Witold MICKIEWICZ, Jerzy SAWICKI
SZCZECIN UNIVERSITY OF TECHNOLOGY, INSTITUTE OF ELECTRONICS, TELECOMMUNICATION AND INFORMATICS

Spatial audio reproduction by headphones using binaural room impulse responses measured individually by the listener

Ph.D. Witold MICKIEWICZ

Adjunct in the Department of Cybernetics and Electronics, Institute of Electronics, Telecommunication and Informatics of Szczecin University of Technology. Dr. Mickiewicz is a specialist in analysis and processing of audio signals, electroacoustic systems and sound engineering.

e-mail: Witold.Mickiewicz@ps.pl

Ph.D. Jerzy SAWICKI

Adjunct in the Department of Cybernetics and Electronics, Institute of Electronics, Telecommunication and Informatics of Szczecin University of Technology. Dr. Sawicki is a specialist in electronic circuits, analysis and processing of audio signals, especially of speech.

e-mail: Jerzy.Sawicki@ps.pl

Abstract

Listening via headphones in opposition to a loudspeaker reproduction introduces changes in the perception of acoustic atmosphere and spaciousness (internalisation effect). This can be changed using the Head Related Transfer Function (HRTF) technology. In the paper there is presented an idea of the headphones processor which uses an individualized Binaural Room Impulse Response (BRIR) measured for a given listener and for a given acoustical environment. There is investigated the influence on the performance of the individualized headphones processor of acoustical properties of the room, length of the BRIR and electroacoustical chain quality. The main goal of this research was to evaluate the minimum requirements, which have to be fulfilled by the processing algorithm to obtain a good subjective performance.

Keywords: acoustical measurements, digital audio signal processing, BRIR, HRTF, 3D sound, virtual acoustics.

Przestrzenna reprodukcja dźwięku przez słuchawki z wykorzystaniem dwuusznych odpowiedzi impulsowych pomieszczenia rejestrowanych indywidualnie przez słuchacza

Streszczenie

Odsłuch słuchawkowy w porównaniu z odsłuchem przez głośniki wprowadza zmiany w postrzeganej atmosferze akustycznej i przestrzenności nagrania (efekt internalizacji). Można to zmienić korzystając z technologii HRTF (ang. Head Related Transfer Function). W niniejszym artykule przedstawiono koncepcję procesora słuchawkowego wykorzystującego dwuuszne odpowiedzi impulsowe pomieszczenia (BRIR) mierzone w konkretnym pomieszczeniu odsłuchowym przez końcowego użytkownika systemu. Przeprowadzono badania dotyczące wpływu akustyki pomieszczenia, długości użytych BRIR i jakości elektroakustycznego toru pomiarowego na skuteczność działania procesora. Głównym celem tych badań było określenie minimalnych wymagań, jakie musi spełnić proponowany algorytm, aby zapewnić zadowalające efekty subiektywne.

Słowa kluczowe: pomiary akustyczne, cyfrowe przetwarzanie sygnałów audio, BRIR, HRTF, dźwięk 3D, akustyka wirtualna.

1. Introduction

The headphone processor considered in this paper shall be an electronic appliance that processes a 2-channel audio signal in such a way, that the signal reproduction through headphones creates auditory sensations, which are normally perceived during listening to a 2-channel audio signal via 2 loudspeakers in a listening room with some acoustical properties. Especially maintaining the "out-of-head" effect is desirable [1]. To realise the externalization task the processing algorithm models physical phenomena, which take place in the listening room and influence the audio signal on its path from loudspeakers to human ears (eardrums). The main effect perceived by the listener is the externalisation of the auditory scene and its shift to a stereo base between the loudspeakers. The headphone processor applies a transfer function, which is a superposition of: a) directional and frequency characteristics of the loudspeakers, b) reflection characteristics of the surfaces in the listening room, c) differences in the arrival time for the direct and reflected sound, and d) HRTFs (Head Related Transfer Functions [2, 3, 4, 5]) for each direction of a direct and indirect sound. Such cumulative impulse responses are called the *Binaural Room Impulse Responses* (BRIRs) or the *Head and Room Related Transfer Functions* (H&RRTFs) [6, 7]. This BRIR philosophy is shown in Fig. 1.

a)
b)

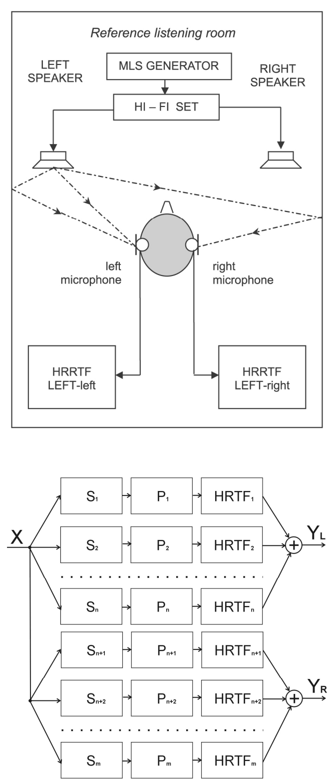

Fig. 1. The concept (a) of the H&RRTFs as a cumulative transfer functions (b) containing information about directional characteristics of the sound source (S_i), transfer functions of sound paths in the listening room (P_i) and the pinna-head-torso influence ($HRTF_i$); X- the input audio signal, Y_L,Y_R –signals near the ear canal entries

Rys. 1. Koncepcja (a) funkcji przenoszenia związanych z głową i pomieszczeniem jako skumulowanych funkcji przejścia (b) zawierających informacje o kierunkowości źródła dźwięku (S_i), funkcji przejścia związanych z poszczególnymi drogami rozchodzenia się dźwięku (P_i) oraz wpływu układu małżowina-głowa-tułów ($HRTF_i$); X-wejściowy sygnał audio, Y_L,Y_R – sygnały w pobliżu wlotu do kanału słuchowego

Transfer functions (impulse responses) are to measure in any listening room (Fig. 2a), where the stimuli are radiated through the loudspeakers and the measurement microphones are placed in the ear canals of the listener (Fig. 2b).

a) b)

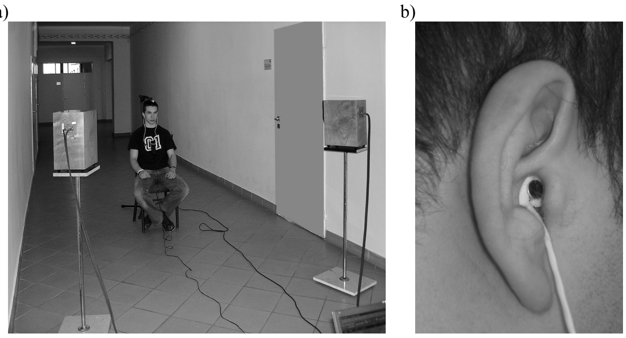

Fig. 2. The example of measurement (and listening) set-up in an acoustically non-adapted environment with difficult sound reproduction conditions (a); the measurement microphone placed in the listeners ear canal (b)
Rys. 2. Przykładowe ustawienie w trakcie pomiaru (i odsłuchu) w nieadaptowanym akustycznie otoczeniu o trudnej akustyce (a); widok miniaturowego mikrofonu pomiarowego umieszczonego w kanale słuchowym słuchacza (b)

The main requirement for a proper spatial reproduction of an audio signal via headphones is the precise measurement of the BRIR. There are many measurement methods described in literature which differ in a) the measuring sensor placement and b) excitation signal. The authors of this paper used the MLS signal as the excitation [8]. The exponential sinusoidal sweep can be also used. An example of one BRIR pair for the test signal emitted by the right loudspeaker is shown in the time and frequency domains in Fig. 3.

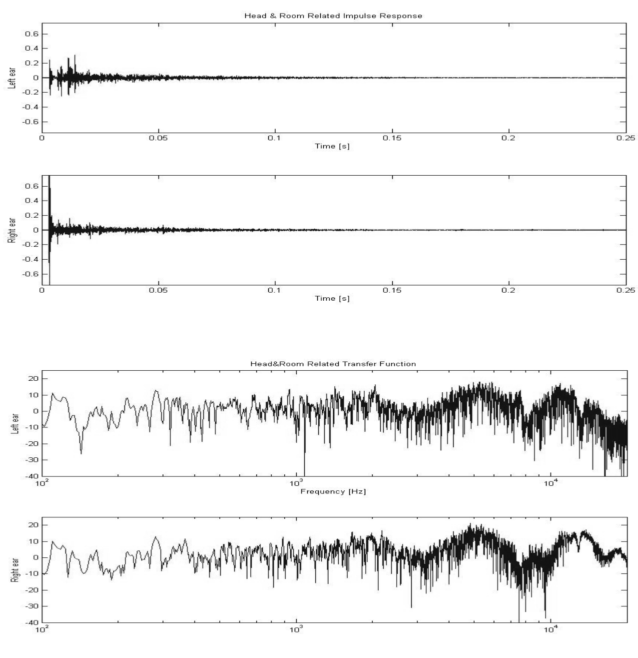

Fig. 3. The example of a BRIR pair (for the left and right ear) in time and frequency domains
Rys. 3. Przykładowe dwuuszne odpowiedzi impulsowe pomieszczenia (dla lewego i prawego ucha) przedstawione w dziedzinie czasu i częstotliwości

The individualized headphones processor convolves an audio signal (e.g. from a CD-player) with the measured BRIRs. The signal flow in the 2-channel processor is shown in Fig. 4. The correction blocks ($CORR_i$) are there to compensate irregularities in frequency characteristics of the BRIR measurement system. It will be discussed later.

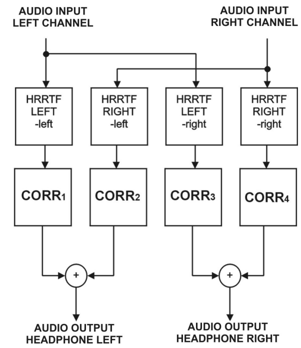

Fig. 4. The signal flow in the 2-channel headphones processor
Rys. 4. Przepływ sygnałów w słuchawkowym procesorze 2-kanałowym

Because of computational complexity of the convolution algorithm, practical realization of the headphones processor as an appliance working in real time should take into consideration some additional factors. In the following part of the paper some experiment results are presented allowing the analysis of the influence of listening room acoustics on the headphones processor performance, the influence of the BRIR shortening on the perceived sound spaciousness and the correction blocks performance used to compensate irregularities in frequency characteristics of the measurement chain, especially of the electrets microphones used as the measurement sensors.

2. Experiments concerning processor performance

2.1. The influence of the room acoustics

To evaluate the efficiency of the headphones processor to simulate different listening conditions via headphones the BRIR measurements were conducted in 3 different environments (rooms) and for 3 listeners. Two environments simulated acoustical properties of typical listening conditions at home. Third one, for the contrast, represented a difficult acoustical environment with strong reflections and long reverberation. The recorded impulse responses were then used to create testing signals. During the tests listeners were asked, if the localization of phantom sound sources generated by headphones is concurrent with the localization of sound sources created by loudspeakers.

The loudspeakers-listener set-up was the same in all three environments. Loudspeakers and listener head's positions created an equilateral triangle with the distance between corners of 1.8m. The height of listener's ears and loudspeaker's tweeters was near the same (about 1.3m 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 canals. The average distance between microphones was 18cm. 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 listener (with normal hearing and some experience in subjective audio tests) four kinds of responses to the 2^{14} samples long MLS signal (the sampling frequency 44.1kHz) were recorded: two responses from the left and right ear to the MLS emitted by the left loudspeaker and two responses from the left and right ear to the MLS emitted by the right loudspeaker. Using the fast Hadamard transform algorithm, impulse responses were calculated [8, 9, 10]. All calculations and processing were carried out using a Matlab software.

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

- original pink noise without any processing, emitted by one loudspeaker (left or right),
- pink noise convolved with the BRIR measured for the actual listener and environment, emitted by headphones.

The experiments have shown, that the sound image created by headphones was correctly externalized and the perceived distance was nearly the same as the actual distance to the loudspeaker. This effect is rarely to achieve in processors being on the consumer market. Averaged results for all listeners are presented in table 1.

Tab. 1. Perceived distance and width of a phantom sound source generated by the headphones processor comparing to the loudspeaker listening
Tab. 1. Postrzegana odległość i szerokość pozornych źródeł dźwięku przy odsłuchu przez procesor słuchawkowy w porównaniu z odsłuchem przez głośniki

Subjective feature	Environment 1		Environment 2		Environment 3	
	Left	Right	Left	Right	Left	Right
Distance to the phantom source	Correct	Correct	Correct	Correct	Correct	Correct
Width of the phantom source	Narrower	Narrower	Correct	Correct	Narrower	Narrower

Our experiments conducted in different acoustical environments prove that the method of the joined measurement of acoustical properties of the room and cumulative influence of the individualized pinna-head-and-torso (measurement of BRIR), works well in externalization of sound image using headphones practically for every acoustical environment.

2.2. The influence of the BRIR's length

By convolving an input audio signal with properly measured BRIRs in certain room and in certain listeners ears the effect of spatialization was practically always achieved. However it required a huge computational effort because even in medium sized rooms the recorded impulse responses were about thousands samples of length using the standard 44,1 kHz sampling frequency. Even using fast convolution algorithms (e.g. partitioned FFT [11]), processing in real time requires very fast (and expensive) signal processors. So the question was, if for decreasing a computational power demands the BRIRs length could be shorten and simultaneously the externalization performance of the headphone processor could be maintain. So the authors wanted to examine, how much the BRIRs can be shorten after measurements to maintain the externalization properties of the processing algorithm. During the experiment 11 listeners were listening the test signal containing 30 pairs of 2-seconds processed pink noise. First signal in each pair was always the master signal – i.e. pink noise convolved with full length BRIRs (15000 samples), the second signal in each pair was calculated using the same pink noise and the BRIR with a variable length (before processing the Hann window with 3000 to 8000 samples was applied to shorten the BRIR). The step length was 500 samples. During the test the pairs were emitted via headphones in a random order. In one test with 30 pairs the signal which was shorten to certain length appeared 3 times on average. The listener task was to answer, if she/he detects the difference in spatial properties (especially the distance to phantom sources) of the differently processed signals in the pair. Cumulative results of the experiment are shown in table 2. The experiment showed, that for a certain acoustical environment there is a boundary BRIR length called the transition region. If the BRIR is longer than the boundary value, the processor performance remains constant, and if the BRIR becomes shorter, the performance of the processor gets worse. In our experiment the boundary length was over 2 times shorter than the full length of the measured BRIR. Therefore there is a possibility to optimize the BRIR length to achieve the lowest computational complexity of the algorithm and still maintaining its highest performance. For the room used during experiment the boundary length corresponded to the time, during which the reverberation decay curve calculated by the back integration of the BRIR decreased of about 20 dB. This relationship will be investigated in further research.

Tab. 2. Perceived differences in sound image (the distance to phantom sources) in function of the BRIR length
Tab. 2. Postrzeganie zmiany wrażenia słuchowego (szczególnie dystansu do pozornych źródeł dźwięku) podczas skracania długości BRIR

BRIR length [samples]	Percentage of perceived differences [%] (0%-nobody detected the difference, 100% - the difference was detected in all trials)	Evaluation
8000	0	Length sufficient for the majority of listeners
7500	0	
7000	11	
6500	18	
6000	50	Threshold of sensitivity (transition region)
5500	26	
5000	61	
4500	60	
4000	83	Length insufficient for the majority of listeners
3500	81	
3000	96	

2.3. The influence of correcting blocks

When the experiments with pink noise were completed, headphone processor was used to externalize real music and speech signals. The listeners reported very good externalization effects. Many listeners had problems to distinguish, when the processed signal is transmitted via headphones and when original signal via loudspeakers. During the processor tests with music and speech signals some problems appeared. The quality of the processed audio signal was lower than the original signal reproduced via headphones. The quality of original signals was higher (flat frequency characteristics and better signal-to-noise ratio) but the undesirable internalization effect was evident. On the other hand the processed signals were correctly externalized but their quality was typical when the signals would be listen via loudspeakers in a flat. So the people accustomed to the difference between loudspeakers and headphones listening complain of the timbre distortion introduced by headphones processor. However these complaints from the author's point of view are unfounded, because our aim was to simulate listening conditions via loudspeakers, which have known limits. Nevertheless some attempts were made to improve the signal timbre without loosing the processor ability to externalize the sound. It was also considered that BRIR measurement equipment was not ideal and it could influence the quality of the final audio signal feed to headphones. Below the assumptions and experiments results concerning the signal timbre improvement are presented.

During the measurement of the BRIR the frequency characteristics of the test signal emitted by loudspeakers is modified when recorded by microphones located in the ear canals. Except the changes resulting from BRIR processing, which carry localization information, non-ideal electroacoustical chain amplifier-loudspeaker-walls-microphones-preamplifiers-A/D converters can distort the frequency characteristics. The most critical element is the microphone. It have to be small (mountable in a ear canal), rather cheap, and because of mounting conditions it is exposed to vibration also as a whole device, not only its membrane. These factors can cause lower fidelity than expected in ideal conditions.

To counteract these disadvantageous effects certain correction filters can by applied at the processor output (correction blocks in Fig. 4).

The main task for these filters is to flatten the transfer function of the processor in a way, that the whole processing algorithm (together with the BRIR measurement) will change only spatial attributes of the sound image and would not change its timbral characteristics.

An adequate selection of the correction filter characteristics is not a trivial problem.

The correction filter should not eliminate the information about the spatial distribution of sound sources but it should only compensate the irregularities in frequency characteristics of measuring microphones and loudspeakers. The transfer function of correction filters should average and inverse the attenuation effects existing during the BRIR measurement process.

The smoothing process should be done in properly selected frequency bands to not to lose the spatial information included in the BRIR details (the BRIR filter is in general non-minimum-phase and after averaging the correction filter will be minimum-phase). So the basis for evaluating the transfer function can be the smoothed BRIR or H&RRTF (each correction function $CORR_i$ is evaluated on the base of the averaged $HRRTF_{I-i}$ from preceding block). The smoothing can be done in different ways in a time and frequency domain. During experiments the efficiency of the correction filters basing on three BRIR's averaging methods in a frequency domain were examined:

1. linear averaging (referred in the table 3 as LIN)
2. averaging in one-third octave bands (referred in the table 3 as THIRD)
3. averaging in critical bands using bark scale [12] (referred in the table 3 as BARK).

So the averaging process was also conducted, when the frequency scale was partitioned using a human auditory system models.

In Fig. 5 the BRIRs after the averaging process are shown. Using the averaged BRIRs correcting filters were synthesised.

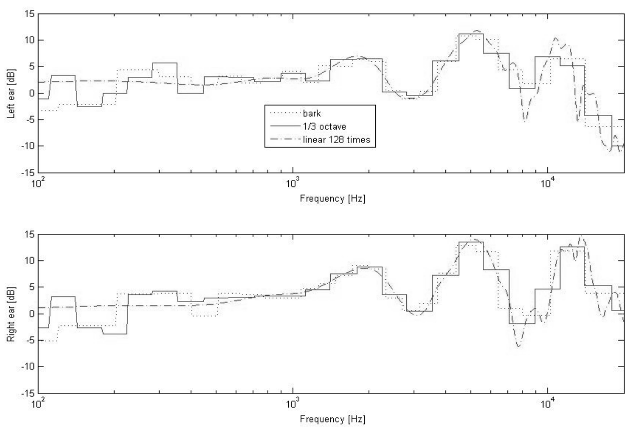

Fig. 5. The BRIR pair after the averaging process in a frequency domain
Rys. 5. Para dwuusznych odpowiedzi impulsowych pomieszczenia po procesie uśredniania w dziedzinie częstotliwości

Subjective tests of the processor with the frequency characteristics correction basing on the described averaging methods did not detect a distinct predominance of only one method. Summary of the subjective test results concerning the timbre correction and the localization correctness is presented in the table 3.

Tab. 3. Influence of correction filters on a sound timbre and source localization (100%=best fit, 0%= no fit).
Tab. 3. Wpływ korekcji na postrzeganą barwę i lokalizację źródeł

Method	Timbre consistency [%]	Localization consistency [%]	Total score [%]
LIN	100	33	67
THIRD	89	56	72
BARK	11	89	50

3. Discussion and conclusions

The concept of the sound spatialization reproduced via headphones has been investigated since 1989 (at least) [3, 13]. Some commercial solutions in the form of a software or hardware-based processors are also available on the market today. However these processors use averaged or modelled BRIRs [14] and the reproduced sound image hardly ever corresponds to the listening experience in real-life situation or by using loudspeakers. The main goal of the authors was to externalize the sound image, which by listening via headphones is normally localized "in the head". It can be done using the headphones processor and BRIRs individually measured in the listening conditions and ears of the end user. An achievement of this target depends on the development of cheap and efficient BRIR measurement methods and cost-effective convolution and correction algorithms.

The authors showed that the headphone processor based on BRIRs measured in a listening room is able to create auditory sensations using headphones that are comparable with real listening conditions using loudspeakers. 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 performs well practically for every acoustical environment.

It is shown, that the satisfactory performance can be achieved also when using BRIRs with a reduced length, what decreases the processor computational load.

The individualized headphones processor ability to externalize headphones sound is high, but for some purists changes introduced to frequency characteristics of the reproduced sounds can be not acceptable and perceived as quality degradation. A solution to this problem can be a development of better correction inverse filters. The correction methods proposed in the paper seem to be not optimal.

4. References

[1] Zahorik P., Auditory display of sound source distance, Proc. of 2002 International Conference on Auditory Display, Kyoto, Japan (2002)

[2] Blauert J., Spatial Hearing (revised Edition), MIT Press, MA, 1997.

[3] 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)

[4] Gilkey R.H., Anderson T.R., Binaural and Spatial Hearing in Real and Virtual Environments, Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[5] Kistler D.J., Wightman F.L., A model of head related transfer function based on principal components analysis and minimum-phase reconstruction, J. Acoust. Soc. Am., March 1992, 1637-1647.

[6] Mickiewicz W., Sawicki J., Headphone processor based on individualized HRTFs measured in listening room, 116th AES Convention, Berlin (2004)

[7] Mickiewicz W.: Subjective performance evaluation of the individualized headphone processor used to auralization 2-channel stereo sound, Archive of acoustics, (2005)

[8] Moon T.K., Stirling W.C., Mathematical methods and algorithms for signal processing, Prentice Hall, 2000.

[9] Elliot D.F, Rao K.R., Fast transforms: Algorithms, analysis, applications, Academic Press, New York, 1982.

[10] Sutter E.E., The fast m-transform: A fast computation of cross-correlations with binary m-sequences, Society for Industrial and Applied Mathematics Journal Comput., Vol.20, No. 4, pp.686-694, (August 1991).

[11] Farina A., Torgel A., Real-time partitioned convolution for ambiophonics surround sound, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2001)

[12] Ozimek E., Dźwięk i jego percepcja. Aspekty fizyczne i psychoakustyczne Wydawnictwo Naukowe PWN, Warszawa - Poznań 2002.

[13] Hawksfords M.O.J., Scalable Multichannel Coding with HRTF Enhancement for DVD and Virtual Sound Systems, J. Audio Eng. Soc., Vol. 50, No. 11, 2002 November

[14] Duda R.O., Modelling head related transfer functions, in Proc. Twenty-Seventh Annual Asilomar Conference on Signals, Systems and Computers, Asilomar CA, November 1993.

Artykuł recenzowany