ARCHIVES OF ACOUSTICS
31, 4 (Supplement), 343–348 (2006)

INFLUENCE OF ACOUSTICAL PROPERTIES OF A LISTENING ROOM ON THE CONSISTENCY BETWEEN SOUND FIELDS REPRODUCED USING STANDARD STEREO TECHNIQUE AND HEADPHONE TECHNIQUE WITH HRTF PROCESSING

J. SAWICKI, W. MICKIEWICZ

Szczecin University of Technology
Institute of Electronics, Telecommunication and Computer Science
26 Kwietnia 10, 71-126 Szczecin, Poland
e-mail: jerzy.sawicki@ps.pl, witold.mickiewicz@ps.pl

(*received June 15, 2006; accepted September 30, 2006*)

The analysis of acoustical properties of three different listening rooms and its influence on the subjective and objective fidelity and consistency between sound fields reproduced using two loudspeakers in standard stereo configuration and headphone feed with the signal processed using a method of sound externalization developed by the authors based on the modified HRTF (Head Related Transfer Function) technology are presented in the paper. The experiments allow more deep insight into the sound source localization by a human in a closed space and an improvement of the processing algorithm for the sound presentation using headphones.

Key words: HRTF, 3D sound, virtual acoustics.

1. Introduction

Authors of the paper are involved in the development of a headphone processor to externalize 2-channel stereo sound reproduced with headphones. The theoretical background of the idea is described in [1, 2]. The main goal of the present research is the development of the measurements procedure to evaluate the parameters of processing algorithm based on fast convolution for individual person and its individual listening room, and not using a laboratory equipment and anechoic conditions as in [2]. All the needed parameters can be represented as the transfer function called by the authors Head & Room Related Transfer Function (H&RRTF) [5, 6] and also known as the Binaural Room Impulse Response (BRIR). Using the individual measurements gives better results than an application of average data or a dummy head.

The goal of the research described in this paper is to find the answer to the question, if the size and the acoustical properties of the listening room influence on the consistency between sound fields reproduced using a standard stereo technique and a headphone technique with H&RRTF processing. It is important from the practical point of view, if the proposed method should be used to design an electronic appliance to externalize headphone sound reproduction, which destination is a wide consumer market. It is also an important and interesting question from a cognitive point of view, to enhance our knowledge how exactly the human evaluates direction and distance of perceived sound [3, 4]?

The results of subjective listening tests are presented in the paper. The tests consisted in comparing perceived sound source localization, when the sounds were emitted by loudspeakers and by headphones with proper processing.

2. Measurements

The set of measured H&RRTF's was collected from 3 different environments (rooms) and for 3 subjects. Environments 1 and 2 were the ordinary office rooms and have the same size: $2.8\ \mathrm{m} \times 5.7\ \mathrm{m} \times 3.1\ \mathrm{m}$ (width × length × height) but different acoustic adaptation: room 1 has wideband sound absorbers on the walls to attenuate first reflections; room 2 has wideband sound absorbers on the rear of the listeners (Fig. 1).

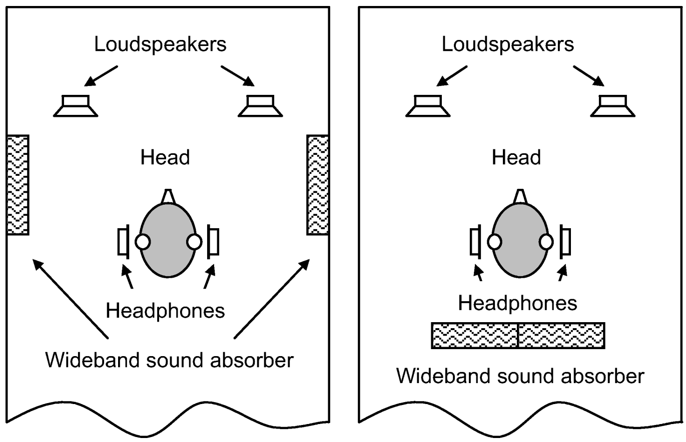

Fig. 1. Wideband sound absorber positioning in environment 1 (left) and environment 2 (right).

In both rooms, there was carpet on the floor and painted plaster on the ceilings. The average reverberation time of both rooms was the same and about 0.45 sec. Environment 3 was a building corridor with size of $2.5\ \mathrm{m} \times 25\ \mathrm{m} \times 3.1\ \mathrm{m}$ (width × length × height) with hard and strong reflecting walls, floor and ceiling (Fig. 2). The reverberation time was about 2.3 sec. The loudspeakers-listener set-up was the same in all three environments. Loudspeakers and listener head's positions created an equilateral triangle with distance between corners of 1.8 m. The height of listener ears and tweeters in the loudspeakers was near the same (about 1.3 m over the floor). The sound pressure level of the test signal measured near the head was 60 dB SPL. The impulse responses were recorded using miniature electrets microphones located in the entry of the ear channels. The average distance between microphones was 18 cm. The signals from each microphone after amplification in a low-noise custom-made microphone preamp was connected to the line inputs of the Digidesign audio interface Digi002. For each subject (with normal hearing and some experience in subjective audio tests) four kinds of responses to 2^{14} samples long MLS signal (sampling frequency 44.1 kHz) were recorded: two responses from left and right ear to MLS emitted by left loudspeaker and two responses from left and right ear to MLS emitted by right loudspeaker. Using fast Hadamard transform algorithm, impulse responses were calculated. All calculation and processing were carried out using Matlab software.

Fig. 2. The measurement (and listening) set-up in environment 3 – building corridor.

3. Listening tests and subjective evaluation

For the listening tests, the pink noise signal was used to create testing sounds. The tests were conducted in two stages. The loudspeakers-listener set-up was the same as during measurements done before. In the first stage, two kinds of sounds were presented to the listener in random sequence:

- original pink noise signal without any processing, emitted by one loudspeaker (left or right),
- pink noise signal convolved with H&RRIR measured for actual subject and environment, emitted by headphones.

Each listener was asked to point out, if the sound was emitted by loudspeakers or by headphones and evaluate the perceived distance to sound source. In most cases the listeners answers were correct, but also a few mistakes were present. Listeners reported that the different timbre of the noise helped to distinguish the source. The sound image created by headphones was externalized, but the perceived distance was shorter then actual distance to loudspeaker. During the experiments, it was noticed that the correctness of the recognition process depends on the test signal order during presentation. The unprocessed sound (pink noise mono signal without any filtering) emitted by headphones before the testing signals mentioned above, influenced results of recognition – the listeners have subjective sensation of much more consistency between localization of phantom sound sources created by emission the previous test signals after the emission of the unprocessed noise. So in the second stage 3 kinds of sounds were presented to the listener. The sequences of test signals as before were always preceded by the unprocessed pink noise mono signal emitted by headphones (called in the next part of the paper as "calibration signal"). It was examined, that a time gap from zero to few seconds between calibration signal and test signals does not influence the results of the experiment.

After the emission of each signal sequence (3 signals), each listener was asked again to point out the source of sound (loudspeaker or headphones) and then asked to perceived distance to sound source. In contradiction to previous experiments, in most cases the listeners could not properly distinguish when the testing sounds were emitted by loudspeaker or when by headphones. In this case, unlike before, the different timbre of the noise not help to distinguish the sources. The sound image created by headphones was externalized and the perceived distance was often the same as the actual distance to the loudspeaker. Detailed results for all listeners are presented in the Table 1.

Table 1. Changes in perceived distance and size of phantom sound source.

Listener	Environment 1		Environment 2		Environment 3	
	Left	Right	Left	Right	Left	Right
JS	$X + \Delta X$	X	$X - \Delta X$	$X - \Delta X$	X	X
	N	N	C	C	N	N
GS	X	X	$X + \Delta X$	$X + \Delta X$	$X + \Delta X$	X
	N	N	C	C	C	N
WM	X	X	$X - \Delta X$	$X - \Delta X$	$X - \Delta X$	$X - \Delta X$
	N	N	C	C	N	N

$X - \Delta X$ – perceived distance sorter than in real conditions;
X – perceived distance the same as in real conditions;
$X + \Delta X$ – perceived distance longer than in real conditions;
N – phantom sound source narrower than in real conditions;
C – phantom sound the same size as in real conditions.

Table 1 contains averaged results for 3 listeners (JS, GS and WM), who have listed to 10 sequences of test signals with random order of sounds emitted by headphones and loudspeakers. The highest consistency between the sound images perceived when test signals were emitted by a loudspeaker and headphones (after processing) was observed in environment 1. Proper acoustical adaptation and especially the decreasing the first reflections from walls using wideband absorbers placed as shown in Fig. 1 (left) could cause it. Practically all listeners evaluated the phantom sound source (generated by headphones processor) as located in the same distance as loudspeaker. It should be noticed, that the size of phantom sound source was evaluated as little smaller then if they were produced by loudspeaker. A higher inconsistency in distance perception was observed in the environment 2. The evaluations were symmetric (the same for left and right channel) but the differences appeared in the evaluation of the distance (the phantom source was closer or further than the real loudspeaker). The perceived sizes of the sound images were the same. It should be noticed, that in the environment 2 the average reverberation time was the same as in environment 1, but different absorbers placement as is shown in Fig. 1 (right) created a different profile of early reflections. Some interesting results were obtained for environment 3. The distance was evaluated differently for each listener but the size of phantom source was evaluated rather as narrower. The listeners reported also sometimes increased distance between phantom images created by headphones for left and right channel. The reason for higher inconsistency of the results in this case was the length of used impulse responses, which from practical reasons (processing time) were shorten to 12000 points for all environments. For the environment 3 with 2.3 sec. reverberation time, about 10 times shorter impulses were not long enough to make the proper distance rendering.

Although it was not the subject of the test, listeners reported sometimes changes in perceived elevation of the phantom sound sources. The consistency in angular reproduction in the horizontal plane was quite fully consistent.

4. Discussion and conclusions

The experiments proved that the headphone processor, based on H&RRIRs measured in different listening environments, can create auditory sensations using headphones that are comparable with real listening conditions using loudspeakers and maintain some spatial attributes of the real sound image created by loudspeakers.

It was observed the important role of the introductory "calibration" of the human hearing system, which in some cases meaningfully improve the consistency between subjective localization of the phantom sound sources generated artificially (using headphones processor) and localization of sound images emitted actually by loudspeakers. It proves that the human hearing short-time memory plays very important role in the sound events perception and localization. In the mechanism of a distance perception very important is the reference information stored in the human memory. A person being in a given environment subconsciously "calibrates" its hearing perception system (also using visual perception system) and on this base evaluates distance to appearing new sound events. The goal for further research can be evaluation of time characteristics of these effects.

Our experiments conducted in different acoustical environments prove that proposed method of joined measurement of acoustical properties of the room and cumulative influence of individualized pinna-head-and-torso (measurement of Head&Room Transfer Related Function), works well in externalization of sound image using headphones practically for every acoustical environment, if the measured impulse responses for convolution process have adequate length. If the user wants to render sound images using headphones processor, that are consistent with these existing in environments with high reverberation, long convolutions have to be done. Finding the exact relation of reverberation time to desired convolution length for consistent distance rendering will be the goal of further research. If the processor should work in real time, it will desire the application of specific fast convolution algorithms and high-power processing hardware.

Acknowledgments

The authors are grateful to Mr. Grzegorz Szubartowski for his support and the listening test participation.

References

[1] BLAUERT J., *Spatial hearing* (revised Edition), MIT Press, MA, 1997.

[2] KISTLER D. J., WIGHTMAN F. L., *Headphone simulation of free-field listening: I. Stimulus synthesis, II. Psychophysical validation*, J. Acoust. Soc. Am., 858–878 (1989).

[3] ZAHORIK P., *Auditory display of sound source distance*, Proc. of 2002 International Conference on Auditory Display, Kyoto, Japan 2002.

[4] OZIMEK E., *Dźwięk i jego percepcja. Aspekty fizyczne i psychoakustyczne*, Wydawnictwo Naukowe PWN, Warszawa – Poznań 2002.

[5] MICKIEWICZ W., SAWICKI J., *Headphone processor based on individualized HRTFs measured in listening room*, 116-th AES Convention, Berlin 2004.

[6] MICKIEWICZ W., *Subjective performance evaluation of the individualized headphone processor used to auralization 2-channel stereo sound*, Archives of Acoustics, **30**, 4 (Supplement), 119–122 (2005).