Reproduction of Phantom Sources Improves with Separation of Direct and Reflected Sounds

Piotr KLECZKOWSKI, Aleksandra KRÓL, Paweł MAŁECKI

Department of Mechanics and Vibroacoustics
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Kraków, Poland; e-mail: kleczkow@agh.edu.pl

In virtual acoustics or artificial reverberation, impulse responses can be split so that direct and reflected components of the sound field are reproduced via separate loudspeakers. The authors had investigated the perceptual effect of angular separation of those components in commonly used 5.0 and 7.0 multichannel systems, with one and three sound sources respectively (KLECZKOWSKI *et al.*, 2015, J. Audio Eng. Soc. 63, 428–443). In that work, each of the front channels of the 7.0 system was fed with only one sound source. In this work a similar experiment is reported, but with phantom sound sources between the front loudspeakers. The perceptual advantage of separation was found to be more consistent than in the condition of discrete sound sources. The results were analysed both for pooled listeners and in three groups, according to experience. The advantage of separation was the highest in the group of experienced listeners.

Keywords: spatial audio, multichannel sound reproduction, phantom sources, auralization, ambisonics abbreviations: IR, impulse response; SIR, spatial impulse response; RT, reverberation time; ANOVA, analysis of variance; DS – direct sound, RSs – reflected sounds.

1. Introduction

Convolution of an anechoic sound with room impulse response (IR), the latter obtained from a measurement or calculation based on physical models, is the basic operation in auralization (KLEINER, 1993; VORLANDER, 2008) or virtual acoustics (VORLANDER, 2008; 2014; WOSZCZYK, 2009; 2012), in both loudspeaker and binaural reproduction. When the simulation of acoustic spaces is determined by perception rather than physical data it is referred to as artificial reverberation, but then it is also the convolution with room IRs that brings realism. Indirect means of rendering spaciousness are the key issue in contemporary audio production of pop music, where original room acoustics are seldom captured during recording.

In multichannel spatial sound reproduction, each loudspeaker should be fed with a signal convolved with its own specific IR. 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. With spaced microphones the procedure may provide a plausible reproduction of acoustic ambience (WOSZCZYK, 2009), but

this is rather perception-based, i.e. it is seen as artificial reverberation. With coincident microphone setups, like the Ambisonic microphone (GERZON, 1973) 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 presented in literature, more advanced can be found in (MERIMAA, PULKKI, 2005; PULKKI, MERIMAA, 2006; ZOTTER, FRANK, 2012).

When room IRs are analysed, it is useful to divide them into three consecutive parts (direct sound, early reflections and reverberation tail), but the division into just two parts: direct sound (DS), and all reflected sounds (RSs) is of special interest in this paper.

A straightforward approach to implement convolution in rendering of acoustic spaces: $y(t) = s(t) * h(t)$, where $s(t)$ is the anechoic signal and $h(t)$ is the room IR, is to use the complete IR, i.e.

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

This is a logical consequence of measuring SIRs with both coincident and spaced microphone techniques, where the DS is received by all microphones. A recent example of this approach in auralization can be found in (Tervo *et al.*, 2014).

Using only RSs in convolution ($h(t) = h_{\mathrm{refl}}(t) = $ RSs) may be a better choice for all channels delivering ambience. Indeed, there is only one direction from which the DS from one sound source reaches the listener, therefore only reflected sounds should arrive from other channels of a multichannel system. Coherent radiation of DS by a number of sound sources may confuse the ear's sense of direction. In lower frequencies the vector summation of sinewaves should work as it does in stereo, but in higher frequencies the cues from the head shadow and the pinna function are conflicting. Coloration from comb filtering, more complex than that present in stereo reproduction, may add to this confusion. Removing DS from IRs has been used in the Ambiophonics system, in its extension supplementing ambience in stereo reproduction (Farina *et al.*, 2001; Glasgal 2001) and in hardware convolvers. It is often available in convolution reverberation plugins (Waves Audio, 2014; Audio Ease, 2014; Christian Knufinke Software, 2014). However, in some conditions the removal of DS may not be appropriate (Farina, 2001; Woszczyk, 2014).

The contrasting operation of using only the DS part in convolution ($h(t) = h_{\mathrm{dir}}(t) = $ DS) may be advantageous in front channels. In live musical events reflections are not concentrated in the same directions as direct sounds, and it is known that reflections from the sides are preferred over reflections from the front (Ando, 1977; Toole, 2008; Imamura *et al.*, 2014). Making DS not spatially coincident with RSs can also bring some release from masking (Moore, 1999). The results in (Kleczkowski, Pluta, 2014) indicate that listeners tend to subjectively prefer sound quality of recordings with less spectral overlap from individual musical instruments. In the practice of mixing, the centre channel is often left dry, and reverberation is put in the left and right channels. The centre channel is dry by default in some commercial plug-ins.

In the introduction to their paper Johnston *et al.* (2010) remarked about direct and diffuse components of the recorded sound: "In the best of all worlds, such signals would be separate, and reproduced via appropriate transducers".

Some examples of complete separation of the DS from the ambience in multichannel systems can be found in literature (Newell, Katz, 2006; Farina, Ayalon, 2003; Pulkki, Merimaa, 2006; Faller, 2006; Grosse, van de Par, 2014).

In view of the above, decisions about the inclusion of the DS or RSs are important in developing virtual acoustics systems and are commonplace in everyday mixing practice, and hence the authors have been investigating their perceptual consequences. In (Kleczkowski *et al.*, 2015a) the following effects have been tested:

- removing DS from all but the centre channel (C) of a standard 5.0 system,
- removing ambience from channel C of a standard 5.0 system,
- applying both of the above operations, i.e. feeding channel C with the DS and the remaining channels with RSs.

The results demonstrated a significant preference for the first of the above options.

After careful analysis of the results a next, extensive series of experiments was carried out (Kleczkowski *et al.*, 2015b), including more experimental variables, like some variants of the experimental options listed above, using SIRs measured in different rooms, performing listening tests in different acoustic environments, using both 5.0 and 7.0 systems, and different methods of evaluation. An improved management of DS and RSs was also used.

The results presented in (Kleczkowski *et al.*, 2015b) further confirmed the listeners' preference for separated reproduction of DS and RSs, especially for complete separation (DS in channel C, RSs in other channels) with one sound source. 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.

In that work, complete separation was also examined in the case of three sound sources and the 7.0 system. The preference was rather selective. It was significant only in one of three auralized rooms and in one of three audio excerpts. In that experimental setup each of the three sound sources was reproduced only by one of the three front channels of the 7.0 system, thus azimuthal localisation of sources corresponded to localisations of the respective loudspeakers. This is rather unrealistic in music production, therefore the authors carried out a further experiment, presented in this paper, where, maintaining most of the experimental conditions unchanged, the three sound sources were amplitude panned around the frontal acoustic scene, so that phantom sound sources were created, thus making the experimental conditions closer to those met in real-world mixing. Such an experiment was suggested by one of the reviewers of (Kleczkowski *et al.*, 2015b).

2. Method

The method was similar to that used in (Kleczkowski *et al.*, 2015b), in order to make the conditions as close as possible. Besides panning of sound sources, some changes were made to the experimental procedure, in order to improve its efficiency.

The new elements will be presented in detail, while only the most important points of the common part of the method will be given. The reader is referred to (KLECZKOWSKI *et al.*, 2015b) for more details.

2.1. Signal processing

The experiment consisted of the perceptual comparison of two schemes of management of the DSs and RSs. The schemes are shown in Fig. 1, with symbolic descriptions of signals.
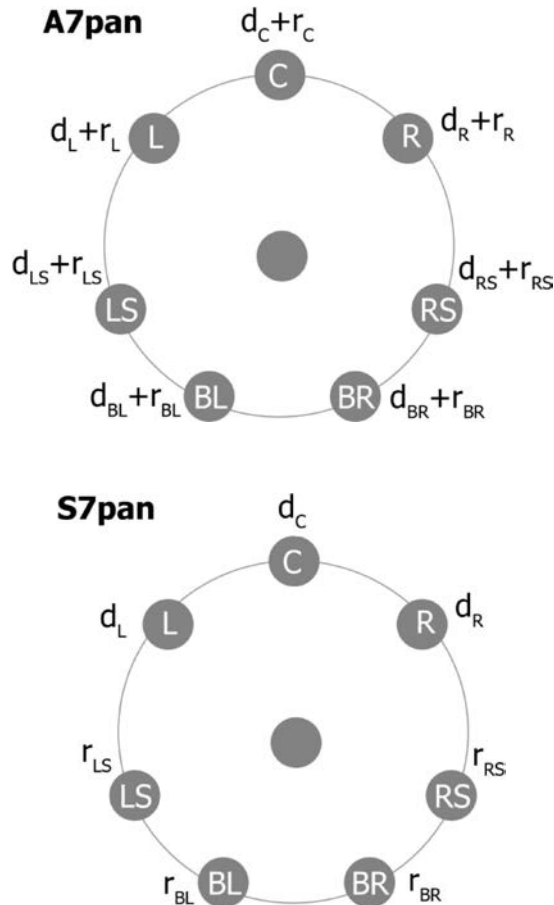

Fig. 1. Schemes of management of direct and reflected sounds compared in the listening experiment. "d_X" denote DS components, and "r_X" denote RSs components in respective loudspeakers of the 7.0 system. The angles in the picture do not represent the actual values. The loudspeaker array conformed to the ITU-R BS. 775 standard.

The scheme denoted "A7pan" (for consistency with abbreviations used in (KLECZKOWSKI *et al.*, 2015b)) is the effect of convolutions of anechoic sounds with full IRs according to (1). In effect, each of the loudspeakers reproduces a sum of DSs and RSs. Particular IRs are specific for each direction. This was the reference scheme in this experiment. In the scheme "S7pan" complete separation of DS from RSs has been imple-

mented so that each loudspeaker reproduces either DSs or RSs.

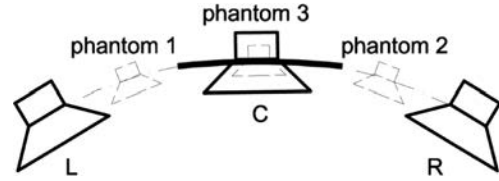The sources were amplitude panned as shown in Fig. 2

Fig. 2. Amplitude panning used in this work. Dashed lines: the phantom sources.

In effect of the panning, each of the loudspeakers was fed by the following proportions of three anechoic signals p_1 (for phantom 1), p_2 and p_3

$$l(n) = 0.89p_1(n) + 0.45p_2(n) + 0.56p_3(n), \quad (2)$$

$$c(n) = 0.61p_3(n), \quad (3)$$

$$r(n) = 0.45p_1(n) + 0.89p_2(n) + 0.56p_3(n), \quad (4)$$

where $l(n)$, $c(n)$ and $r(n)$ are respectively the signals fed to the loudspeakers L, C, R. The signal p_3 was localised in the centre of the sound stage but was apparently widened by panning it also to L and R channels (FRANK, 2013; ZOTTER, FRANK, 2013). The respective coefficients multiplying p_1, p_2 and p_3 were derived from the tangent panning law (BENNETT *et al.*, 1985)

$$\frac{\tan\varphi}{\tan\varphi_0} = \frac{g_L - g_R}{g_L + g_R}, \quad (5)$$

where φ is the phantom source's angle of offset from the centre of the virtual image, φ_0 is the azimuth angle of L and R loudspeakers (30° in our case), g_L and g_R are gains in L and R channels, for respective signals p_i.

Hence

$$g_L = \frac{1}{\left[\left(\frac{1 - \frac{\tan\varphi}{\tan\varphi_0}}{1 - \frac{\tan\varphi}{\tan\varphi_0}}\right)^2 + 1\right]} \quad (6)$$

as from the condition of constant power

$$g_R = \sqrt{1 - g_L^2}. \quad (7)$$

Thus, in A7pan scheme

$$d_L + r_L = l(n) * h_L(n), \quad (8)$$

$$d_C + r_C = c(n) * h_C(n), \quad (9)$$

$$d_R + r_R = r(n) * h_R(n), \quad (10)$$

where $h_X(n)$ are IRs appropriate for respective loudspeakers.

In (KLECZKOWSKI *et al.*, 2015b) the surround and back channels of the 7.0 system (LS, RS, BL, BR) were each fed by the same sum of signals of three sound sources, but that sum was convolved with a different IR, specific for each channel. The same paradigm was used in this experiment, and thus

$$d_{\text{LS}} + r_{\text{LS}} = (l(n) + c(n) + r(n)) * h_{\text{LS}}(n). \qquad (11)$$

The same pattern was followed in the other surround and back channels.

In order to prepare signals for the S7pan scheme, a number of factors presented below must have been involved.

The separation of DS and RSs was performed by dividing the IRs into the respective parts. The first 3 ms of the IR was assumed as representing the DS and the rest as representing the RSs (KLECZKOWSKI *et al.*, 2015b). Moreover, in the loudspeakers reproducing DSs only, the measured DS part of the IR was replaced by the scaled Kronecker delta, to avoid the measurement blur. The authors were aware that this way some useful information like floor reflection was removed too.

In channel L, the scaling coefficient for the Kronecker delta (i.e. the coefficient to multiply the anechoic signal $s(n)$) was obtained as

$$k_{a\text{L}} = \frac{(l(n) * h_{\text{Ldir}}(n))_{\text{RMS}}}{l(n)_{\text{RMS}}}, \qquad (12)$$

where $h_{\text{Ldir}}(n)$ is the DS part of the IR in channel L. The same pattern was followed in channels C and R.

An important experimental condition in the preparation of audio samples in this research is that the proportion of direct to reverberated sound has to be kept constant in all compared options, otherwise any perceptual evaluations would be strongly biased by a subject's preference towards more or less reverberant reproduction. In this work the same approach as used in (KLECZKOWSKI *et al.*, 2015b) was taken: the energy of DSs deleted from surround and back channels (when compared to the state in the scheme A7pan) must be compensated for in the front channels. The compensation was performed by increasing the energy of DSs in respective channels and not by actually moving sound waves from surround and back channels. The same rule was used in removing the energy of RSs from the front channels.

In the following, the symbol $f(n) = l(n) + c(n) + r(n)$ will be used.

Hence, the following signals were used in channel L in the scheme S7pan:

$$d_{\text{L}} = k_f k_{a\text{L}} l(n) \qquad (13)$$

and

$$k_f = \frac{\sqrt{x_{\text{dir}}}}{\sqrt{y_{\text{dir}}}}, \qquad (14)$$

where

$$
\begin{aligned}
x_{\text{dir}} = & (l * h_{\text{Ldir}})^2_{\text{RMS}} + (c * h_{\text{Cdir}})^2_{\text{RMS}} \\
& + (r * h_{\text{Rdir}})^2_{\text{RMS}} + (f * h_{\text{LSdir}})^2_{\text{RMS}} \\
& + (f * h_{\text{RSdir}})^2_{\text{RMS}} + (f * h_{\text{BLdir}})^2_{\text{RMS}} \\
& + (f * h_{\text{BRdir}})^2_{\text{RMS}}, \\
y_{\text{dir}} = & (l * h_{\text{Ldir}})^2_{\text{RMS}} + (c * h_{\text{Cdir}})^2_{\text{RMS}} + (r * h_{\text{Rdir}})^2_{\text{RMS}}
\end{aligned}
$$

and h_{Xdir} is the DS part of the IR in the respective channel. The notation of discrete time signals $x(n)$ was omitted in both formulae above 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_f 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 signals in C and R channels were computed from (13) following the same pattern.

A formula for the computation of signals for surround and back channels is given by way of example for channel LS

$$r_{\text{LS}} = f(n) * k_s h_{\text{LSrefl}}(n) \qquad (15)$$

and

$$k_S = \frac{\sqrt{x_{\text{refl}}}}{\sqrt{y_{\text{refl}}}}, \qquad (16)$$

where

$$
\begin{aligned}
x_{\text{refl}} = & (l * h_{\text{Lrefl}})^2_{\text{RMS}} + (c * h_{\text{Crefl}})^2_{\text{RMS}} \\
& + (r * h_{\text{Rrefl}})^2_{\text{RMS}} + (f * h_{\text{LSrefl}})^2_{\text{RMS}} \\
& + (f * h_{\text{RSrefl}})^2_{\text{RMS}} + (f * h_{\text{BLrefl}})^2_{\text{RMS}} \\
& + (f * h_{\text{BRrefl}})^2_{\text{RMS}}, \\
y_{\text{refl}} = & (f * h_{\text{LSrefl}})^2_{\text{RMS}} + (f * h_{\text{RSrefl}})^2_{\text{RMS}} \\
& + (f * h_{\text{BLrefl}})^2_{\text{RMS}} + (f * h_{\text{BRrefl}})^2_{\text{RMS}}
\end{aligned}
$$

and h_{Xrefl} is the RSs part of the IR in the respective channel.

The SIRs used were measured in two Orthodox churches by one of the authors (PM) (MAŁECKI, 2013) with the sweep sine method and the first order Ambisonics microphone Soundfield ST350 and were the same as used in (KLECZKOWSKI *et al.*, 2015b). The rooms had RT of 1.1 and 2.9 s and will be further denoted as room Z and room T, respectively. Decoding from the SIRs in the B-format, i.e., the W, X, Y, and Z signals (GERZON, 1973) to five IRs to be reproduced in five channels of the 5.0 system was performed according to a procedure described in (FARINA *et al.*, 2001).

2.2. Listening experiment

The listening conditions were the same as those in the relevant part of (KLECZKOWSKI *et al.*, 2015b).

The experiment took place in the anechoic chamber of the AGH University (PILCH, KAMISIŃSKI, 2011). The radius of the 7.0 system (ITU-R, 2012) was 2.5 m, and the surround and back channels were positioned at angles of ±90° and ±150° respectively. Active monitors Genelec 6010A were used. The same audio samples were replayed from the computer's hard disk at a peak level of 80 dB(A). The experiments were run with a custom script written in Matlab, providing a screen and mouse user interface. For most listeners the experiment took from 15 to 25 minutes.

Thirty-three participants took part in the experiment. None of the listeners reported any hearing deficiencies. According to (KLECZKOWSKI, PLUTA, 2012), there is no effect of the listener's audiometric threshold on their performance in listening tasks with test material well above threshold.

The experiment was a full factorial design, with three independent variables: the scheme (A7pan and S7pan), the SIR (rooms Z and T) and three anechoic audio excerpts, each using different musical instruments (KLECZKOWSKI *et al.*, 2015b), giving $2 \times 2 \times 3 = 12$ combinations. The excerpts will be referred to as "instruments" in the rest of this paper. The combinations were blocked according to schemes, so there were six blocks.

Each block included two schemes of reproduction to be perceptually compared: A7pan and S7pan. Reproductions of audio excerpts were rated, according to four attributes, on a rating scale from 0 through 100. Various attributes are in use in the evaluation of audio quality (BECH, ZACHAROV, 2006; KIN, PLASKOTA, 2011). In this work, the authors chose attributes according to the effects observed in their pilot tests. The attribute "localisation of sound source" was meant as the definitions of: angular localisation, its apparent distance and its apparent width. This was explained in a written instruction for the listeners. The descriptions were attached to the poles of the scale in order to provide guidance on the actual perceptual range of the scale. The attributes and anchors are given in Table 1. There was one common user interface screen for each block, containing two activation keys denoted X and Y and an array of four rows, each for one attribute, containing software sliders running from 0 through 100.

Table 1. Attributes and verbal descriptions of the poles of the scales.

Attribute	The lowest grade – 0	The highest grade – 100
Localisation 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

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 to keys X and Y was random. The actual test was preceded by the evaluation of two of the six blocks in the training session.

The order of presentation of the six blocks was randomised for each listener. Thus the experiment was fully randomised.

3. Results

3.1. Pooled results for all listeners

The application of the scale from 0 through 100 fulfils one of the assumptions for applying parametric statistical tests. It states that the dependent variable is measured on a continuous interval or ratio scale (BECH, ZACHAROV, 2006).

The assumptions of normality and homogeneity of variance were tested. Normality according to the Shapiro-Wilk's test, and homogeneity (Bartlett's test) were both satisfied at the significance level of 0.05.

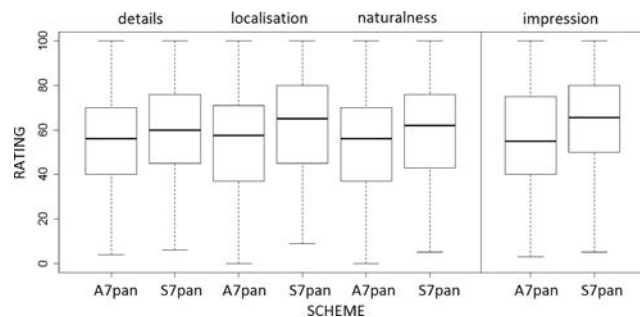Figure 3 presents the box plot of all results, showing quantiles.

Fig. 3. The box plot of the results for individual attributes. The ends of whiskers are 12.5/100 and 87.5/100 quantiles.

Table 2 presents the results of three-way ANOVA of the experimental data. The effect of the scheme was significant in all four attributes, as was the effect of the instruments. The effect of the room was significant in "localisation" and "details" attributes, but not in the other two. None of the interactions of "scheme" with either "instruments" or "room" were significant.

There was some correlation between the attributes, ranging from 0.35 (localisation vs. details) through 0.69 (naturalness vs. impression), with highest values for correlations of impressions with each of the other three analytic attributes. Besides, the attribute "impression" has an integrative character and hence was considered most representative of all four. Figures 4 and 5 present the interactions scheme × instruments and scheme × room respectively, for the attribute "impression".

Table 2. The results of three-way ANOVA of data
in the experiment.

Attribute	Source of variance	Df	F value	p-value
Details	Scheme	1	4.13	0.04
	Instruments	2	3.97	0.02
	Scheme × Instruments	2	0.20	0.81
	Scheme × Room	1	0.18	0.67
Localisation	Scheme	1	9.44	<0.005
	Instruments	2	4.70	<0.006
	Scheme × Instruments	2	1.77	0.17
	Scheme × Room	1	2.26	0.13
Naturalness	Scheme	1	4.86	0.03
	Instruments	2	5.75	<0.005
	Scheme × Instruments	2	1.62	0.20
	Scheme × Room	1	0.10	0.75
Impression	Scheme	1	11.06	<0.001
	Instruments	2	3.17	0.04
	Scheme × Instruments	2	1.34	0.26
	Scheme × Room	1	0.00	1.00

Figures 4 and 5 illustrate, that for the attribute impression there is a consistent perceptual advantage of the separated scheme S7pan over the scheme A7pan without separation for all three audio excerpts, in both rooms.
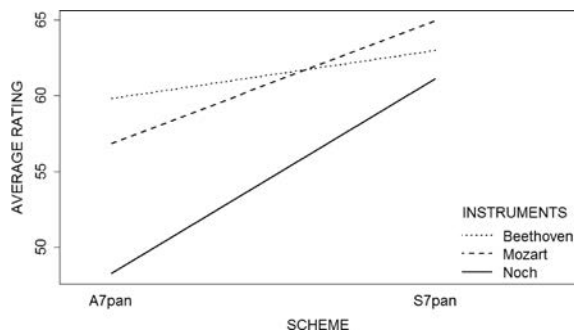

Fig. 4. Interaction scheme × instruments for the attribute "impression".
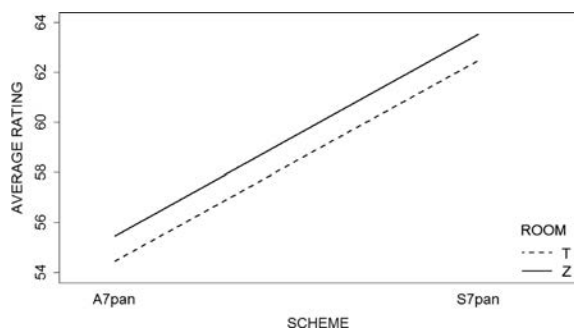

Fig. 5. Interaction scheme × room for the attribute "impression".

The scheme with separation S7pan was perceptually preferred over the A7pan scheme in all of the six conditions examined. The preference was statistically significant for the excerpt "Mozart" in room Z, and for the excerpt "Noch" in both rooms Z and T, as demonstrated by appropriate the t-tests.

Figure 6 presents interactions scheme × attribute, which were insignificant according to the four-way ANOVA, obtained by using attributes as the fourth factor.
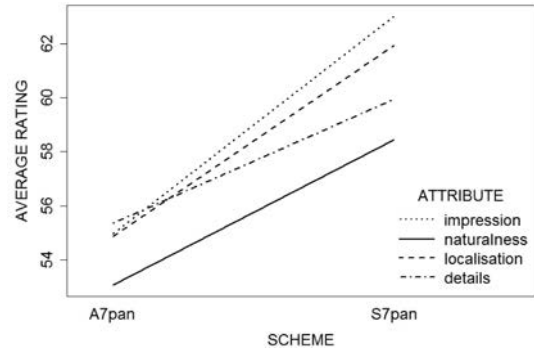

Fig. 6. Interaction scheme × attribute.

3.2. Results in groups of listeners

For the purpose of this analysis, the listeners were divided in three groups. Group H, five participants, were highly experienced listeners. Each of them held at least a M.Sc. degree in acoustics or a related field, had extensive experience in various listening tests, at least some experience in audio mixing, and had musical background. Group M consisted of 19 students of the degree in acoustical engineering, with basic knowledge on acoustics, psychoacoustics and audio mixing and elementary experience in listening tests. Group L were nine listeners without any experience in listening tests and with no relation to the audio field.

Figures 7–10 present the quantiles of results in each of the three groups of listeners for each of the four attributes.
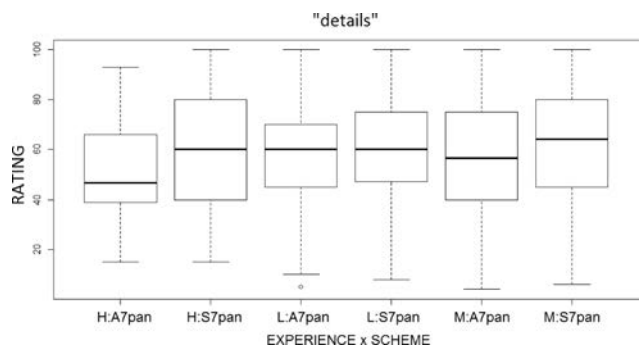

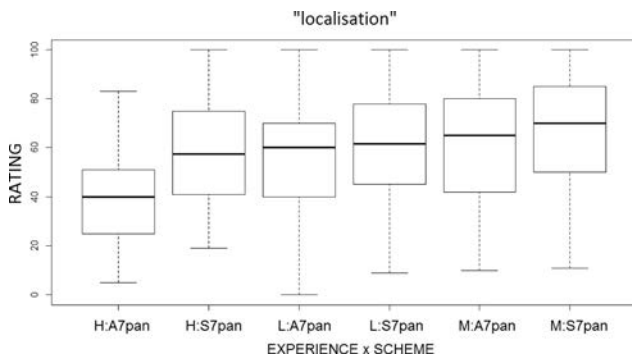Fig. 7. The box plot of the results for the attribute "details". Dots denote outliers.

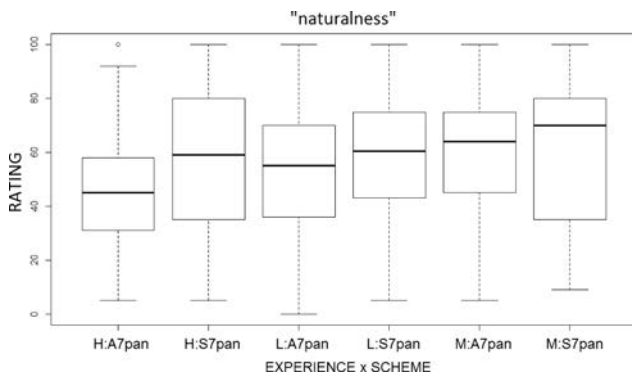Fig. 8. The box plot of the results for the attribute "localisation".

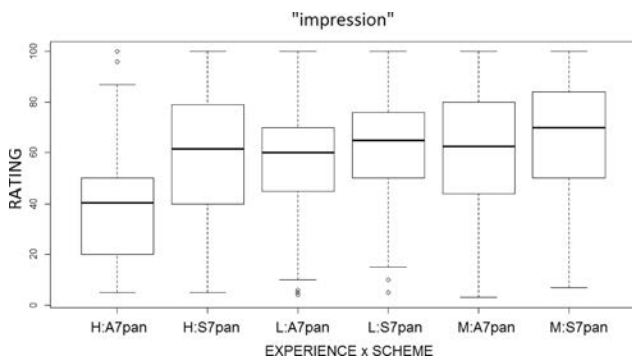Fig. 9. The box plot of the results for the attribute "naturalness". Dots denote outliers.

Fig. 10. The box plot of the results for the attribute "impression". Dots denote outliers.

Highly experienced listeners were more exacting in their evaluations, giving lower scores. This group was most sensitive in detecting the differences between the schemes. The difference in medians of scores in this group was 19 points of the scale, when averaged over all attributes, on the ground of some correlation between them. Higher scores, still with a considerable difference in medians can be observed in the group of inexperienced listeners. Moderately experienced listeners were least sensitive in detecting the differences, but even this group slightly preferred the S7pan scheme. The evaluations in each of the groups

demonstrate a similar spread of the scores, it is only the evaluations of A7pan version in the experienced group M that had noticeably lower spread, and noticeably lower scores, than can be observed in other boxes.

Despite some outliers, the mean values obtained in the above groups follow the same pattern. Table 3 presents mean scores of schemes A7pan and S7pan in particular groups, for individual attributes.

Table 3. Results averaged over the scheme and the group of listeners for each of the four attributes.

Attribute	Group of listeners	A7pan	S7pan
Details	L	59	62
	M	56	59
	H	49	61
Localisation	L	61	66
	M	56	62
	H	42	60
Naturalness	L	62	67
	M	55	60
	H	41	60
Impression	L	61	61
	M	51	57
	H	46	58

The significance of differences between the means was tested with ANOVA. The data met the appropriate assumptions (see Subsec. 3.1). The results of ANOVA are presented in Table 4. There was a significant effect of the experience on evaluations in all three attributes except "details", but no significant interaction between scheme and experience. The last interaction, for the results averaged over all attributes, is shown in Fig. 11.

Table 4. The results of ANOVA of scores in particular groups and for each of the four attributes.

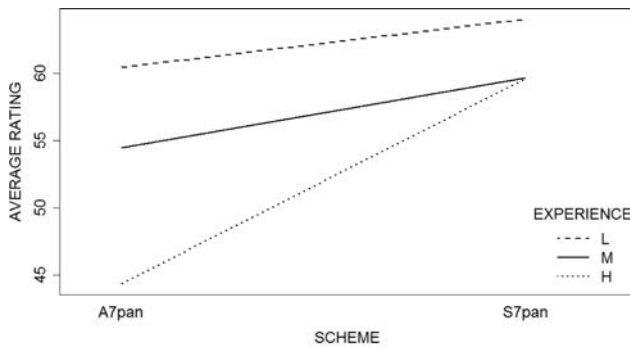| Attribute	Source of variance	Df	F value	p-value
Details	Scheme	1	4.07	0.04
	Experience	2	1.18	0.31
	Scheme $\times$ Experience	2	1.06	0.35
Localisation	Scheme	1	9.43	<0.005
	Experience	2	7.57	<0.001
	Scheme $\times$ Experience	2	2.29	0.10
Naturalness	Scheme	1	4.86	0.03
	Experience	2	3.69	0.03
	Scheme $\times$ Experience	2	0.97	0.38
Impression	Scheme	1	11.35	<0.001
	Experience	2	5.30	<0.006
	Scheme $\times$ Experience	2	1.67	0.19

Fig. 11. Interaction scheme × experience.

The post hoc analysis was also performed, for the results averaged over the attributes. The Tukey test indicated significant differences in mean evaluations between particular groups of participants. It also confirmed, that participants in group H preferred the S7pan scheme at the significance level of less than 0.0001, and participants in group M preferred that scheme at the significance level of 0.01.

4. Discussion

In all four attributes evaluated, the separated scheme S7pan was perceptually preferred over the reference (non-separated) scheme A7pan. Furthermore, the preference was consistent for all six conditions investigated.

Thus the preferences were more consistent than in the version with no panning (KLECZKOWSKI *et al.*, 2015b). In that work, the preference was significant only for one of three rooms investigated (room T). Two of those rooms were included in the current research, the third one was excluded, as separation had little effect in that room. It had a longest RT (4.6 s) and, unlike in rooms Z and T, the measurement of SIR was performed close to the back of the room, far behind the critical distance. It also obtained lowest ratings with both separated and non-separated schemes.

The current and previous (KLECZKOWSKI *et al.*, 2015b) experiments are not consistent in the effect of the audio excerpt on the results. In the previous one, only one out of the three audio excerpts ("Mozart") was found to bring significant perceptual advantage for the separated scheme. In the current work, both "Mozart" and "Beethoven" were scored higher in the S7pan scheme, but the highest advantage was obtained for the "Noch" excerpt. The latter excerpt did not have an effect on preference in the previous work.

The question whether it is amplitude panning or some changes in the experimental procedure that brought these differences remains unanswered. The changed elements of the procedure are given below.

1. The rating scale, 0–100 in this work, and 1–5 in the previous work.

2. Full randomization in this work, and previously a blocked design with only the assignment of software keys to samples randomized.

There seem to be no indications that any of the above differences might bias the results towards either of the compared schemes. They could affect the power of the test. The goal of the change of the rating scale was to meet the assumptions needed to apply parametric statistical tools.

Although the attributes investigated concerned different perceptual constructs, their assessments by listeners were correlated to some extent, and therefore a part of the analysis was performed on results averaged over the attributes.

The obtained average differences of scores between the schemes, in the order of 5%, are not large, but it is natural that the scores for audio of very similar quality, with the poles of the scales defined as in Table 1, tend to concentrate in a fairly narrow range of the scale. However, the difference of medians of the scores in the experienced group (H) was 19 points of the scale.

It should be considered, that the separated scheme had some inherent limitations. 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. An accurate arrangement in this experiment (and in (KLECZKOWSKI *et al.*, 2015b) as well) would require three different SIRs for three directions of sound sources. This could likely improve the quality of both schemes, but presumably the separated scheme could gain more. This option was not yet investigated.

The reason why the listeners' preference for the separated scheme was more pronounced with panning used perhaps can be explained by the increase in apparent source width (ASW) due to panning. In the previous experiment (KLECZKOWSKI *et al.*, 2015b), with each of the instruments reproduced by just one dedicated loudspeaker and no reflections from the frontal semicircle between $-90°$ and $+90°$, the ASW must have been narrower.

There were considerable differences between the results in three groups of participants. The most experienced listeners were considerably more sensitive to the differences between the schemes, which is a result to be expected. Higher sensitivity of the inexperienced group over the moderately experienced group is difficult to explain.

5. Conclusion

In all four attributes evaluated and in all six experimental conditions investigated, the scheme of reproduction with DSs and RSs separated was perceptually preferred over the reference scheme without separation. The preference was particularly high in the

group of experienced listeners. ANOVA demonstrated significance of the scheme in the results.

In contrary to the previous experiment (KLECZ-KOWSKI *et al.*, 2015b), where the preference for separated reproduction was demonstrated in a not quite realistic arrangement of an individual loudspeaker dedicated to the reproduction of just one sound source, phantom sources were used in the current work. The results indicate that separated reproduction can be widely applied in multichannel systems to improve sound quality.

Acknowledgment

References

1. ANDO Y. (1997), *Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo*, J. Acoust. Soc. Am., **62**, 1436–1441.

2. Audio Ease, *Altiverb 7 Manual*, www.audioease.com/Pages/Altiverb/Altiverb-7-manual.pdf.zip, accessed 2014 Nov. 10.

3. BECH S., ZACHAROV N. (2006), *Perceptual audio evaluation: theory, method and application*, John Wiley, Hoboken, New York.

4. BENNETT J.C., KEITH B., FREDERICK O.E. (1985), *A new approach to the assessment of stereophonic sound system performance*, J. Audio Eng. Soc., **33**, 314–321.

5. Christian Knufinke Software, *SIR2 Manual*, www.knufinke.de/sir/downloads/SIR2_Manual.pdf, accessed 2014 Jul. 10.

6. FALLER C. (2006), *Multiple-Loudspeaker Playback of Stereo Signals*, J. Audio Eng. Soc., **54**, 1051–1064.

7. FARINA A., AYALON R. (2003), *Recording Concert Hall Acoustics for Posteriority*, Proceedings of the 24th International Conference: Multichannel Audio, The New Reality, Canada.

8. FARINA A., GLASGAL R., ARMELLONI E., TORGER A. (2001), *Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music*, Proceedings of the 19th AES Conference on Surround Sound, Techniques, Technology and Perception, Schloss Elmau, Germany.

9. FRANK M. (2013), *Source Width of Frontal Phantom Sources: Perception, Measurement, and Modeling*, Archives of Acoustics, **38**, 3, 311–319.

10. GERZON M.A. (1973), *Periphony: With-Height Sound Reproduction*, J. Audio Eng. Soc., **21**, 2–10.

11. GLASGAL R. (2001), *Ambiophonics. Achieving Physiological Realism in Music Recording and Reproduction*, Proceedings of the 111th AES Conv., USA.

12. GROSSE J., VAN DE PAR S. (2014), *Perceptual Optimization of Room-in-Room Reproduction with Spatially Distributed Loudspeakers*, Proceedings of the EAA Joint Symposium on Auralization and Ambisonics, Germany.

13. IMAMURA H., MARUI A., KAMEKAWA T., NAKAHARA M. (2014), *Influence of directional differences of first reflections in small spaces on perceived clarity*, Proceedings of the 134th AES Conv., Germany.

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

15. JOHNSTON J.D., JOT J., FEJZO Z., HASTINGS S. (2010), *Beyond Coding – Reproduction of Direct and Diffuse Sounds in Multiple Environments*, Proceedings of the 129th AES Conv., USA.

16. KIN M.J., PLASKOTA P. (2011), *Comparison of Sound Attributes of multichannel and Mixed-Down Stereo Recordings*, Archives of Acoustics, **36**, 2, 333–345.

17. KLECZKOWSKI P., KRÓL A., MAŁECKI P. (2015a), *On the Management of Direct and Reflected Sounds in the 5.0 Surround Sound Reproduction System*, Acta Physica Polonica A, **128**, 1A, A11–A16.

18. KLECZKOWSKI P., KRÓL A., MAŁECKI P. (2015b), *Multichannel Sound Reproduction Quality Improves with Angular Separation of Direct and Reflected Sounds*, J. Audio Eng. Soc., **63**, 6, 428–443.

19. KLECZKOWSKI P., PLUTA M. (2012), *Normally Hearing Subjects Have no Advantage of Better Audiograms in Listening Tasks*, Acta Phys. Pol. A, **121**, A120–A125.

20. KLECZKOWSKI P., PLUTA M. (2014), *Perceptual Evaluation of the Effect of Threshold in Selective Mixing of Sounds*, Acta Physica Polonica A, **125**, 4-A, A117–A121.

21. KLEINER M., DALENBÄCK B.I., SVENSSON P. (1993), *Auralization – An Overview*, J. Audio Eng. Soc., **41**, 861–875.

22. MALECKI P. (2013), *Evaluation of objective and subjective factors of highly reverberant acoustic field* [in Polish], Doctoral thesis, AGH University of Science and Technology, Krakow.

23. MERIMAA J., PULKKI V. (2005), *Spatial Impulse Response Rendering I: Analysis and Synthesis*, J. Audio Eng. Soc., **53**, 1115–1127.

24. MOORE B.C.J. (1999), *Controversies and mysteries in spatial hearing*, Proceedings of the 16th AES International Conference on spatial sound reproduction, Germany.

25. NEWELL P.M., KATZ S.M. (2006), *Discrete Layered Sound*, Proc. Ist. Acoust., **28**, 8, 169–178.

26. PILCH A., KAMISIŃSKI T. (2011), *The effect of geometrical and material modification of sound diffusers on their acoustic parameters*, Archives of Acoustics, **36**, 1, 955–966.

27. PULKKI V., MERIMAA J. (2006), *Spatial Impulse Response Rendering II: Reproduction of Diffuse Sound and Listening Tests*, J. Audio Eng. Soc., **54**, 3–20.

28. TERVO S., LAUKKANEN P., PATYNEN J., LOKKI T. (2014), *Preferences of Critical Listening Environments Among Sound Engineers*, J. Audio Eng. Soc., **62**, 300–314.

29. TOOLE F. (2008), *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*, Focal Press, Oxford.

30. VORLÄNDER M. (2008a), *Auralization, Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer Verlag, Berlin-Heidelberg.

31. VORLÄNDER M. (2008b), *Virtual acoustics: opportunities and limits of spatial sound reproduction*, Archives of Acoustics, **33**, 4, 413–422.

32. VORLÄNDER M. (2014), *Virtual acoustics*, Archives of Acoustics, **39**, 3, 307–318.

33. 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.

34. WOSZCZYK W. (2014), private communication.

35. WOSZCZYK W., BEGHIN T., DE FRANCISCO M., KO D. (2009), *Recording Multichannel Sound Within Virtual Acoustics*, Proceedings of the 127th AES Conv., USA.

36. WOSZCZYK W., KO D., LEONARD B. (2012), *Virtual Acoustics at the Service of Music Performance and Recording*, Archives of Acoustics, **37**, 1, 109–113.

37. ZOTTER F., FRANK M. (2013), *Efficient Phantom Source Widening*, Archives of Acoustics, **38**, 1, 27–37.

38. ZOTTER F., FRANK M. (2012), *All-Round Ambisonic Panning and Decoding*, J. Audio Eng. Soc., **60**, 807–820.