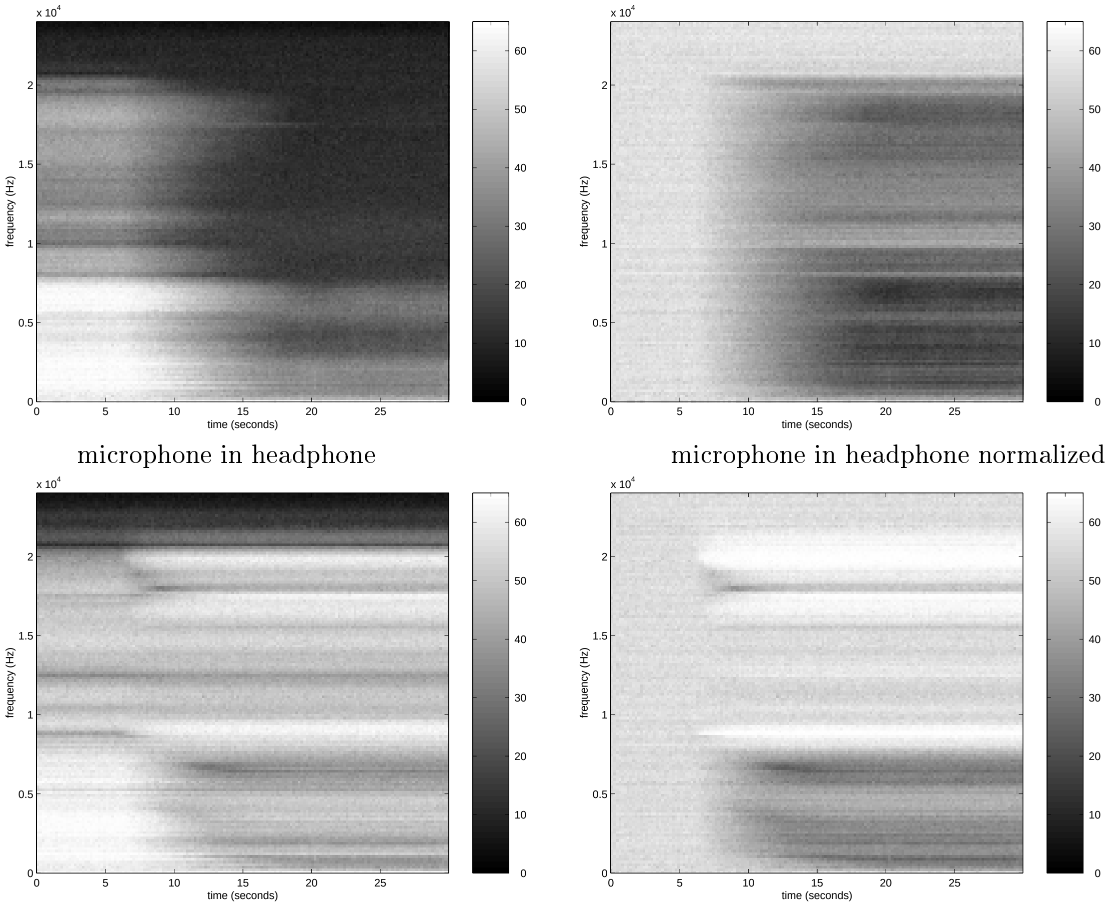
$d = 10$ mm, $f = 5000$ Hz

$d = 23$ mm, $f = 5000$ Hz

$d = 10$ mm, $f = 10000$ Hz

$d = 23$ mm, $f = 10000$ Hz

Fig. 8: Noise cancellation simulation, the noise source is at position $x = 1.00$ m, $y = 1.75$ m (at the lower right of the origin, this is the left front loudspeaker in a ITU configuration), the error microphone is at $x = 0.017$ m, $y = 0$ m (*left graphs*) and $x = 0.004$ m, $y = 0$ m (*right graphs*), corresponding to a tube length of 10 mm and 23 mm respectively. The headphone is at position $x = 0.027$ m, $y = 0$ m. Only the noise suppression at the origin is of interest, this is where the ear is located. The color bar has a dB scale. Without the secondary sound source the noise level at the origin is approximately 35 dB.



microphone in headphone

microphone in headphone normalized

microphone in dummy head

microphone in dummy head normalized

Fig. 11: Spectogram showing Active Noise Cancellation convergence for left side of the HP890 headphone prototype. The top-left one corresponds to the microphone in the headphones and the lower-left one corresponds to the microphone inside the dummy head. The horizontal axis corresponds to time (s) and the vertical axis corresponds to frequency (Hz). The color indicates the power of the signals (dB). The adaptation is started at about 7 s.