Acoustic and Biomedical Engineering

Subjective Assessment of the Multi-Channel Auralizations

P. Małecki*, J. Wiciak and J. Wierzbicki

AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics
Department of Mechanics and Vibroacoustics, al. A. Mickiewicza 30, 30-059 Krakow, Poland

The article presents the course and the results of an experiment, which aimed at the subjective assessment of the multi-channel impulse responses. The assessment was conducted considering the usefulness of the received responses for the conducting operation of the digital convolution. The resulting sound material is generated for the simulation of the characteristics of the room. In a medium-sized, rectangular reverberation room (74 m^3) a number of measurements of impulse responses were conducted with the use of multi-channel microphone techniques and with the use of SoundField type microphone. In identical conditions the raw sound material was recorded (in conditions of free field). Next, the convolution was performed between the raw material and the recorded impulse responses. The group of experts, whose members had at least 5 years of experience in the field of sound engineering, was subjected to the psychoacoustic tests aiming at comparison of the sound materials achieved in the convolution and in the recording.

PACS: 43.55.Hy, 43.55.Mc, 43.66.Lj

1. Introduction

The multi-channel impulse responses find their application mainly in the audio-visual industry. In the music and movies production they allow production of any space, room and sound conditions. It should be precised that in most cases the goal of the producers is a creation of an expected space instead of real re-creation of the existing or historical one. The assumption made in this work is an analysis how signal operations, used for simulation of the space, influence the subjective reception of the space.

The main function used for simulation of the room is the technique of the impulse response convolution of any room with the sound signal recorded in conditions of free field. Large capabilities in the field of re-creation and creation of the space lie in the systems of 3D microphones such as the SoundField microphone. They engage additional signal operations that may influence the sound and the space receptions. In the second part of the article the course of the measurements is presented together with the course of the experiment and the hearing tests that is aimed at analysis of the presented problem.

2. The assumptions for the experiments

The main assumption for the described tests is the analysis if the discrete convolution and other signal processing influence the subjective reception of the given space. The difficulty in testing such influence is the fact that while listening to the modified signal that is after convolution or other processing (post-processing) there is no reference signal [1]. The listener can only reference to their own music memory, which in such cases is very unreliable. If the person listened to some recording in the X room then the sound engineer, who presents the recording with the simulation of the X room, has no pattern that could be presented as the reference for the comparison. There are few possibilities that could be used, but each is loaded with consequences. The use of the physical source (e.g. a musical instrument) in the room as the pattern and presentation of the same source recorded in the free field conditions generates the possibility that there will be more differences emerging from the performance of the musician or the method of recording. The biggest problem in such case is the need for movement between the listening room and the tested and compared room. All the combinations and modifications of the presented method will be exposed to such factors, which could add to the ambiguity of the experiment.

This problem has been considered by other scientists [2–4] but their main goal was different than of authors of this paper. Papers [2] and [3] show subjective comparison between two different processing from B-format ambisonic raw audio material. Similar experiment to proposed in this paper was conducted by Kearney and Levison in [4] but using different sound source which caused other problems such as proximity effect etc. that can be avoided in proposed methodology.

To minimise as much as possible the danger of the ambiguity the experiment was proposed in this paper consisting in recording of the pattern and presenting it in the reference listening room together with the signal gener-

* corresponding author; e-mail: pawel.malecki@agh.edu.pl

ated by numerical operations. Previously prepared sound material recorded in the free field conditions was recorded with several different multi-channel systems and the impulse response of the systems was measured. Each system consists of the same source (speaker set) and the spatial microphone system. The precise description of the measurement stand is presented in the following Section.

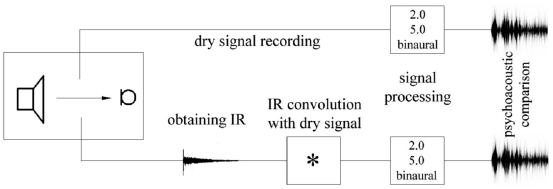

Fig. 1. The schematic of the concept of experiment.

Figure 1 presents the brief concept of the experiment. There were two sets of signals prepared. First set was obtained through recording of dry signal in examined room (upper branch of Fig. 1). Second set of signals was prepared through convolution of the same dry signal and impulse response of the same examined room (lower branch of Fig. 1). The multichannel measuring chain (described in detail in the following section) was identical for both the recording and measuring impulse responses. During recording the signal was received that directly included the response of the room and the electroacoustic chain. Also the convolution of the impulse responses measured with the raw material was performed as well as other necessary signal and edition operations were performed. The achieved sound materials were subjected to the psychoacoustic tests to verify the differences between them.

3. The measurement of the impulse responses and recording of the sound material

Figure 2 presents the schematic of the measurement stand. The experiment was conducted in the medium-sized (74 m^3) rectangular room with small amount of furniture and hard bordering surfaces. The measured reverberation time on the day of recording and measurement is averaged to about 1 s making the impulse response clearly audible especially early reverberations.

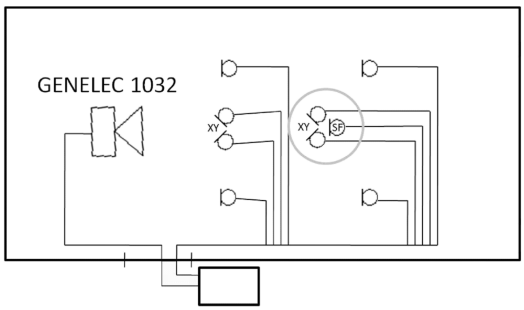

Fig. 2. The schematic of the measurement stand.

As a sound source, both for measurement and recording, the high quality dual-way monitor was used, produced by Genelec, model 1032. The most important parameters of the applied speaker were its high sound pressure level and flat characteristics in the band 42–21 kHz, where maximal deviation does not exceed 2.5 dB [5].

The positions of the microphones are presented in Fig. 2. The SF symbol denotes the SoundField ST350 portable microphone system. The microphone contains four sub-cardioid capsules mounted in a tetrahedral arrangement. The special mutual positioning of the capsules allows measuring the signal of any direction characteristics [6]. Also, in the room, there were installed a number of stereophonic systems, both XY and AB configurations. The stereophonic techniques allow capturing of the apparent spatial image and its re-creation on properly configured listening system [7]. For all the configurations the RODE M3 microphones were used. The microphones are commonly used in the audio-visual industry and are characterised by good parameters and cardioidal directional characteristics without introduction of unwanted audible distortions. The measurements of multi-channel impulse responses were conducted by means of the EASERA 1.1 PRO software. The test signal was 2.7 s long Log-Sweep sine. All the impulse responses (from all the channels) were measured within one measurement with the sampling frequency of 96 kHz and the resolution of 24 bits. The recording was performed in the Samplitude 11 DAW system. The following sounds were recorded:

— male speech (a sentence in English),

— acoustic guitar (a piece from J.S. Bach's Bourree e-moll),

— xylophone (a piece from A. Khachaturian's The Sabre Dance),

— trumpet (a piece from H. Purcell's Trumpet Voluntary).

The used sound samples were recorded earlier in the free field of the anechoic chamber and are called the raw signal. The length of the recording was between 10 and 20 s.

4. Post-processing of the recorded material

After the preliminary analysis of the material for further research and tests it was decided to use the signal from the SoundField microphone and the pair of Rode M3 microphones placed in the XY stereophonic system (two microphones with cardioidal characteristics with perpendicular membranes placed as close to each other as possible).

The first stage was the convolution of the raw signals with the measured impulse responses. The raw samples were monophonic so that each sample was convoluted with six impulse responses. The SoundField microphone

measured four impulse responses W, X, Y, Z, where the letters denote particular directions (W — omnidirectional). Two of the responses were taken from the Rode M3 microphones in the XY configuration, where X, Y are the impulse responses of the channels of the system presented in Fig. 1. The convolution in the discrete domain was conducted using Eq. (1), implemented in discrete form (2).

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)\,\mathrm{d}\tau, \tag{1}$$

$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{+\infty} x(k)h(n-k), \tag{2}$$

where x — raw signal, h — impulse response.

In the following stage of processing the conversion was performed from the B-format [6], received from the recording with the SoundField and the convolutions with the impulse responses of the microphone. The generation of the signal with any characteristics of the virtual microphone could be performed using Eq. (3) [6]:

$$V(\boldsymbol{r}) = \frac{2-\kappa}{2}W + \frac{\kappa\sqrt{2}}{4}(r_x X + r_y Y + r_z Z), \tag{3}$$

where W, X, Y, Z — SoundField B-Format signals, κ — the coefficient of the directional characteristics of the virtual microphone, r — versor of the direction of the virtual microphone in the Cartesian coordinates.

To give the value of the versor r more intuitively, in Eq. (4) the components were split into trigonometric function of the horizontal and vertical angle

$$\boldsymbol{r} = [r_x\ r_y\ r_z] = [\cos\theta\cos\varphi\ \ \sin\theta\cos\varphi\ \ \sin\varphi], \tag{4}$$

where θ — horizontal angle of direction of the virtual microphone, φ — vertical angle of direction of the virtual microphone.

The coefficient κ in Eq. (3) is contained in the range $(0 < \kappa < 2)$ and for extreme values the directional characteristics is omnidirectional for $\kappa = 0$, and bi-directional for $\kappa = 2$. For middle values, e.g. $\kappa = 1$, the virtual microphone has cardioidal characteristics and for other vales the characteristics is more omnidirectional or more bi-directional. That feature could be compared to the physical construction of the microphone with variable directional characteristics, where the resulting characteristic depends on the share of the gradient effect and the pressure effect on the acoustic field. The values of gain on the selected directions depending on the spatial angle γ are defined by Eq. (5):

$$g(\gamma, \kappa) = \frac{2-\kappa}{2} + \frac{\kappa}{2}\cos(\gamma), \tag{5}$$

where γ — spatial angle of the maximal gain, g — maximal gain of the signal in given direction.

Considering the above equations and the ITU recommendations [9] regarding the spatial systems of speaker 5.1, the calculations were performed for simulation of the virtual microphones for the system. The 5.0 system, despite the critical opinions of the scientific environment for its poor performance in re-creation of the spatial sounds, is often used by the listeners for the sake of availability of both the listening equipment and the sound material. Figure 3 presents the schematic of the used system of virtual microphones.

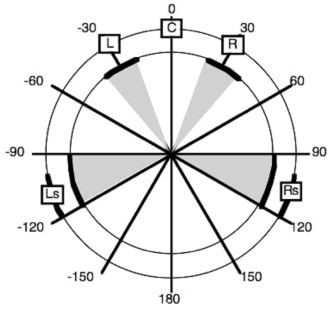

Fig. 3. The schematic of the used system of virtual microphones [8].

From the discussion presented above one could conclude that the generation of the signal of the virtual microphones is limited to simple numerical calculations. All the simulations and computations were performed in the Matlab environment. The values of angles for particular microphones are determined by 5.1 setup [9]. Directional characteristic is cardioidal so $\kappa = 1$ for all virtual microphones.

The last stage of processing of the material was typical edition in the time and amplitude domain. The signals convoluted and recorded were equalised and had their loudness level set to equal level. The particular samples were equalised regarding the average level of root mean square. Also the equal levels of crossfade were set.

5. The listening test with method AB

To perform the assumed verification of the influence of the abovementioned factors the psychoacoustic tests were performed with the use of AB method. The recommendations of International Telecommunication Union (ITU) [10] regarding the psychoacoustic tests were used in possibly best faithful manner, but were slightly modified because of the specifics of the AB method. The test was performed in the listening room of the recording studio in DMV, AGH-UST. The measured level of acoustic background was 38 dBA. The reverberation time of the listening room is sufficiently short and is about 0.4 s. The test speaker system was installed and calibrated according to the recommendations [9].

The test comprised three series:

a) listening on the surround system 5.0,

b) listening on the stereophonic system,

c) headphone listening.

The surround and stereophonic listenings were performed by use of the studio sound monitors Genelec 8030 [5]. For headphone listening the closed reference headphones were used, model Bayerdynamic DT770 PRO.

The group of experts consisted of 8 people with different musical or sound engineering experience. All listeners had at least few years of experience in the range of listening tests. The expert group included two people with more than 6 year experience (EXP6) and two people with more than 4 year experience (EXP4) in working with sound as the sound engineer. Two people were musicians (MUZ) with more than 10 years of education and musical practice and two people were amateurs (AMA) that work with a sound as a hobby. The people tested were 2 women and 6 men, aged 22 to 27. All listeners were characterised by normal ear in terms of audiology. The brackets contain the abbreviations used in Fig. 4.

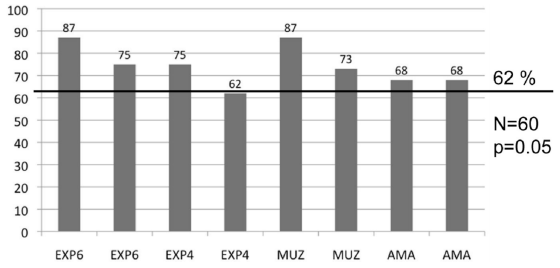

Fig. 4. The results of the listeners. The vertical axis contains the percentage of recognition, and the horizontal axis contains the abbreviations for particular listeners.

In each series each musical sample was tested 5 times, where singular presentation consisted of two sounds that could be the same (the same two recordings or convolutions — identical files) or different (recording then convolution or conversely). The task for the listener was to state the difference between the samples. During each sample the listener also defined how big was the difference between the samples if they stated the difference — this question was not obligatory. So the test was composed of 3 series, with 4 types of sounds and 5 tries (3 series $\times 4$ types $\times 5$ sample pairs $= 60$ samples). The whole test lasted no longer than 30 min including the pauses between the samples.

Since there was no assumption which signal was the reference signal (recorded or convoluted), the listeners assessed how much the signals are mutually distorted according to the following scale (1–5):

1 — the differences are hard to notice,

2 — differences based on the noise and crackles,

3 — small differences based on the quality and sound,

4 — big differences in sound,

5 — very big differences.

6. The test results

The received answers were subjected to the statistic analysis. The results were of the binomial distribution. The null hypothesis assumed stated that the probability of giving the correct answer was 50%, that is it was assumed that the listeners did not hear the differences between the presented samples. The hypothesis was rejected in all the cases with standard level of significance $p = 0.05$. In Figs. 4, 5 and 6 the critical value is given (62% for Fig. 4, 57% for Fig. 5, and 58% for Fig. 6), which does not allow the hypothesis to be accepted. The figures include also the value of N sample ($N = 60$ for Fig. 4, $N = 160$ for Fig. 5, and $N = 120$ for Fig. 6), for which the analysis was performed.

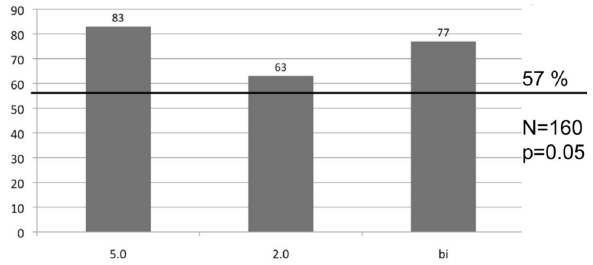

Fig. 5. Results for particular listening systems. The vertical axis contains the percentage of recognition, and the horizontal axis contains the abbreviation denoting particular listening systems (bi denotes the headphone).

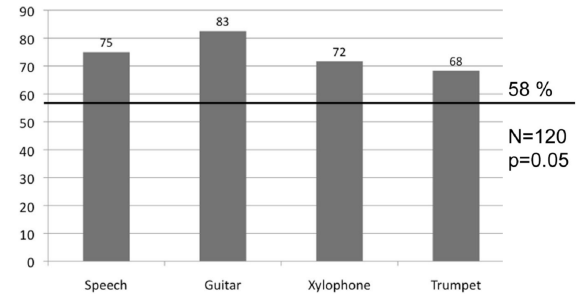

Fig. 6. Results related to the sound samples. The vertical axis contains the percentage of recognition, and the horizontal axis contains the type of the sample.

Figure 4 presents the results of the recognition achieved by particular listeners. Based on the results it could be stated that the differences between the recordings and convolutions were big enough so that all the listeners noticed the differences. The results of the particular listeners were different and high level of recognition (almost 90%) was achieved by only two listeners. Noticeable but minor difference was between the amateurs and experienced listeners.

Figure 5 allows the statement of the subsequent conclusions. The general recognition level for the surround system is much higher than for the stereophonic system,

but the level of 63% is still much higher than the value which allows acceptance of the null hypothesis. The remarkable difference between recognition level between the stereophonic system and the headphone system is very significant since the samples for both the systems were identical. Such difference means that the recognition level depended strongly on the type of the listening system. It is likely that with closed headphone it is possible to hear details that were in some way masked in the speaker system. The levels of the sound were calibrated on all systems to the same level of 80 dB SPLA, by measuring the pink noise of −6 dB FS level. The higher level of recognition for the 5.0 system might be the result of few factors. The SoundField microphone, used for the recording, of which the impulse response was used for convolution, has a very high sensitivity. Additionally the material was presented on 5 speakers, so the noise level increased by 4 dB compared to the stereo signal.

Based on Fig. 6 it could be stated that the type of the sound was not the factor remarkably affecting the recognition level of the differences. Only the sound of guitar achieved a little higher result than other samples. The listeners stated several times that the sound of the guitar was the most pleasant for listening as opposed to the sound of the trumpet, which was stated to be annoying. Probably that subjective factor inspired such differences.

The average assessment of differences stated by the listeners is also important. Regarding the significance of the above results it is justified to assess the listening systems as stated below:

a) surround system: 2.06,

b) stereophonic system: 1.75,

c) headphone system: 2.35.

The detailed definition of the assessment scale is found in Sect. 5. The average rate shows that the listeners noticed the differences between the samples based on the noise and crackles and not on the tone, timbre or other factors, that add to the reception of the space. The rate for the stereophonic system is the lowest, which agrees with the level of recognition of the differences.

7. Conclusions

The performed research opens the field for further research on the use of multi-channel impulse responses in room acoustics. The presented experiments and results allow us stating some conclusions about the method of that type of experiments and about the thesis included in this article, that is the assessment of multi-channel auralizations. More of the listening channels expose the acoustic background (barely audible broadband noise) of the recording compared to the convolutions.

The assumption made before the test, that the differences are barely audible, was false so the method of the test was not chosen properly. The preliminary tests indicated that one should expect smaller differences. Nevertheless conclusions from subjective listening tests are:

— declared by listeners experience in the range of listening tests should be verified (differences between two listeners both EXP6 and MUZ),

— the biggest influence on assessment of auralization is caused by type of multi-channel sound systems (differences between stereophonic and headphone system during the same signals presentation),

— there is influence of testing signal content on assessment of auralization (differences between guitar and other instruments).

The final conclusion could be stated as follows:

— it is necessary to find objective method of assessment based on the listening tests.

Such conclusion is very important because of fact that auralization is more often used for a modelled room acoustic evaluation.

References

[1] A. Farina, R. Ayalon, in: *Proc. 24th AES Conf. on Multichannel Audio, Banff (Canada)*, Audio Engineering Society 2003, Paper 38.

[2] P. Martignon, A. Azzali, D. Cabrera, A. Capra, A. Farina, in: *Proc. 118th AES Convention, Barcelona (Spain)*, Audio Engineering Society 2005, Paper 6485.

[3] F. Wanderley, J. Sousa, in: *Proc. 130th AES Convention, London 2011*, Audio Engineering Society, Paper 8374.

[4] G. Kearney, J. Levison, in: *Proc. 124th AES Convention, Amsterdam 2008*, Audio Engineering Society, Paper 7326.

[5] `http://www.genelec.com/` dated on 7 IX 2011.

[6] A. Farina, R. Glasgal, E. Armelloni, A. Torger, in: *Proc. 19th AES Conf. on Surround Sound, Techniques, Technology and Perception, Schloss Elmau (Germany)*, Audio Engineering Society 2001, Paper 155.

[7] M. Williams, *The Stereophonic Zoom*, Rycote Microphone Windshields Ltd and Human Computer Interface, Gloucestershire (UK) 2002.

[8] K. Hiyamaand, S. Komiyama, in: *Proc. 113th AES Convention, Los Angeles 2002*, Audio Engineering Society, Paper 5674.

[9] I.T.U. The, Radiocommunication Assembly, "Recommendation ITU-R BS.775 Multichannel stereophonic sound system with and without accompanying picture", 2006.

[10] I.T.U. The, Radiocommunication Assembly, "Recommendation ITU-R BS.1284-1 General methods for the subjective assessment of sound quality", 2002.