Journal of the Audio Engineering Society
Vol. 63, No. 6, June 2015 (© 2015)
DOI: http://dx.doi.org/10.17743/jaes.2015.0054

Multichannel Sound Reproduction Quality Improves with Angular Separation of Direct and Reflected Sounds*

PIOTR KLECZKOWSKI, *AES Member*, **ALEKSANDRA KRÓL,** *AES Student Member*, **AND PAWEL MALECKI,** *AES Member*

Department of Mechanics and Vibroacoustics, Audio Engineering Group, AGH University of Science and Technology, 30-059 Krakow, Poland.

Arguments can be put forward for the separation of direct and reflected components of the sound field and reproducing them through appropriate transducers, but there is no definite tested outcome available about that. In this work the perceptual effect of angular separation in commonly used 5.0 and 7.0 multichannel systems was investigated. Four listening experiments were performed involving several schemes of separation and a variety of experimental conditions. The listeners consistently preferred schemes involving separation to schemes without separation.

0 INTRODUCTION

In the introduction to their paper [1] Johnston et al. made a remark about the direct and diffuse components of recorded sound: "In the best of all worlds, such signals would be separate and reproduced via appropriate transducers."

This idea concerns the improvement of spatial audio reproduction and can be implemented in multichannel systems. The direct component of sound is available in most cases as the original room acoustics are seldom captured during recording. The diffuse component is freely available from large libraries of room impulse responses (IRs) and when a particular space needs to be accurately reproduced, its IR can be measured.

In fact, this idea has been applied to a greater or lesser degree, but, to the best knowledge of the authors, no investigation concerning its perceptual effects has been made.

At the implementation level the actual division between direct and diffuse components must be specified. This division is a core operation in a multichannel method of rendering spatial impression of sound called Spatial Impulse Response Rendering (SIRR) [2]. When the acoustics of a room are analyzed its IRs are divided into three consecutive parts: direct sound, early reflections, and late reflections (the reverberation tail). If the IR is divided into only two parts, then early reflections can be included in the first or second part or split between them. In this work the most unambiguous approach to the division was taken: direct sound (DS) as opposed to all reflected sounds (RSs), where the latter comprises both early and late reflections.

Realistic multichannel methods of spatial reproduction are based on the convolutions of dry (or anechoic) sounds with different IRs characterizing one room, reproduced via individual channels. Original sets of IRs (often referred to as multichannel IRs) of existing spaces are obtained from measurements with a variety of microphone systems, either spaced or coincident, or from calculations based on a physical model. With spaced microphones, the procedure may provide a plausible reproduction of acoustic ambience [3, 4], but the procedure is rather perception-based. With coincident microphone setups, like the Ambisonic microphone [5], the effects are closer to an accurate reproduction of a particular acoustic sound field. The sets of IRs obtained with coincident microphone setups are often referred to as Spatial Impulse Responses (SIRs). Various methods of encoding a measured SIR for reproduction through systems with different numbers of loudspeakers have been developed. Advanced examples are presented in [2, 6–9].

A straightforward approach to implement convolution in the rendering of acoustic spaces: $y(t) = s(t)*h(t)$, where $s(t)$

*A part of this work was originally presented at 138th AES Convention, Warsaw, Poland, 2015 May 7–10.

is the anechoic signal and $h(t)$ is the room IR, is to use the complete IR, i.e.,

$$h(t) = h_{dir}(t) + h_{refl}(t) = \text{DS} + \text{RSs}. \quad (1)$$

This is a natural consequence of measuring spatial IRs, where the DS is captured by all capsules or microphones, although with different amplitudes. A recent example of this approach in auralization [10, 11] can be found in [12].

It can be argued that using only RSs in convolution ($h(t) = h_{refl}(t) = \text{RSs}$) is a better choice for all channels delivering ambience. Indeed, there is only one direction from which the DS reaches the listener. If other directions are represented in a multichannel system, then only reflected sounds should arrive from them. Direct sounds reproduced by a number of sound sources may confuse the ear's sense of direction and may bring coloration from interference. Removing DS from IRs has been used in the Ambiophonics system, in its extension supplementing ambience in the reproduction of stereo recordings [13, 14], and in hardware convolvers. This option is often available in convolution reverberation plug-ins; for examples see [15–17]. It is also used in the SIRR method, where non-diffuse (DS plus early reflections) sounds are radiated from selected directions only, while the diffuse sounds are emitted by all loudspeakers [2]. However, removal of the DS is not always appropriate and this is largely dependent upon the actual DS part of the IR, the method and conditions of its recording, and its relation to the actual dry sound of the source [18]. Farina et al. [13] noticed that depriving IRs of the DS (in fact, they removed DS and some early reflections) caused the reproduction of the sound through the Ambisonics rig to be excessively diffuse, resulting in poor localization of the sound arriving from the frontal stage. They also performed a brief informal test employing IRs containing also the direct sound and noticed that the difference was not very evident.

Another related option can be considered, namely using only the DS part in convolution ($h(t) = h_{dir}(t) = \text{DS}$) in frontal channels (abstracting from their actual number and placement for now), i.e., deleting the RSs part from them. Reflections from the sides (especially from directions around +/– 60°) are preferred over reflections from the front, as they produce lower interaural cross correlation and hence increase the apparent source width and spatial impression [19–21]. Another justification is that there is some release from masking when sound sources (the dry sound and its reverberation in this case) are not spatially coincident [22], thus the clarity of reproduction of dry sound should be increased. In the practice of mixing, the center channel is often left dry and reverberation is put in the left and right channels, but this is sometimes debated. The center channel is dry by default in some commercial plug-ins.

Some examples of complete separation of the DS from the ambience in multichannel systems can be found in the literature. Farina and Ayalon [23] recommend that in the Ambiophonics system, the stereo-dipole loudspeakers should provide only the DS and early reflection from the stage, while the surround loudspeakers should provide late reflections. Pulkki and Merimaa [6] designed and tested two modifications of the Ambisonics system, consisting of full separation, for further comparison with their SIRR method. The Ambitail option differed from Ambisonics in reproducing direct sound only from the L or R loudspeaker of the 5.0 system. The Omnitail option also had the DS in the L or R channel but used the same IR (B-format W component) in all five channels. Faller [24] developed a decomposition of the stereo signal into ambient and localized components and proposed the up-mixing of these components respectively to the side and front loudspeakers. Grosse and van de Par [25] proposed a perception-based system for rendering acoustics of a recording room to compensate for the acoustics of the playback room, where the DS was reproduced by a pair of stereo loudspeakers, while the diffuse part was reproduced by a pair of dipole loudspeakers.

In view of the above synopsis, decisions about the inclusion of the DS or RSs are important in developing virtual acoustics systems. They are also commonplace in everyday mixing practice, therefore the authors decided to investigate their perceptual consequences in commonly used multichannel systems, i.e., 5.0 and 7.0. The alternative divisions of sound mentioned earlier in this Introduction, such as including some early reflections with the DS, were not investigated in this work.

The paper explores the perceptual effects of four types of management of the direct and reflected parts of SIRs:

- Using complete IRs in all channels,
- Removing the DS from all but the center channel in a standard 5.0 system,
- Removing ambience from the center channel in a standard 5.0 system,
- Applying both of the above operations, i.e., separating DS from RSs in particular channels.

The investigation was planned so that the results could be generalized. In other terms, some priority was given to external validity, or so-called ecological validity (see, for example, [26]), over internal validity. It was considered necessary that several variations of the above effects should be examined, that the SIRs used should be measured in different rooms, that the listening tests should be performed in different acoustic environments, using different multichannel systems, different number of sound sources, different groups of listeners, and different methods of evaluation. The disadvantage of this ecological validity was that it was not possible to use a common experimental design, so the results of particular experiments cannot be directly compared.

1 GENERAL METHOD

1.1 Division of Impulse Responses

The DS parts in measured IRs are not perfect impulses as they are blurred for various reasons, and they mainly represent the IRs of the measuring equipment used. Therefore, the division of the IR into DS and RSs parts is not straightforward. Kuster in his works [27, 28] used two

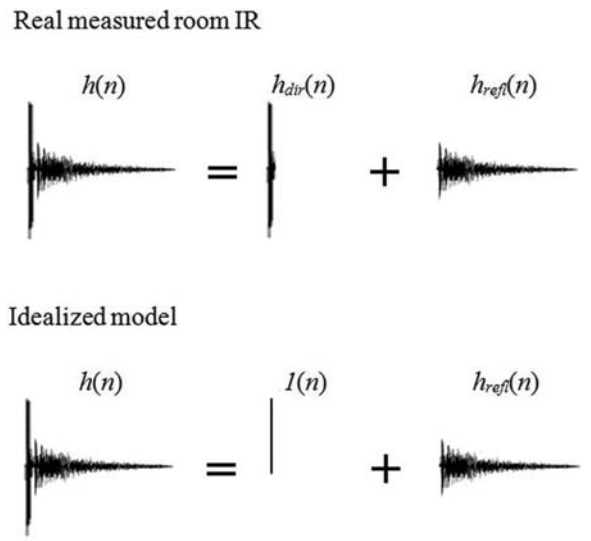

Fig. 1. The division of measured room IR into direct $h_{dir}(n)$ and reflected $h_{refl}(n)$ parts and an idealized model implemented in the paper.

estimations for the duration of the DS: 2.5 ms and 1 ms, the latter value was based on the estimation of loudspeaker and microphone IR. According to the data in [29], the time span of the DS in a control room is around 2 ms. In [30] the point of division was set as the minimum point in the envelope function derived from the first 10 ms of the IR. In [6] the first 4 ms of IRs were chosen as the DS. The authors of the present study measured the IR of their measurement setup in an anechoic chamber (details in Secs. 1.4 and 1.5) and obtained the value of about 3 ms. The startpoint of the DS (after the source to receiver latency) was determined as the first sample in $h(n)$ whose magnitude exceeded the level of –20 dB relative to the maximum magnitude in the entire IR. The endpoint corresponding to a duration of 3 ms should be 143 samples later (at the sampling frequency of 48 kHz, used throughout this work). In order to reduce high frequency spectral leakage to the separated parts, the endpoint was set at the nearest sample with a value close to that of the startpoint. The startpoints and endpoints were set on the W (omnidirectional) signal of the Ambisonic microphone (details in Sec. 1.4). The same points were used for dividing the X, Y, and Z signals.

The work investigated if any perceptual improvement could be obtained by the separation of the DS, therefore it was considered natural to introduce the improvement of the DS itself, by replacing it with an impulse (scaled Kronecker delta) as illustrated in Fig. 1, thus avoiding considerable magnitude and phase distortion introduced by the measured IR.

This was only done in channels assigned to reproduce only the DS. In channels where the complete IR was used: $h(t) = h_{dir}(t) + h_{refl}(t)$, the original measured $h_{dir}(t)$ was used. The impulse replacing the DS was situated at the position of the sample with maximum magnitude in the DS, denoted n_{max}. When implemented in channels assigned to the reproduction of the DS only, instead of performing the convolution: $y(n) = s(n)*h_{Cdir}(n)$, the output signal was obtained as: $y(n) = s(n\text{-}n_{max})$.

1.2 General Paradigm of the Experiments

The following main independent variables were involved in the experiments:

1) The type of management of the DS and RSs components of the sound field reproduced with the 5.0 and 7.0 systems.
2) The auralized space, i.e., the rooms from which the SIRs were collected.
3) The type of listening room.
4) The audio excerpt reproduced.

The basic types of management listed in the Introduction are presented in Fig. 2 in the case of the 5.0 system.

Scheme A is the reference scheme where the input signal to each channel is the result of the convolution: $s(n)*h_X(n)$, where $h_X(n)$ denotes complete IR and X denotes standard designations of the 5.0 system's channels (L, C, R, LS, RS).

Scheme B differs from A in that the center channel is now fed with the DS only. The impulse is used instead of $h_{dir}(n)$ and hence DS $= s(n)$.

Scheme C differs from A in that the DS is deleted from all but the center channel so that the other channels contain RSs only.

Scheme S (for separation) differs from A in that the DS and RSs feed separate channels respectively: the DS feeds the center channel and the RSs feed the front side channels and the rear channels.

The basic paradigm of the experiments consisted of the perceptual comparison of various schemes with the reference scheme within the standard 5.0 system, with only one sound source localized in the center channel. In Experiment 4, this was done within the 7.0 system and with three sound sources.

The spaces from which SIRs were taken are presented in Sec. 1.4 and the listening rooms in Sec. 1.5. In each experiment, several anechoically recorded audio excerpts were used. Each contained music played on a different instrument (Experiments 1–3) or set of instruments (Experiment 4).

1.3 The Devil Is in the Detail

The schemes presented in Fig. 2, except for reference scheme A, are models only and should not really be used in an unbiased perceptual experiment. The definitions of signals given in Fig. 2 in schemes B, C, and S are missing appropriate scaling factors. First, if the DS is replaced with the Kronecker delta then the latter needs appropriate scaling, so that the amplitude of the anechoic (direct) sound in channel C matches that in the reference scheme. Second, the proportion of direct to reverberated sound of scheme A must be kept in all other schemes, otherwise any perceptual evaluations would be strongly biased by a subject's preference towards more or less reverberant reproduction. The need for adjustment of the energy ratio between the

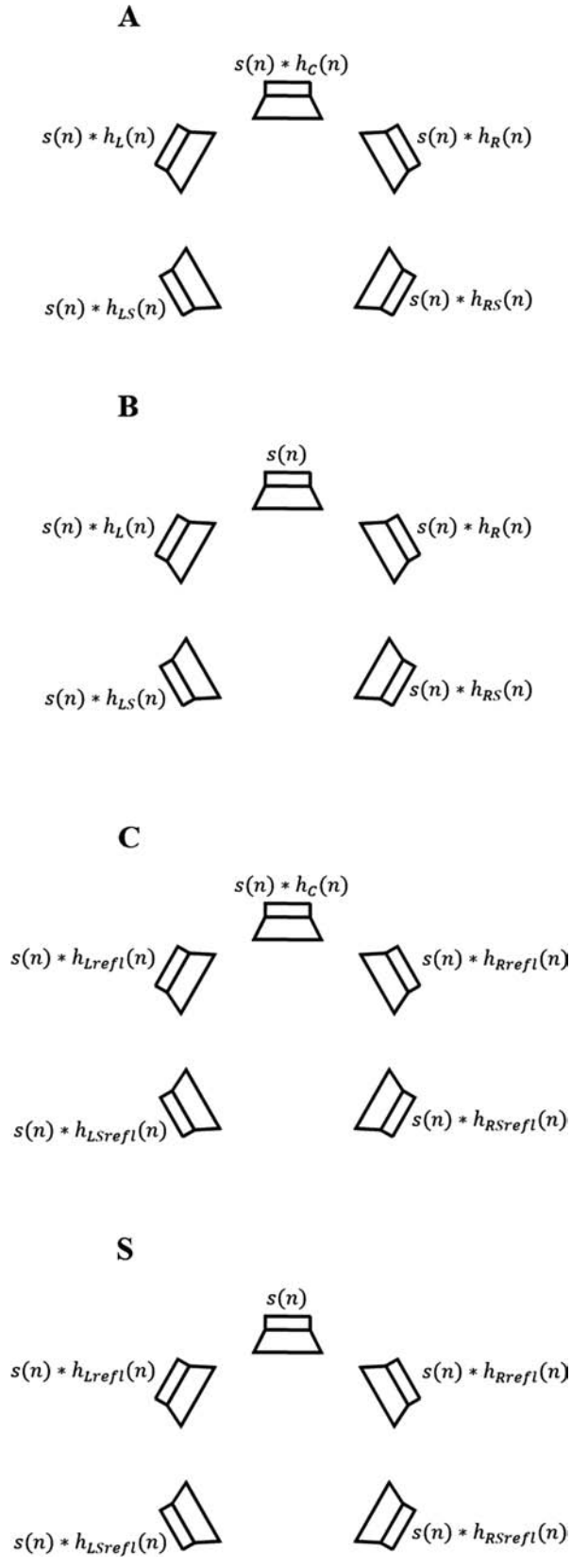

Fig. 2. Basic types of management within the standard 5.0 system.

DS and RSs was noted in [6]. Third, there may be different strategies for maintaining that proportion, each requiring different scaling.

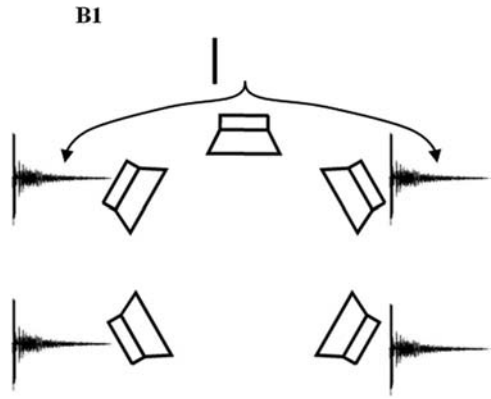

Fig. 3. Schematic illustration of scheme B1. The energy of the RSs part of the IR of the center channel is transferred to both front side channels. Note that the IR of the center channel is now reduced to an impulse.

The scaling factor of the Kronecker delta, i.e., the scaling factor of $s(n)$ was found by setting the RMS amplitude of the scaled $s(n)$ equal to the RMS amplitude of the DS obtained from the convolution

$$k_a \cdot s(n)_{RMS} = (s(n) * h_{Cdir}(n))_{RMS} \tag{2}$$

where k_a is the scaling factor and $h_{Cdir}(n)$ is the DS part of the IR in channel C.

Hence,

$$k_a = \frac{(s(n) * h_{Cdir}(n))_{RMS}}{s(n)_{RMS}}. \tag{3}$$

The above rule was used in all experiments.

1.3.1 Scheme B

In scheme B, according to Eq. (3) and the delay of the impulse mentioned in Sec. 1.1, the output signal to the center speaker was $k_a s(n\text{-}n_{max})$.

The energy of RSs deleted from the center channel (when compared to the state in scheme A) must be compensated for in the other channels. A simple approach (further referred to as B1) is to shift this energy in equal amounts to both front side channels, as shown schematically in Fig. 3. If the RSs signal of the center channel were added to the L and R channels then interference in the low frequencies could occur, as there is some correlation between the IRs of frontal channels, due to the low spatial resolution of the first order Ambisonics microphone. This low spatial resolution can be used as an advantage in a better solution: instead of moving the actual RSs waveform, the approximation of its energy is moved. The RSs energy contained in IRs of L and R channels was increased, simulating the transfer of energy from channel C. Consequently, the arrows in Figs. 3, 4, 6, and 7 denote shifts of energy and not the actual waveforms.

In order to find the scaling factor to increase RSs in the front side channels, the total reflected energy from all three front channels in the reference scheme (A in Fig. 2) is set equal to the scaled total reflected energy from the front side

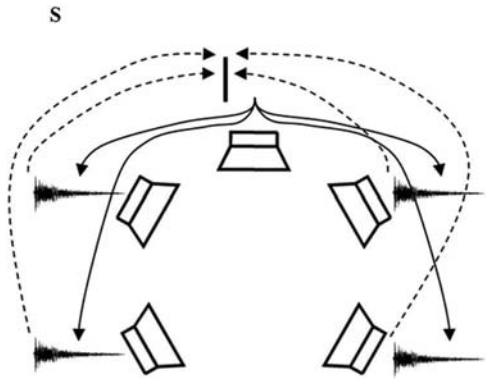

Fig. 4. Schematic illustration of scheme S. The energy of the RSs part of the IR of channel C is transferred to the remaining channels (solid lines). The energy of the DS part of the IRs of front side and rear channels is transferred to channel C (dashed lines). Note that the IR of channel C is now reduced to an impulse, while IRs of the remaining channels are deprived of the DS.

channels (i.e., after the shift of energy), leading to

$$k_{B1} = \frac{\sqrt{(s * h_{Lrefl})_{RMS}{}^2 + (s * h_{Crefl})_{RMS}{}^2 + (s * h_{Rrefl})_{RMS}{}^2}}{\sqrt{(s * h_{Lrefl})_{RMS}{}^2 + (s * h_{Rrefl})_{RMS}{}^2}} \tag{4}$$

where k_{B1} is the scaling factor, h_{Xrefl} is the RSs part of the IR in the respective channel. The notation of discrete time signals $x(n)$ was omitted to streamline the formula. In the above, the assumption is made that the signals involved are uncorrelated. As this assumption is not met at low frequencies, the k_{B1} coefficient is underestimated in that frequency range. However, there was not much low frequency energy in the audio excerpts used in the listening tests.

The actual IR to be used in convolution in channel L in scheme B1 was

$$h_{B1L}(n) = h_{Ldir}(n) + k_{B1} h_{Lrefl}(n) \tag{5}$$

and the output signal in this channel was

$$y_{B1L}(n) = s(n) * (h_{B1L}(n)) \tag{6}$$

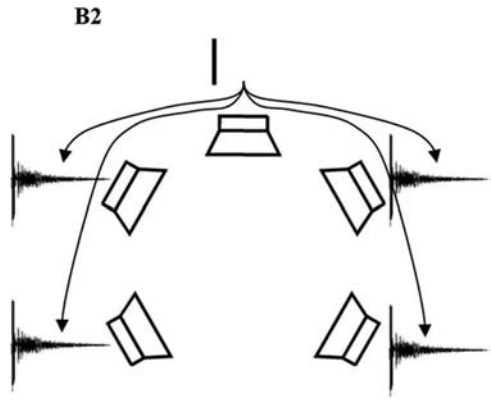

Fig 6 . Schematic illustration of scheme B2.

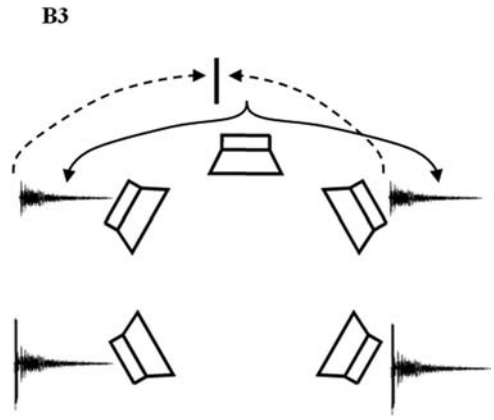

Fig 7. Schematic illustration of scheme B3. The energy of the RSs part of the IR of channel C is transferred to both front side channels (solid lines). The energy of the DS part of the IRs of front side channels is transferred to channel C (dashed lines).

as it was only h_{Lrefl} (RSs part) of the IR that was increased. The same rule was used in channel R. The signals in the rear channels were exactly as defined in panel A of Fig. 2.

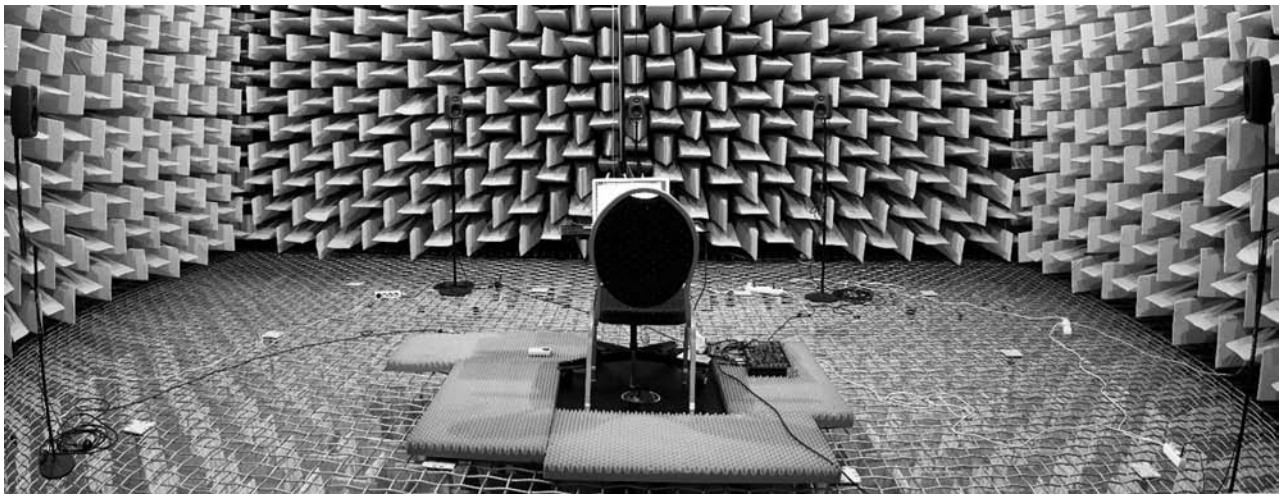

Fig. 5 . The setup for Experiment 1 in the anechoic chamber.

1.3.2 Scheme C

In scheme C the energy of the DS deleted from the front side and rear channels must be compensated for in channel C, so that the total energy of the DS in the reference scheme A is equal to the DS energy in the center channel of scheme C. Then

$$k_C = \frac{\sqrt{\begin{array}{c}(s*h_{Ldir})_{RMS}{}^2+(s*h_{Cdir})_{RMS}{}^2+(s*h_{Rdir})_{RMS}{}^2\\ +(s*h_{LSdir})_{RMS}{}^2+(s*h_{RSdir})_{RMS}{}^2\end{array}}}{(s*h_{Cdir})_{RMS}}. \tag{7}$$

The output signal in channel C was:

$$y_{CC}(n) = s(n)*(k_C h_{Cdir}(n) + h_{Crefl}(n)). \tag{8}$$

The signals in all but the center speaker were exactly as defined in panel C of Fig. 2.

1.3.3 Scheme S

For scheme S, the shifts of energies presented schematically in Fig. 4 were chosen. The RSs energy of the center channel was shifted to the front side and rear channels. The DS energy of the front side and rear channels was shifted to the center channel. The scaling factor for the DS in the center channel was calculated according to Eq. (7) and the actual output signal in this channel was

$$y_{SC}(n) = k_C k_a s(n - n_{\max}) \tag{9}$$

as the DS was replaced by the Kronecker delta.

The scaling factor to compensate for the RSs deleted from the center channel was

$$k_S = \frac{\sqrt{\begin{array}{c}(s*h_{Lrefl})_{RMS}{}^2+(s*h_{Crefl})_{RMS}{}^2+(s*h_{Rrefl})_{RMS}{}^2\\ +(s*h_{RSrefl})_{RMS}{}^2+(s*h_{LSrefl})_{RMS}{}^2\end{array}}}{\sqrt{\begin{array}{c}(s*h_{Lrefl})_{RMS}{}^2+(s*h_{Rrefl})_{RMS}{}^2\\ +(s*h_{LSrefl})_{RMS}{}^2+(s*h_{RSrefl})_{RMS}{}^2\end{array}}} \tag{10}$$

and the output signals to all but the center channel were of the form

$$y_{SX}(n) = s(n)*(k_S h_{Xrefl}(n)). \tag{11}$$

After all test signals have been computed according to the above rules final normalization was performed so that each evaluated audio excerpt in each scheme had the same total RMS level computed from all channels. This was adequate for the normalization of loudness, since samples of the same audio excerpt had nearly the same spectral content.

1.4 Acquisition of Spatial Impulse Responses

The SIRs were taken from the library developed by one of the authors (Malecki) [31]. Three SIRs were used, measured in three Orthodox churches further referred to as Z, T, and H. Room Z was a small wooden church with an RT60 of 1.1 s and an approximate volume of 750 m^3, room T had an RT60 of 2.9 s and a volume of 3,000 m^3, and room H had an RT60 of 4.6 s and a volume of 6,000 m^3. The last church is a regular host of a choir music festival.

The SIRs were measured with the sweep-sine method. EASERA Pro software [32] was employed as well as algorithms developed by Malecki and implemented in Matlab. The ISO3382 [33] standard was followed during the measurements as far as possible, with the exceptions that a first order Ambisonics microphone, the Soundfield ST350, was used, and, instead of an omni-directional source, an active two-way loudspeaker JBL EON 515 was used. The RME Fireface800 served as the AD/DA converter and the microphone preamplifier.

Table 1. Assignment of listening environments to experiments. AC – anechoic chamber, LR – listening room.

EXPERIMENT NO.	LISTENING ENVIRONMENT
1	AC
2	LR
3	LR
4	AC

In all of the churches the sound source was placed in front of the iconostasis, a typical place for a priest or a choir during the concerts. In two rooms, Z and T, the microphone was respectively at about six and eight meters from the source towards the church center, where the best seats for the audience were. In the biggest room (H), the microphone was placed at the rear of the church, about 15 meters from the sound source and about 5 meters from the rear wall, in order to obtain a SIR in a more diffuse field. The tweeter of the loudspeaker and the microphone were both placed at 1.6 m above the floor.

Decoding from the SIRs in the B-format, i.e., the W, X, Y, and Z signals to five IRs to be reproduced in five channels of the 5.0 system was performed according to a procedure described in [13]. It is based on the conversion of the B-format signals to signals from an array of virtual directional microphones. Each virtual microphone points toward one speaker of a 5.0 system, thus the procedure derives the proper impulse response corresponding to the position of this speaker in the multichannel array. Cardioid characteristics of virtual microphones were assumed.

In order to achieve the best S/N ratio, all SIRs were measured using a test signal of 10 s duration. All IRs were truncated at the time of theoretical 90 dB signal decay (calculated from the RT values) plus a fade out of 100 ms.

1.5 Conditions of the Listening Experiments

Two listening environments were used. Their assignment to the experiments is given in Table 1. The anechoic chamber of the AGH University is a cube of internal volume of 465 m^3. The volume between the wedges is 342 m^3, and the accessible floor dimensions are 6.7 m x 7.1 m. It is formed with a steel grid and is placed 0.5 m above wedges covering the real floor. The SPL of the background noise is –2 dB(A). The cut-off frequency is 80 Hz [34].

The listening room has the following dimensions: 3.9 m x 6.7 m x 2.8 m. The shorter walls are covered by thick curtains, placed 1.25 m from them. These walls were the side walls during the listening tests. The room is acoustically treated with three kinds of APAMA acoustic foam absorbers: panels of 5 cm thickness, 40 cm long

wedges, and low frequency cuboids (2 m x 1 m x 0.4 m) behind the curtains as low frequency absorbers. Average RT20 is 0.15 s, calculated from the RT20 at 500 Hz and 1,000 Hz. The room meets most of the ITU-R criteria [35] or has parameters very close to recommended. It meets the criteria for floor area and the ratio of dimensions. The RT is much shorter than the recommended 0.56 s and the noise level is 19 dB(A) and meets the most stringent rating, NC15.

The standard 5.0 system setup [36] was used at +/–120° angles for surround speakers. The radius of the system was 2.5 m in the anechoic chamber and 2 m in the listening room. Active two-way monitors were used with their tweeters positioned at 1.2 m above the floor. The output levels of all monitors were calibrated with the use of pink noise at 80 dB(A) and the Svantek SVAN 959 sound pressure level meter. Calibration accuracy was ±0.25 dB.

The listener's chair was precisely positioned and the listener's head was exactly in the center of the 5.0 setup. In the anechoic chamber, the listener only had a small stiff pad on his lap to be used with a wireless mouse.

In the listening room the video display was placed on the wall behind the central speaker. The listeners had a computer keyboard and a mouse ergonomically placed in front of them on a small angled table with the dimensions of 55 cm x 32 cm.

The model of monitors used in the anechoic chamber was the Genelec 6010A, while in the listening room it was the Genelec 8030A. The audio interface in Experiments 1 and 4 was the RME Fireface 800 and in Experiments 2 and 3 it was the PreSonus FireStudio Lightpipe.

The experiments were run with a custom script written in Matlab, providing a screen and mouse user interface. The audio excerpts were pre-computed and replayed from the computer's hard disk at a peak level of 80 dB(A). Each listener was given written instructions before the experiment.

1.6 Participants

In each test, the listeners represented a diverse group as regards their age and experience in listening tests. Inexperienced listeners were included in order to increase the ecological validity of the research. Among the entire panel of 65 subjects, those aged 20 to 25 dominated; 35% of the subjects were older. The groups in experiments were rather disjoint, but 25% of them took part in more than one experiment, mainly the experienced subjects. Thirty-eight percent of the listeners were students studying for a degree in acoustical engineering or faculty members of the Department of Mechanics and Vibroacoustics at the AGH University. Most of them had some experience in listening tests. None of the listeners reported any hearing deficiencies. According to [37], there is no effect of the listener's audiometric threshold on his performance in listening tasks with test material well above threshold.

Table 2 presents numbers of listeners and proportions of experienced listeners in the particular experiments.

Table 2. Characteristics of experience of listeners in particular experiments.

	EXP. 1	EXP. 2	EXP. 3	EXP. 4
Lack of experience	9	4	5	23
1–3 years experience	4	7	7	9
More than 3 years experience	3	8	3	6
Total	16	19	15	38

1.7 Quantification of Impression and Statistics

In Experiments 1 and 2, pairs of stimuli were evaluated by a pairwise preference test, hence the analysis of the results was based on the test of equal proportions.

In Experiments 3 and 4, individual stimuli were evaluated with a rating scale according to attributes of sound quality. It was found in pilot tests that in most cases the differences from changing the reproduction scheme were clearly observable, also to inexperienced listeners; however, translating them into sensory attributes involving aesthetical preference was more difficult. Both the pilot tests and the earlier experience of the authors indicated that, with this type of cognitive task, listeners would rather use just a few categories or values on a rating scale than a scale with better resolution. A theoretical support for this supposition might be found in psychophysics: when the range of a stimulus can be divided into just a few just noticeable differences, then a subject would naturally choose a rating scale with a corresponding number of levels. Gescheider [38] gives an example of astronomers, who judged the apparent brightness of stars on a scale from 1 to 6.

However, such a choice was in conflict with the assumptions for applying parametric statistical tests. One of them is that the dependent variable is measured on a continuous interval or ratio scale. According to the literature [39], a scale can be considered continuous when it consists of at least 11 points.

The final choice of the scale was made after consulting some listeners, who chose a simple discrete 5-point rating scale. This scale is recommended in the ITU-T. P.800 standard [40] together with the use of ANOVA, with a comment that ANOVA's usual underlying assumptions are sufficiently nearly satisfied.

ANOVA was performed, but the final conclusions were based on nonparametric tests.

2 EXPERIMENT 1

2.1 Details of Design

This experiment compared schemes A and C. The stimuli in particular channels were processed according to Fig. 2A for scheme A and as explained in Sec. 1.3.2 for scheme C.

The experiment was run as a series of two-interval forced choice comparisons (2IFC), where each pair consisted of one audio excerpt reproduced according to schemes A and C. The listener activated a pair of sounds by a software button and could listen to it only once. The break between the intervals in a pair was 100 ms. Next, he was asked the question: "Which version sounded better?" His choice was

Table 3. Results of hypothesis testing in Experiment 1

Hypothesis	Instrument	z	p-value
H_1: $p_C > p_A$	Guitar	−3.586	<0.0002
	Trumpet	−3.825	<0.0001
	Cello	−2.869	0.002

saved by the click of one of two appropriately designated software buttons. Each trial was repeated ten times. Initially this experiment was planned to involve six different pairs of schemes, therefore 60 trials were presented to the listener. The sequence was randomized, as was the sequence within each interval.

This procedure was repeated for three audio excerpts. The first was a phrase from J. S. Bach's "Bourrée" played on the guitar and lasting 9 s, the second was a phrase from H. Purcell's "Trumpet Voluntary" lasting 5 s, both taken from the Bang & Olufsen's *Music for Archimedes* anechoic recording CD [41]. The last was a phrase from W. A. Mozart's aria of Donna Elvira from the opera *Don Giovanni*, played on the cello (6 s), from the anechoic recordings of symphonic music [42]. Each listener evaluated 60 trials x 3 sources = 180 intervals, which took nearly an hour. The experimental setup in the anechoic chamber is shown in Fig. 5.

All audio samples were auralized with the SIRs of room Z.

2.2 Results

During the listening test, the trials containing the pair A and C or the pair C and A were randomly distributed among ten other pairs. On that basis, individual presentations of these pairs were assumed independent.

It would not be correct to combine responses of an individual subject to all three audio excerpts as they were dependent, hence the results were analyzed for each excerpt. The null hypothesis was that listeners did not prefer either of the schemes A or C (H_0: $p_C = p_A$), and the alternative hypothesis was that they preferred C to A (H_1: $p_C > p_A$). The test of equal proportions based on normal statistics was performed for each sample consisting of the results of one audio excerpt, i.e., 16 (listeners) x 10 (repetitions) = 160 results. The required number of subjects was $n \geq 20$. With all three instruments, listeners significantly preferred scheme C to scheme A. The results are given in Table 3.

The inter-excerpt consistency of the results was investigated. The effect of changed experimental condition (the excerpt) on evaluations was tested with the Cochran's test [43], appropriate for dependent trials with numbers of observations in samples n ≥ 20 and binomially distributed results. The test was based on a chi-squared distribution. The null hypothesis (H_0: the changing audio excerpt does not affect the evaluations) was not rejected ($\chi 2 = 1.53$, df = 2, p = 0.46). This indicates consistency between the excerpts and hence further supports the results given in Table 3.

3 EXPERIMENT 2

3.1 Details of Design

This experiment was a detailed study of the general scheme B of Fig. 2. Three options of this scheme were implemented: B1, B2, and B3.

In both B1 and B2 schemes the output signal to the center speaker was $k_a s(n)$.

Option B1 is shown in Fig. 3 and the details were given in Sec. 1.3.2. Option B2 is presented in Fig. 6. B2 differed from B1 in that the RSs energy removed from channel C was shifted to all of the other channels according to

$$k_{B2} = \frac{\sqrt{\begin{array}{c}(s*h_{Lrefl})_{RMS}{}^2 + (s*h_{Crefl})_{RMS}{}^2 + (s*h_{Rrefl})_{RMS}{}^2 \\ +(s*h_{LSrefl})_{RMS}{}^2 + (s*h_{RSrefl})_{RMS}{}^2\end{array}}}{\sqrt{\begin{array}{c}(s*h_{Lrefl})_{RMS}{}^2 + (s*h_{Rrefl})_{RMS}{}^2 \\ +(s*h_{LSrefl})_{RMS}{}^2 + (s*h_{RSrefl})_{RMS}{}^2\end{array}}}. \tag{12}$$

The output signals to all but the center channel were of the form

$$y_{B2X}(n) = s(n) * (h_{Xdir}(n) + k_{B2} h_{Xrefl}(n)). \tag{13}$$

In scheme B3 the DS and RSs were separated in the frontal channels, and in the rear channels they remained mixed as in the reference scheme A. This can be seen in Fig. 7.

The output signal to the center speaker was $k_{B3} k_a s(n)$, where

$$k_{B3} = \frac{\sqrt{(s*h_{Ldir})_{RMS}{}^2 + (s*h_{Cdir})_{RMS}{}^2 + (s*h_{Rdir})_{RMS}{}^2}}{(s*h_{Cdir})_{RMS}} \tag{14}$$

and the output signal to the front side speakers was as given by Eqs. (5) and (6), but with $h_{Ldir}\ (n) = 0$.

The experiment consisted of paired preference tests of audio excerpts reproduced with the following pairs of schemes: A vs. B1, A vs. B2, and A vs. B3.

The method of comparison (2IFC) and other details of the procedure were those used in Experiment 1. Each listener evaluated 3 pairs x 10 repetitions (randomized) and this was repeated for 3 sound sources, altogether 90 trials, which took about half an hour for most listeners, in accordance to the maximum time of 30 to 40 minutes recommended in [39]. The audio excerpts and the SIRs were the same as in Experiment 1.

3.2 Results

At the first stage of analysis, repeated evaluations from individual subjects were converted to one of two categories: a subject preferred Bx (B1, B2 or B3) to A, or a subject did not prefer Bx to A. The responses of a subject were qualified to the first category if they had chosen Bx in at least eight trials out of ten. It was assumed that this proportion corresponded to 80% correct responses to a stimulus in the determination of a psychometric function. The number of trials was lower than usually recommended [38] but it was partly compensated by a higher threshold than the usual 75%.

Table 4. Results of hypothesis testing in Experiment 2, χ^2 statistics at df = 1, * - significant at $p < 0.05$.

Hypothesis	Instrument	χ^2 value	p-value
H_{1B1}: $p_{B1} > p_A$	Guitar	2.083	0.074
	Trumpet	0.364	0.27
	Cello	0.000	0.5
H_{1B2}: $p_{B2} > p_A$	Guitar	2.083	0.074
	Trumpet	2.083	0.074
	Cello	0.125	0.36
H_{1B3}: $p_{B3} > p_A$	Guitar	4.923	0.013*
	Trumpet	3.063	0.04*
	Cello	1.125	0.1

At the second stage, three evaluations for three investigated schemes were performed independently. Consequently, there were three null hypotheses, formulated in an analogous way as in Experiment 1, H_{0B1}: $p_{B1} = p_A$, H_{0B2}: $p_{B2} = p_A$, H_{0B3}: $p_{B3} = p_A$. The alternative hypotheses were H_{1B1}: $p_{B1} > p_A$, H_{1B2}: $p_{B2} > p_A$, H_{1B3}: $p_{B3} > p_A$.

The statistical procedure of Experiment 1 could not be applied because the condition for the size of the test sample ($n \geq 20$) was not met. In each of the nine tests, the actual number was less than 19 participants, as only those qualified in the first stage were included. Therefore, the test of equal proportions based on chi-squared statistics was performed. The summary of the results is presented in Table 4. In all schemes the null hypotheses were not rejected, but there was a tendency to prefer scheme B1 with the guitar, scheme B2 with the guitar and trumpet, and scheme B3 with all instruments at $p \leq 0.1$. The procedure of Experiment 1 was performed for comparison and the results were similar: preferences for B1 with the guitar, for B2 with the guitar and trumpet at $p < 0.01$, and for B3 with all instruments at $p < 0.05$.

At the third stage, inter-instrument consistency was tested as in Experiment 1, this time with the McNemar's test [44], appropriate for dependent trials and small numbers of observations ($n < 20$). This was performed for all three perceptual comparisons: A vs. Bx. The null hypothesis H_0 was the same as in Experiment 1. For all three comparisons, H_0 was not rejected. The respective statistics were: in B1 ($\chi^2 = 2.89$, df = 2, $p = 0.24$), in B2 ($\chi^2 = 1.14$, df = 2, $p = 0.56$), in B3 ($\chi^2 = 0.2$, df = 2, $p = 0.90$), indicating that there was no effect of changing excerpt in any of the evaluated schemes. This is further evidence for the preferences given in Table 4.

4 EXPERIMENT 3

4.1 Details of Design

In this experiment complete separation of the DS from RSs was evaluated according to scheme S (Fig. 4) and details given in Sec. 1.3.3. It was also considered worthwhile to evaluate another condition, where the same separation scheme was applied, but only one IR was used for convolutions in all channels: the omnidirectional component of the B-format SIR. This option is referred to as S1.

Table 5. Attributes and verbal descriptions of anchors in Experiment 3.

	The lowest grade - 1	The highest grade - 5
Localization of sound source	Not defined	Well defined
Naturalness of space	Unnatural	Perfectly natural
How detailed are sound sources?	Muddy	Very detailed
General impression	Unacceptable	Very good

A method of evaluation different in every respect from the one used in the previous two experiments was chosen. Reproductions of audio excerpts were rated, according to four attributes, on a 5-point rating scale. The use of descriptions attached to the points of the scale would be difficult with the chosen attributes and is debatable [39], but descriptions were attached to the poles of the scale in order to provide guidance on the actual perceptual range of the scale. Three of the attributes were chosen according to the effects observed in pilot tests. The attributes and anchors are given in Table 5.

The experiment was a full factorial one with the following independent variables: the reproduction scheme (three), an auralized room (two), and an audio excerpt (four). The stimuli were randomized regarding the reproduction scheme (the effect of interest), but were blocked according to the other two variables. SIRs of rooms Z and T were used. During the first half of the experiment all stimuli associated with room Z were presented and then the stimuli associated with room T. This was considered correct in view of the ability of humans to adapt to the environment they are presently in [39] and because of the considerable effect of reverberation time found in the pilot test. The order of the rooms was not randomized, as perceptual differences were more pronounced with shorter reverberation time, i.e., in room Z, so this helped listeners to adapt to the more difficult task in the second part. The lower level of blocking was due to the audio excerpt. This design was considered superior to full randomization, as it helped listeners to concentrate on perceptual differences between schemes with the other two independent variables fixed and reduced the risk of sequential contraction bias [39].

There were eight blocks altogether. Each block included three schemes of reproduction, A, S, and S1. There was one common user interface screen for each block containing three activation keys denoted X, Y, and Z and an array of four rows, each for one attribute, containing software buttons designated 1 to 5. The listener was free to activate the keys in any order and to repeat any of them at will. The names of the attributes and the anchors were given on screen (in Polish). The assignment of schemes A, S, and S1 to keys X, Y, Z was random.

Two audio excerpts were the same guitar and trumpet phrases as used in Experiments 1 and 2. The third was a female soprano voice singing a 9 s fragment from S. Moniuszko's opera *The Haunted Manor* recorded in an anechoic chamber at the AGH University using the Schoeps

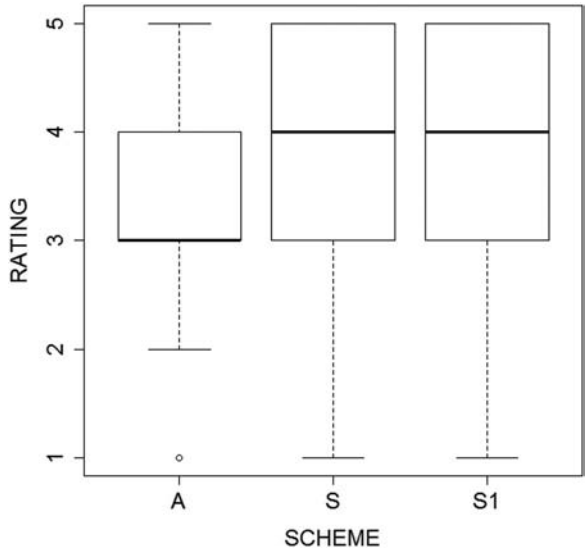

Fig 8. Results of Experiment 3 presented as medians and interquartile ranges of all ratings for schemes A, S, and S1.

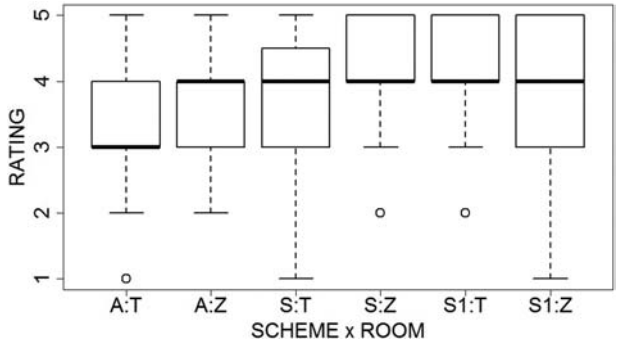

Fig 9. Ratings for schemes A, S and S1 in room T and in room Z.

MK2 microphone with RME UFX ADA and its preamp, and the fourth was a 3 s phrase from Khachaturian's "Sabre Dance" played on the xylophone [41].

4.2 Results

Medians and quartiles are appropriate to demonstrate the results, given in Figs. 8 and 9, because of some outliers. The rating for scheme A averaged over all other factors was 3.36 and the average ratings for schemes S and S1 were equal at 3.98. Average ratings for room Z were: A – 3.51, S – 4.21, S1 – 3.85, and for room T: A – 3.20, S – 3.73, S1 – 4.10.

The assumptions of normality and homogeneity of variance were tested. Normality according to the Shapiro-Wilk's test was satisfied only in 30% of the groups. Homogeneity (Bartlett's test) was satisfied only when computed with larger groups (for two selected conditions with other factors blocked): a scheme and a room, at $p = 0.73$, a scheme and an instrument at $p = 0.76$, a scheme and an attribute at $p = 0.42$. Therefore, it was decided to perform a nonparametric Kruskal-Wallis rank sum test as a primary evaluation and to calculate ANOVA, in view of its high robustness when applied to groups of the same size.

The results of the Kruskal-Wallis test are given in Table 6.

Table 6. Results of Kruskal-Wallis test in Experiment 3.

Source of variance	df	χ^2 value	p-value
Scheme	2	28.137	< 0.0001
Attribute	3	1.262	0.74
Instrument	3	3.258	0.35
Room	1	2.327	0.13
Scheme x Attribute	11	32.424	< 0.001
Scheme x Instrument	11	34.024	< 0.001
Scheme x Room	5	39.260	< 0.0001

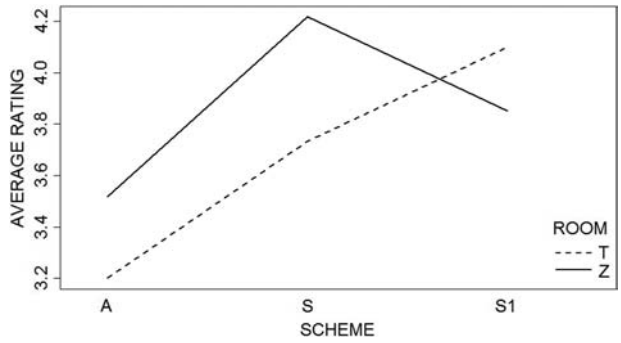

Fig. 10. Interaction scheme x room in Experiment 3.

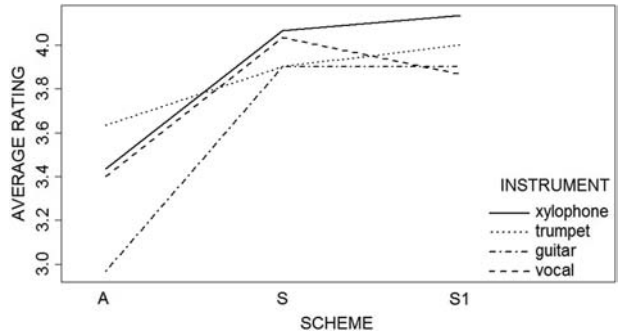

Fig 11. Interaction scheme x instrument in Experiment 3.

The effect of the scheme was highly significant, the others were insignificant. Pairwise interactions were significant.

These results were confirmed by four-way ANOVA. The scheme was highly significant [$F(2, 264) = 14.15$, $p < 10^{-5}$]. Other main effects were insignificant. One significant interaction was found between the scheme and room [$F(2, 264) = 4.12$, $p < 0.05$] (Fig. 10), almost all of it occurring where schemes S and S1 are compared. Another significant interaction was between the scheme, instrument, and attribute [$F(18, 264) = 2.03$, $p < 0.01$]. This can be seen by observing the plot of pairwise interactions shown in Fig. 11, where the largest difference of ratings between scheme A and the two others occurs for the guitar (0.9 point), then for the xylophone and vocal, and the lowest is for the trumpet (0.3 point), and the plot in Fig. 12 where that difference is largest for details (1 point) and lowest for localization (0.4 point).

The HSD Tukey post-hoc analysis confirmed a significant difference between the means of the results of A and S schemes and A and S1 schemes and no difference between the means of S and S1. The differences with their confidence levels are shown in Fig. 13.

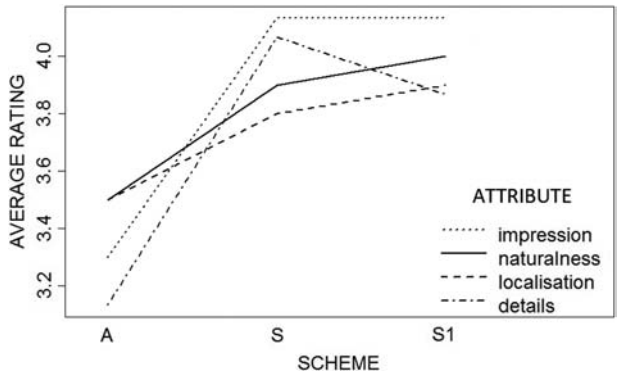

Fig 12. Interaction scheme x attribute in Experiment 3.

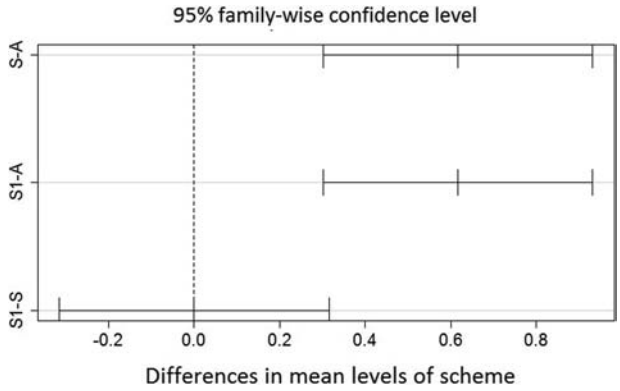

Fig. 13. HSD Tukey post-hoc analysis of differences between the means.

The Tukey multiple comparison of means, performed on all pairs of conditions within factors, demonstrated weak differences either between the instruments or between the attributes themselves and some (but still insignificant) differences between the rooms. The differences of 0.62 points between schemes A and S and schemes A and S1 were both highly significant at $p < 0.0001$ and there was no difference between S and S1.

5 EXPERIMENT 4

5.1 Details of Design

The objective of this experiment was to evaluate the complete separation of the DS from RSs with three separate sources. Each of the dry sound sources was reproduced through one of the three frontal channels of the 7.0 system. An accurate arrangement would require three different SIRs for three directions of sound sources. This was not implemented because frontal reflections were not reproduced anyway, and the complexity of maintaining the constant ratio of direct to reflected energy would increase considerably. The SIRs of one frontal direction were used as in previous experiments and decoding to seven channel IRs was performed according to a similar rule [13].

In the reference scheme referred to as A7, the output signals to frontal channels were of the form

$$y_{A7X}(n) = s_X(n) * h_X(n) \tag{15}$$

where $s_X(n)$, i.e., $s_L(n)$ or $s_C(n)$ or $s_R(n)$ are anechoic signals driving L, C, and R channels. Even with this simplified approach, and in the simple reference scheme, the other channels can be managed in a number of ways. The following scheme was adopted. The surround and back channels (LS, RS, BL, and BR) were positioned at angles of +/–90° and +/–150° respectively. The output signals to LS, RS, BL, and BR were of the form

$$y_{A7XX}(n) = (s_L(n) + s_C(n) + s_R(n)) * (h_{XX}(n)). \tag{16}$$

In the separated scheme, referred to as S7, the output signals to frontal channels were of the form

$$y_{S7X}(n) = k_f k_a s_X(n) \tag{17}$$

where

$$k_f = \frac{\sqrt{\begin{array}{l}(s_L * h_{Ldir})_{RMS}{}^2 + (s_C * h_{Cdir})_{RMS}{}^2 \\ \quad + (s_R * h_{Rdir})_{RMS}{}^2 + (f * h_{LSdir})_{RMS}{}^2 \\ \quad + (f * h_{RSdir})_{RMS}{}^2 + (f * h_{LBdir})_{RMS}{}^2 \\ \quad + (f * h_{RBdir})_{RMS}{}^2\end{array}}}{\sqrt{\begin{array}{l}(s_L * h_{Ldir})_{RMS}{}^2 + (s_C * h_{Cdir})_{RMS}{}^2 \\ \quad + (s_R * h_{Rdir})_{RMS}{}^2\end{array}}} \tag{18}$$

f is given by

$$f = s_L(n) + s_C(n) + s_R(n) \tag{19}$$

and k_a is given by a version of Eq. (3), where $h_{CdirC}(n)$ can be substituted by $h_{Ldir}(n)$ or $h_{Rdir}(n)$ depending on the channel actually calculated.

The output signals to surround and back channels were of the form

$$y_{S7XX}(n) = \big(s_L(n) + s_C(n) + s_R(n)\big) * (k_{S7} h_{XXrefl}(n)) \tag{20}$$

where

$$k_{S7} = \frac{\sqrt{\begin{array}{l}(s_L * h_{Lrefl})_{RMS}{}^2 + (s_C * h_{Crefl})_{RMS}{}^2 \\ \quad + (s_R * h_{Rrefl})_{RMS}{}^2 + (f * h_{LSrefl})_{RMS}{}^2 \\ \quad + (f * h_{RSrefl})_{RMS}{}^2 + (f * h_{LBrefl})_{RMS}{}^2 \\ \quad + (f * h_{RBrefl})_{RMS}{}^2\end{array}}}{\sqrt{\begin{array}{l}(f * h_{LSrefl})_{RMS}{}^2 + (f * h_{RSrefl})_{RMS}{}^2 \\ + (f * h_{LBrefl})_{RMS}{}^2 + (f * h_{RBrefl})_{RMS}{}^2\end{array}}} \tag{21}$$

A similar experimental method to that of Experiment 3 was used. It was different in the numbers of independent variables: three rooms Z, T, and H, presented in this order, three audio excerpts, and two reproduction schemes (A7 and S7). Consequently, the interface screen for each block contained two activation keys denoted by X and Y. Twenty (out of 38) participants of this test were guest students of various engineering programs in Europe—all inexperienced listeners.

The audio excerpts were different than those of the previous experiments. Each of them consisted of three anechoic tracks, reproduced separately by each of the three frontal channels. One was an 8 s phrase with the cello, violin I, and violin II from W. A. Mozart's aria of Donna Elvira from the opera *Don Giovanni*, further referred to as "Mozart," the

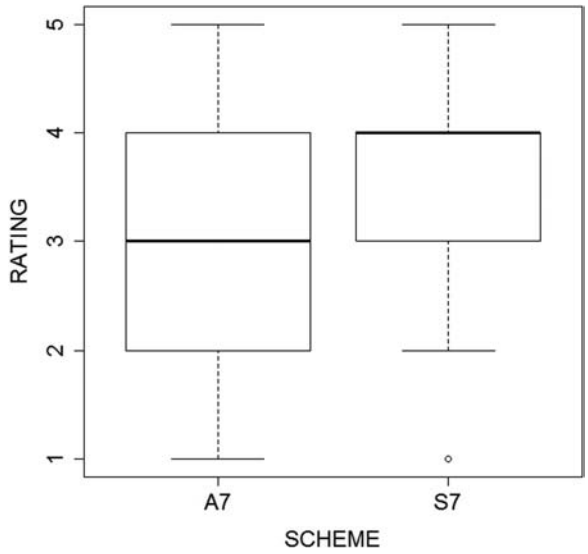

Fig. 14. Results of Experiment 4 presented as medians and interquartile ranges of all ratings for schemes A7 and S7.

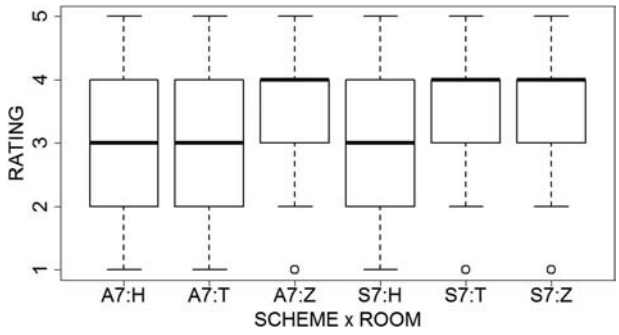

Fig. 15. Ratings for schemes A7 and S7 in rooms H, T and Z.

second was a 4 s phrase with the clarinet, oboe, and bassoon from L. van Beethoven's *Symphony No. 7*, 1st Movement, referred to as "Beethoven," both taken from [42]. The third was a female voice singing a 7 s phrase from Tina Karol's Nochenka song, recorded in the anechoic chamber at the AGH University with the Schoeps MK2 microphone and accompanied by playing guitar and viola samples from an NI Kontakt virtual instrument on a MIDI controller. These tracks were recorded using the multitracking method.

5.2 Results

The same methods as those in Experiment 3 were used. The medians and quartiles are given in Figs. 14 and 15. The rating for scheme A7 averaged over all other factors was 3.25 while it was 3.48 for scheme S7. Average ratings for room Z were: A7 – 3.65, S7 – 3.71; for room T: A7 – 3.22, S7 – 3.70; for room H: A7 – 2.90, S7 – 3.04.

The results of the Kruskal-Wallis test are given in Table 7. All the main effects and all pairwise interactions were highly significant, with the exception of the scheme x attribute interaction that was less significant than the others. The Games-Howell non-parametric test, appropriate for analyzing two populations, demonstrated a high significance

Table 7. Results of Kruskal-Wallis test in Experiment 4.

Source of variance	df	χ^2 value	p-value
Scheme	1	32.398	< 0.0001
Attribute	3	11.581	0.009
Instrument	2	25.356	< 0.0001
Room	2	183.704	< 0.0001
Scheme x Attribute	7	46.973	< 0.0001
Scheme x Instrument	5	105.939	< 0.0001
Scheme x Room	5	235.399	< 0.0001

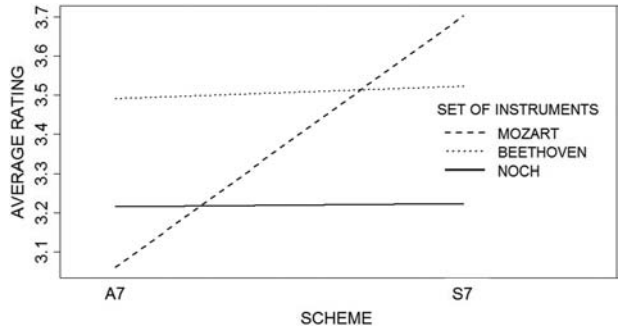

Fig. 16. Interaction scheme x set of instruments in Experiment 4.

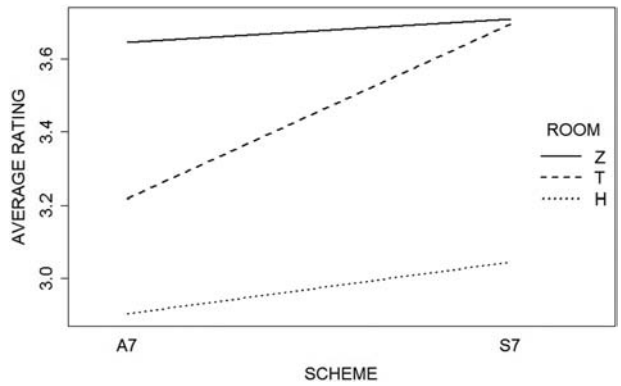

Fig. 17. Interaction scheme x room in Experiment 4.

of the difference in the averaged mean ratings between schemes A7 and S7 at $t = 5.6$, $df = 2721$, $p < 10^{-7}$.

Out of significant two-factor and three-factor interactions, two highly significant, scheme x set of instruments and scheme x room, are shown in Figs. 16 and 17. It can be observed that the ratings of Beethoven and Nochenka were insensitive to the reproduction scheme, while Mozart was quite sensitive yielding 0.7 point of the difference. The Mozart excerpt differed from the other two in that it was played exclusively with string instruments. It can be speculated that these instruments benefit more from separation than those in other excerpts. There was little sensitivity of ratings to the scheme in rooms Z and H, with H gaining considerably lower ratings (about 0.7 point) while there were moderate differences (0.4 point) in room T.

The deviations from assumptions for ANOVA were bigger than in Experiment 3. Normality was found neither in

the small groups nor in the bigger blocked groups. Homogeneity was satisfied in only one blocked group: a scheme and an attribute, at $p = 0.18$.

The effect of the scheme was highly significant [$F(1, 2664) = 38.21$, $p < 10^{-9}$] and so were the effects of the room and of the instrument.

The post-hoc Tukey multiple comparison of means was performed on all pairs of conditions within factors. The difference between A7 and S7 was highly significant ($p = 0$ returned by the software), as were differences in the rooms.

6 DISCUSSION

In Experiment 1 the preference for scheme C was demonstrated for each of the three sound sources. Although the test assumptions were not met (sample size $n = 16$, 20 required) a high significance level obtained leaves enough margin for reduced reliability of the test. Listeners consistently chose reproduction with the DS eliminated from all but the center channel as better sounding.

In Experiment 2 a generally weak preference, dependent on the option, was found. A weak and insignificant tendency for preference was found in scheme B1 where RSs energy removed from channel C was shifted to the front side channels. More, but still insignificant, preference was observed in scheme B2, where RSs energy was shifted from channel C to all other channels. The reason could be either in partial masking of the dry sound in channel C by increased RSs at its sides, leading to the weakening of the positive effect of separation, or by the listeners' preferring the more uniform spatial distribution of RSs to their concentration mainly in the front side channels. Significant preference with two of three instruments occurred in scheme B3, consisting in complete separation of the DS and RSs in the three frontal channels.

More significant preference found in scheme C than schemes Bx indicates that the first type of partial separation is more advantageous. The share of replacing the measured DS with an impulse in the effects found in schemes Bx still remains to be found.

Complete separation between channel C and other channels, investigated in Experiment 3, demonstrated a highly significant preference for both schemes involving separation. Figs. 8 and 9 indicate that the ratings for scheme A were lower but more consistent than for schemes with separation. A surprising result was that scheme S1 implemented with just one omnidirectional IR gained exactly as much average preference to scheme A, as scheme S implemented with the SIR. The straightforward interpretation is that separation itself provides as much improvement to 5.0 sound as spatialization of the IR. Fig. 10 shows the interaction between the scheme and the room. There is a similar increase of preference when scheme A is changed to scheme S in both rooms Z and T, indicating there is no interaction, but there is considerable interaction when scheme S1 is involved. Interactions shown in Fig. 11 indicate that the positive effect of separation (regardless of whether the SIR or the IR was used) is more pronounced with instruments having no sustain phase (guitar, xylophone) than with instruments having it (vocal, trumpet). This might be accounted for by the masking of spatial effects of sound onset by the sustain.

Experiment 4 indicated moderate preference for complete separation between all three frontal channels and surround/back channels. The preference was pronounced in room T of medium RT and weak in rooms Z and H. The weak effect in room H (long RT of 4.6 s) was coincident with distinctively lower average ratings for this room when compared to the others, despite its renown for choir music festivals. A possible explanation is that in low rated artificial space the separation is of little help. The analysis of interactions reveal also that the preference occurred only with the Mozart audio excerpt and not with the other two excerpts (Fig. 16), which is difficult to interpret. These results are not as convincing as those of Experiment 3, but it should be remembered that scheme S7 had its limitation. All RSs were radiated only from surround and back channels, so that RSs did not arrive at all from the frontal semicircle between –90° and +90°. Losing all spatial information from this wide angular range could lower perceptual evaluations. Some realism could also be lost by not implementing three different SIRs (see Sec. 5.1). The weaker effect found in Experiment 4 in relation to Experiment 3 was not due to the listening environment, as it is reflections that could blur the effect and not the anechoic conditions.

The results of both nonparametric and ANOVA tests yielded similar, often highly significant results, thus increasing the chances that ANOVA was valid despite not meeting its assumptions.

In Experiments 1 and 3, a more pronounced preference for separation was found than in Experiments 2 and 4, indicating that the type of listening room did not play a significant role (cf., Table 1).

The values of average preference obtained in Experiments 3 and 4 (0.6 and 0.23 points in favor of separated schemes) can be compared to recent results of general perceptual comparison of stereo and 5.1 surround [45], made with a similar 5-point scale, where the results ranged from 0 to just 0.4 on the 5-point scale in favor of the surround. This demonstrates the perceptual significance of separation.

It is worthwhile comparing the current results with perceptual evaluations presented by Pulkki et al. [6], although conditions differed considerably in some details, e.g., the SIRs were obtained in different ways. In [6] the attribute "naturalness" was graded in several reproduction systems, including the SIRR system, on a continuous 5-point scale. The average score of SIRR reproduced with the 5.0 system was 3.2 points and first-order Ambisonics decoded to 5.0 scored 2.7 points. Their Ambitail modification of Ambisonics (see Introduction), with the DS obtained from a measured SIR (first 4 ms), scored 2.9 points. Their Omnitail scored 2.2 points, distinctively less than Ambitail, while the corresponding S and S1 schemes of the current work received the same average grade.

The main inaccuracy of the method used in this work can probably be attributed to imperfections in the acquisition and reproduction of SIRs. Should this process be more

spatially selective, it could be expected that perceptual results of separation might be even more pronounced. Another inaccuracy could be in the method of maintaining the constant proportion of DS to RSs energy, as the assumption of channel IRs being non-correlated did not quite hold (as mentioned in Sec. 1.3.1).

Although perceptual advantage of separation has been demonstrated, the validity of the results is limited to one or three sound sources and therefore its direct applicability is limited. The authors are planning to repeat Experiment 4 with phantom sound sources between the frontal channels. It is worth mentioning here that the rule used to maintain the constant proportion of DS to RSs energy does not limit the applications in any way, since that proportion can be set by a mixing engineer (in most mixing and recording scenarios) independently of separation. In order to be more useful, the observed preference should be confirmed in systems with more channels. It is also interesting to investigate whether separation is advantageous beyond the sweet spot.

7 CONCLUSIONS

Each of the four experiments conducted demonstrated that in 5.0 and 7.0 reproduction systems the separation of direct and reflected sound sources improve perceived sound quality. This effect was found with one and three sound sources, in several auralized acoustic spaces, in two different listening environments, with various audio excerpts, and with different methods of evaluation.

Partial separation consisting of the removal of the DS from side and rear channels auditioned in anechoic conditions provides a solid advantage. Some advantage can be observed from the removal of the RSs from the center channel in a treated listening room.

Complete separation was perceived as advantageous in both listening environments too. In the case of the 5.0 system, separation was preferred in both auralized rooms with RTs of 1.1 s and 2.9 s, for all four audio excerpts and according to all four attributes evaluated. The advantage was rather selective with the 7.0 system. Although all four attributes indicated the advantage of separation, it was significant only in the auralized room with an RT of 2.9 s but not with RTs of 1.1 s and 4.6 s. The advantage was distinctive with one of the audio excerpts used but not with the two others. In cases where separation did not bring a significant advantage, it brought an insignificant advantage or no effect, but in no case did it deteriorate perception.

The results may be useful as a hint in multichannel rendering of recordings or in further research.

8 ACKNOWLEDGMENT

This research was partly supported by AGH University grant no. 11.11.130.995. The authors are grateful to two anonymous Reviewers for their very helpful comments and to all participants of the listening tests for their patience.

9 REFERENCES

[1] J. D. Johnston, J. Jot, Z. Fejzo, and S. Hastings, "Beyond Coding—Reproduction of Direct and Diffuse Sounds in Multiple Environments," presented at *the 129th Convention of the Audio Engineering Society* (2010 Nov.), convention paper 8314.

[2] J. Merimaa and V. Pulkki, "Spatial Impulse Response Rendering I: Analysis and SynthesisSpatial Impulse Response Rendering I: Analysis and Synthesis," *J. Audio Eng. Soc.*, vol. 53, pp. 1115–1127 (2005 Dec.).

[3] W. Woszczyk, T. Beghin, M. de Francisco, and D. Ko, "Recording Multichannel Sound Within Virtual Acoustics," presented at *the 127th Convention of the Audio Engineering Society* (2009 Oct.), convention paper 7856.

[4] W. Woszczyk, B. Leonard and D. Ko, "Space Builder—An Impulse Response-Based Tool for Immersive 22.2 Channel Ambiance Design," presented at the *AES 40th International Conference: Spatial Audio: Sense the Sound of Space* (2010 Oct.), conference paper 7-1.

[5] M. A. Gerzon, "Periphony: With-Height Sound ReproductionPeriphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, pp. 2–10 (1973 Jan./Feb.).

[6] V. Pulkki and J. Merimaa, "Spatial Impulse Response Rendering II: Reproduction of Diffuse Sound and Listening Tests," *J. Audio Eng. Soc.*, vol. 54, pp. 3–20 (2006 Jan./Feb.).

[7] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *J. Audio Eng. Soc.*, vol. 60, pp. 807–820 (2012 Oct.).

[8] S. Tervo, J. Patynen, A. Kuusinen, and T. Lokki, "Spatial Decomposition Method for Room Impulse Responses," *J. Audio Eng. Soc.*, vol. 61, pp. 17–28 (2013 Jan./Feb.).

[9] D. Scaini and D. Arteaga, "Decoding of Higher Order Ambisonics to Irregular Periphonic Loudspeaker Arrays," presented at the *55th AES Conference on Spatial Audio* (2014 Aug.), conference paper 3-2.

[10] M. Kleiner, B. I. Dalenbäck, and P. Svensson, "Auralization—An Overview," *J. Audio Eng. Soc.*, vol. 41, pp. 861–875 (1993 Nov.).

[11] M. Vorländer, *Auralization, Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality* (Springer Verlag, Berlin-Heidelberg, 2008).

[12] S. Tervo, P. Laukkanen, J. Patynen and T. Lokki, "Preferences of Critical Listening Environments Among Sound Engineers," *J. Audio Eng. Soc.*, vol. 62, pp. 300–314 (2014 May).

[13] A. Farina, R. Glasgal, E. Armelloni, and A. Torger, "Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music," presented at the *AES 19th International Conference: Surround Sound, Techniques, Technology and Perception* (2001 Jun.), conference paper 1875.

[14] R. Glasgal, "Ambiophonics. Achieving Physiological Realism in Music Recording and Reproduction," presented at *the 111th Convention of the Audio Engineering Society* (2001 Nov.), convention paper 5426.

[15] Waves Audio, "IR-1, IR-L and IR-360 User's Guide," http://www.waves.com/1lib/pdf/plugins/ir-convolution-reverb.pdf, accessed 2014 Nov. 10.

[16] Audio Ease, "Altiverb 7 Manual," www.audioease.com/Pages/Altiverb/Altiverb-7-manual.pdf.zip, accessed 2014 Nov. 10.

[17] Christian Knufinke Software, "SIR2 Manual," www.knufinke.de/sir/downloads/SIR2_Manual.pdf, accessed 2014 Jul. 10.

[18] W. Woszczyk, private communication (2014).

[19] Y. Ando, "Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo," *J. Acoust. Soc. Am.*, vol. 62, pp. 1436–1441 (1977 Dec.). http://dx.doi.org/10.1121/1.381661

[20] F. Toole, *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms* (Focal Press, Oxford, 2008).

[21] H. Imamura, A. Marui, T. Kamekawa and M. Nakahara, "Influence of Directional Differences of First Reflections in Small Spaces on Perceived Clarity," presented at *the 134th Convention of the Audio Engineering Society* (2014 Apr.), e-Brief 134.

[22] C. J. Moore, "Controversies and Mysteries in Spatial Hearing," presented at the *AES 6th International Conference: Spatial Sound Reproduction* (1999 Apr.), conference paper 16-023.

[23] A. Farina and R. Ayalon, "Recording Concert Hall Acoustics for Posteriority," presented at the *AES 24th International Conference: Multichannel Audio*, The New Reality (2003 Jun.), conference paper 38.

[24] C. Faller, "Multiple-Loudspeaker Playback of Stereo Signals," *J. Audio Eng. Soc.*, vol. 54, pp. 1051–1064 (2006 Nov.).

[25] J. Grosse and S. van de Par, "Perceptual Optimization of Room-in-Room Reproduction with Spatially Distributed Loudspeakers," *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany (2014 Apr.).

[26] F. Rumsey, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm," *J. Audio Eng. Soc.*, vol. 50, pp. 651–666 (2002 Sep.).

[27] M. Kuster, "Reliability of Estimating the Room Volume from a Single Room Impulse Response," *J. Acoust. Soc. Am.*, vol. 124, pp. 982–993 (2008 Aug.). DOI: http://dx.doi.org/10.1121/1.2940585

[28] M. Kuster, "Multichannel Room Impulse Response Rendering on the Basis of Underdetermined Data," *J. Audio Eng. Soc.*, vol. 57, pp. 403–412 (2009 Jun.).

[29] M. Dunn and D. Protheroe, "Visualization of Early Reflections in Control Rooms," presented at *the 137th Convention of the Audio Engineering Society* (2014 Nov.), convention paper 9157.

[30] F. Menzer, C. Faller, and H. Lissek, "Obtaining Binaural Room Impulse Responses from B-Format Impulse Responses Using Frequency-Dependent Coherence Matching" *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 396–405 (2011 Feb.). http://dx.doi.org/10.1109/TASL.2010.2049410

[31] P. Malecki, *"Evaluation of Objective and Subjective Factors of Highly Reverberant Acoustic Field (in Polish)*," *Doctoral thesis* (AGH University of Science and Technology, Krakow, 2013).

[32] W. Ahnert and W. Schmidt, Fundamentals to Perform Acoustical Measurements (Appendix to EASERA, Berlin 2005).

[33] ISO 3382:1997, "Acoustics – Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters," International Organization for Standardization Geneva, Switzerland.

[34] A. Pilch and T. Kamisiński, "The Effect of Geometrical and Material Modification of Sound Diffusers on Their Acoustic Parameters," *Arch. Acoust.*, vol. 36, pp. 955–966 (2011 Dec.). DOI://dx.doi.org/10.2478/v10168-011-0065-1

[35] ITU-R BS.1116:1997, "Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems," International Telecommunication Union, Geneva, Switzerland, Tech. Rep.

[36] ITU-R BS.775-3: 2012, "Multichannel Stereophonic Sound System with and without Accompanying Picture," International Telecommunication Union, Geneva, Switzerland, Tech. Rep.

[37] P. Kleczkowski and M. Pluta, "Normally Hearing Subjects Have No Advantage of Better Audiograms in Listening Tasks," *Acta Phys. Pol. A*, vol. 121, pp. A120–A125 (2012 Jan.).

[38] G. A. Gescheider, *Psychophysics: The Fundamentals* (Lawrence Erlbaum Associates, Malwah, New Jersey, 1997).

[39] S. Bech, N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application* (John Wiley, Hoboken, New Jersey, 2006). http://dx.doi.org/10.1002/9780470869253

[40] ITU-R P.800: 1996, "Methods for Subjective Determination of Transmission Quality," International Telecommunication Union, Geneva, Switzerland, Tech. Rep.

[41] Spectra s.r.l, "Music for Archimedes," http://www.ramsete.com/Public/Aurora_CD/Anecoic/Archimedes/CD-cover/Archimedes.htm, accessed 2014 Nov. 10.

[42] J Pätynen, V. Pulkki and T. Lokki, "Anechoic Recording System for Symphony Orchestra," *Acta Acustica u. Acustica*, vol. 94, pp. 856–865 (2008 Nov.). http://dx.doi.org/10.3813/AAA.918104

[43] E. Manoukian, *Mathematical Nonparametric Statistics* (Gordon & Breach, London, UK, 1986).

[44] N. E. Breslow and N. E. Day, *Statistical Methods in Cancer Research: Volume 1: Analysis of Case Control Studies* (International Agency for Research in Cancer, Lyon, France, 1980).

[45] M. Schoeffler, S. Conrad and J. Herre, "The Influence of the Single/Multi-Channel-System on the Overall Listening Experience," presented at the *AES 55th International Conference: Spatial Audio* (2014 Aug.), conference paper 5-4.

THE AUTHORS

Piotr Kleczkowski

Aleksandra Krol

Pawel Malecki

Piotr Kleczkowski received a M.Sc. (Tech.) in electronic engineering from the AGH University of Science and Technology in Cracow, Poland, and has been working there since 1981. His first areas of interest were sound synthesis and implementations of first DSP processors. He was a visiting researcher at Cardiff University in the UK in 1985–1986. He received a Ph.D. degree from the AGH University in 1990. Currently he is a professor at the AGH University in the Department of Mechanics and Vibroacoustics and head of the Audio Engineering Group. In the 1990s, he managed his own firm that developed and manufactured quality audio interfaces for PC computers and associated software, which found widespread use in Polish broadcast and recording studios and was nominated for Niptel Media Prize in 1997. In the term of 1990–1994 he served as a member of the City Council in Cracow. His current research interests include psychoacoustics, with a focus on simultaneous perception of sounds, and processing of audio signals, with a focus on time-frequency processing. He wrote a monograph on sound mixing in the time-frequency domain, a textbook on psychoacoustics, and co-authored a textbook on environmental protection. He was the main founder of the field of study in acoustical engineering in the AGH University and now he is its coordinator. He lectures on principles of audio engineering, psychoacoustics, digital audio, and environmental protection. He has been holding several positions in the Polish Section of the AES, and he is a member of the Polish Acoustical Society and of the Polish Society for Electroacoustic Music.

•

Aleksandra Król received a M.Sc. degree in mathematics from the Jagiellonian University in Cracow and a M.Sc. (Tech.) degree with Highest Distinction in acoustical engineering from the AGH University in Cracow, Poland. She currently is a Ph.D. student in the Dept. of Mechanics and Vibroacoustics, AGH University. She is a pianist with a title of musician instrumentalist received from secondary school of music in Rybnik, Poland. Her interests include room acoustics, spatial hearing, listening tests, and applications of mathematics in acoustics, in particular statistical analysis of the results of scientific research. She has worked as an auditor of statistical analysis of the results of research in medicine in HTA Audit company. She is also a practicing sound engineer with two-years experience of work in the Cracow Scena Stu theater and a recording studio.

•

Pawel Malecki was born in Gorlice, Poland, in 1985. He received his M.Sc. (Tech.) in mechanics, and Ph.D. degree in vibroacoustics from the AGH University in Cracow, Poland, 2013. His thesis “Evaluation of Objective and Subjective Factors of Highly Reverberant Acoustic Field” focused on psychoacoustic aspects of choral music perception in the diffused field. His main research tools were auralization techniques combined with ambisonics. He works at the AGH University and as a freelancing sound engineer and acoustic consultant. Dr. Malecki has designed the acoustics of the recording studio in the new AGH University Media Center. He is an officer of the Polish Acoustical Society.