Audio Engineering Society
Conference Paper
Presented at the 21st Conference
2002 June 1–3 St. Petersburg, Russia

Some Rules and Methods for Creation of Surround Sound

A. Czyzewski, A. Kornacki, and P. Odya

Sound & Vision Engineering Department, Gdansk University of Technology, 80-952 Gdansk, Poland

kido@sound.eti.pg.gda.pl

ABSTRACT

The problem of selection of an adequate surround sound life recording and reproduction methods is still open. Alternative methods of organizing this process are discussed in the paper. Some experimental recording sessions employing the 5.1 format were made with the use of various miking techniques and the convolution-based multichannel audio processing algorithm. The results were submitted to some subjective assessments and then compared. Conclusions resulting from performed experiments are derived and discussed.

INTRODUCTION

Currently there is no clear answer as how to record sound for multichannel systems. Producers use various microphone systems, often designed by themselves, like double ORTF, INA and many others [5] [9] [10]. The method discussed in the paper is completely different. It uses the multi-convolution process for the creation of surround sound.

The convolution technique is well known in audio acoustics since many years. The most popular application of the convolution process is to convolve the input signal with the impulse response of a system. In this way, it is possible to simulate the room acoustics. The acoustical signal, usually recorded anechoically, is convolved with the impulse response of a given room. The impulse response could be recorded in a real room or simulated by a computer software modeling acoustical parameters of rooms. The obtained signal should almost perfectly simulate sound in the real or in the virtual room. Consequently, it should be possible to recreate the sound image of a given room in an electroacoustic surround system using some registered impulse responses. Meanwhile, with this approach there should be possible to produce sound in a concrete surround format, e.g. 5.1. In the proposed method referring to this case, five impulse responses are used, thus five signal versions are obtained. Each one is directly associated with a concrete speaker of the surround sound system.

The main difference between the proposed method and systems described in some papers [3] [6] is the direct relation of microphone technique to the common 5.1 system architecture or loudspeaker arrangement in this system. Earlier methods need additional loudspeakers and they are more computationally complex. For example, the so-called Ambiophonic system requires stereo-dipole loudspeaker pair and cross-talk cancellation. The only problem with the proposed method is the influence of room acoustics in which the sound is reproduced. However, as was shown in subjective tests, this problem pertaining also other sound recording/reproduction systems could be minimized in practice.

The main goal of this project was to find the most suitable arrangement of microphones for recording of the impulse responses for the surround sound reproduced by the 5.1 system. Various techniques of mixing signals were tested to achieve this goal. In contrast to most surround sound recording systems the directional microphones were employed in the proposed solution. In order to optimize the result a number of listening tests were conducted. Their results and conclusions derived on this basis are included in the paper.

EXPERIMENTAL SETUP

The impulse response must be recorded in the real room using very short and very loud excitation. The firecracker explosions were used in the project, sound volume of which reached about 110 dB. All the impulse responses were recorded on the hard disc recorder Fostex with sampling frequency 48 kHz and 20-bit resolution.

The recordings were made with five identical microphones (AKG C4000B) placed according to a few different arrangements described later. Mentioned microphones have selectable polar patterns: omnidirectional, cardioid and hypercardiod, but only the two first were used during the recordings. All used audio cables had similar length to eliminate possible phase shifts.

In the first stage of the recordings, some preliminary tests were performed to ensure whether all channels have equal gains and delays. In order to do that, all microphones were placed close to themselves in the middle of the recording room and pink noise of a high level was generated from the loudspeaker. In this way, gains at the console were adjusted, thus signals in all channels had equal level. Then firecracker explosion was used to check whether all channels had identical delays. Only small differences were observed, possibly caused by size of the microphones and related problems with placing them in precisely one place. After this simplified "calibration", no changes in the settings of the console were made. In this way, in further recordings all impulse responses had different audio levels and time delays that were caused by various distances between microphones and place of explosion.

To add point of reference for subjective tests, two "traditional" multichannel recordings were made. In this case, sound was played through a loudspeaker (Yamaha MSP5) positioned at the place for the performer. Sound excerpts were identical with those used in the convolution process (described later). Two different microphone techniques were used: well-known Double ORTF and experimental one. Double ORTF (Fig. 1) was chosen due to the best assessments it ensures in many different kinds of tests [7]. The experimental system used five microphones (AKG C4000B) placed like in Fig. 2.

Major headings (see above) are in capital letters (upper case), Helvetica font, 8 point size, bold style, left justified. All headings are preceded by a blank line. Body text, like this, is Times Roman font, 8 point size, plain style. It should be fully justified to be flush left and flush right.

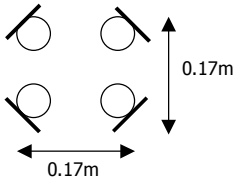


Fig. 1. Double ORTF technique lay-out

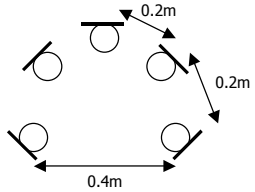


Fig. 2. Experimental microphone technique lay-out

Recorded sound excerpts were then copied into a computer equipped with eight channel sound card. Because of limitations of this soundcard, files were converted to the 16-bit resolution. Outputs of the card were connected to the surround sound amplifier. The listening room (Fig. 3) was equipped with the Paradigm Studio loudspeaker set. The reverberation time of the listening room was equal to about 0.2s.

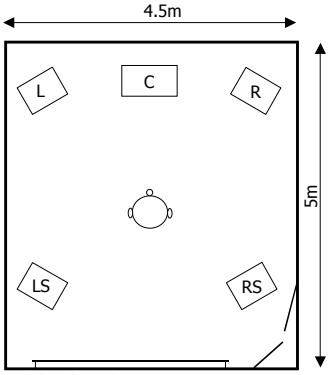


Fig. 3. Lay-out of the listening room

Test sound files were created in the CoolEdit Pro 1.2 software using the plugin "Convolve with clipboard" from the Aurora set of plugins [4]. Mentioned plugin uses "select-save" algorithm, explained in the Fig. 4. All calculations we done with floating-point math that produces a very high sound quality, with no comparison to sound achievable with fixed-point math (e.g. convolution algorithms implemented in the Mathematica system). The Aurora convolution plugin allows for the preservation of information included in the impulse responses such as levels and time delays. Additionally, it was possible to use batch processing that accelerated and made easier all equations.

In experiments were used two "dry" (anechoic) signals: singing male voice and guitar play. These two recordings produced the most diversified convolved sound excerpts. Some preliminary tests showed that e.g. mouth organ recording produces sound not distinctive enough for these tests.

PRELIMINARY EXPERIMENTS

Preliminary recordings were made in the Auditory Hall No. 1 at the Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology. The volume of the Auditory Hall is about 1000m^3. The view of the hall with marked places where microphones were positioned during recordings is presented

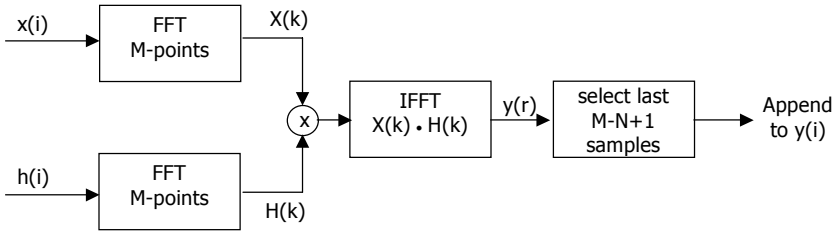


Fig. 4." Select-save" algorithm [2]

in Fig. 5a. Its time reverberation characteristics (Tp) is presented in Fig. 5b.

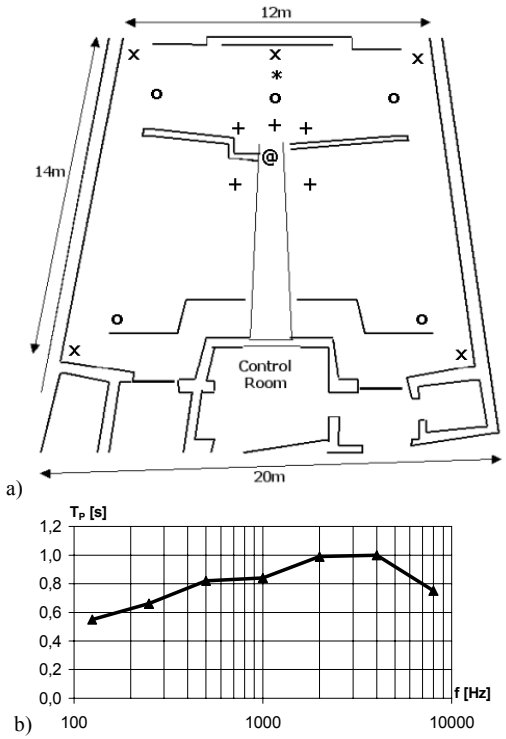


Fig. 6. Layout of the Auditory Hall (a) and its reverberation time characteristics (b)
x - microphones placed at the corners of the room, about 10cm from walls
o - microphones placed at the corners of the room, about 180cm from walls
+ -microphones placed at the center of the room, distances between microphones similar to the distances between loudspeakers in the listening room
* - loudspeaker position
@- experimental and Double ORTF technique

Material

In the Auditory Hall recordings were made using the following microphone arrangements:
-microphones placed at the corners of the room positioned, about 10cm from walls
-microphones placed at the corners of the room, about 180cm from walls
-microphones placed at the center of the room, distances between microphones similar to the distances between loudspeakers in the listening room
-microphones placed at the center of the room , arrangement similar to a small home theater system (the experimental technique)
-microphones located according to the Double ORTF technique

Furthermore, two polar patterns of microphones were used: omnidirectional and cardioid. In the latter case, the main axis of the microphones was directed toward the walls or toward the center of the room. The firecracker explosions were caused also in two different places: at the center of each system and near the center microphone.

Below all recordings made are listed with a short description:
A- arrangement x, omnis, the shot from the central microphone
B- arrangement x, omnis, the shot from the center of the arrangement
C- arrangement x, cadioids, main axis aimed at the center, the shot from the central microphone
D- arrangement x, cardioids, main axis directed to the center, the shot from the center of the arrangement
E- arrangement x, cardioids, main axis directed to walls, the shot from the central microphone
F- arrangement x, cardioids, main axis directed to walls, the shot from the center of the arrangement
G- arrangement o, omnis, the shot from the central microphone
H- arrangement o, omnis, the shot from the center of the arrangement
I- arrangement o, cadioids, main axis directed to the center, the shot from the central microphone
J- arrangement o, cardioids, main axis directed to the center, the shot from the center of the arrangement
K- arrangement o, cardioids, main axis directed to walls, the shot from the central microphone
L- arrangement o, cardioids, main axis directed to walls, the shot from the center of the arrangement
M- arrangement +, omnis, the shot from the central microphone
N- arrangement +, omnis, the shot from the center of the arrangement
O- arrangement +, cadioids, main axis directed to the center, the shot from the central microphone
P- arrangement +, cardioids, main axis directed to the center, the shot from the center of the arrangement
R- arrangement +, cardioids, main axis directed to walls, the shot from the central microphone
S- experimental technique, omnis, the shot from the place for the performer
T- experimental technique, cardioids, the shot from the place for the performer
U- Double ORTF, omnis, the shot from the place for the performer
W- Double ORTF, cardioids, the shot from the place for the performer

Some information about the impulse responses was gathered before starting the convolution processes. While looking at the waveform view, it is noticeable that time shifts between the impulse responses are preserved (exemplary waveform and spectral views presented in Fig. 6-10 came from the recording G). As it was expected, increase in distance caused bigger delays. Similarly, responses from further microphones reveal decreased levels. Spectral views did not show significant differences. Some differences in waveform and spectral view are visible only in the response of the central loudspeaker. It could be caused by the proximity effect (the shot was very close to the microphone). In the next set of pictures (Fig. 11-16) are presented original signal and signals obtained after the convolution process are presented. Changes in the waveform views are visible, especially decrease in volume in case of surround channels. Different time shifts are also observable. Significant changes occurred in spectral views. Convolved files have reduced higher frequencies in comparison to the original sound excerpt. Large diaphragms of microphones led to minimal coloration, audible at low frequencies (the proximity effect). Some efforts were put to correct this effect, but results were rather poor.

Some problems occurred with the signal of the central channel. There was no clear answer whether to use only “dry” signal or convolved one. After some listening, decision was made about employing the convolution process to that end.

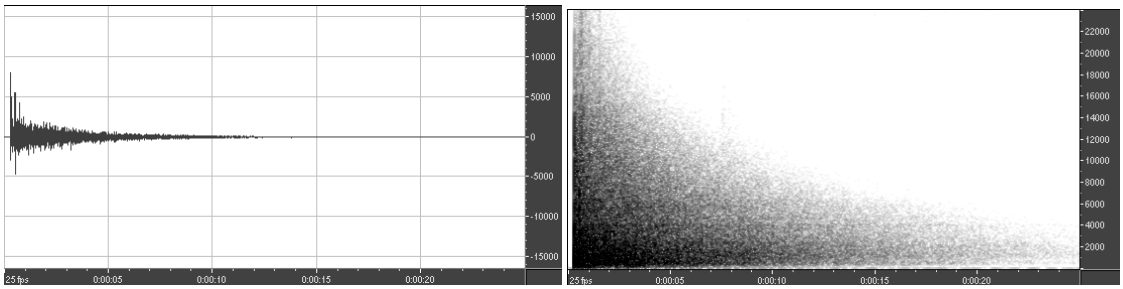


Fig. 6. Waveform and spectral views of the impulse response for the left front loudspeaker

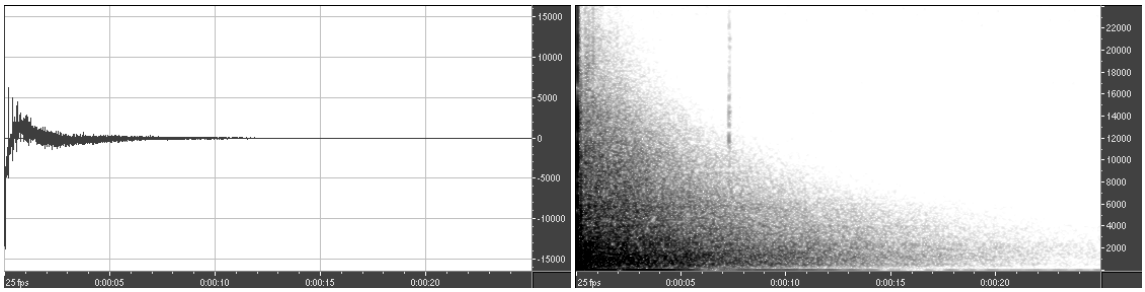


Fig. 7. Waveform and spectral views of the impulse response for the central front loudspeaker

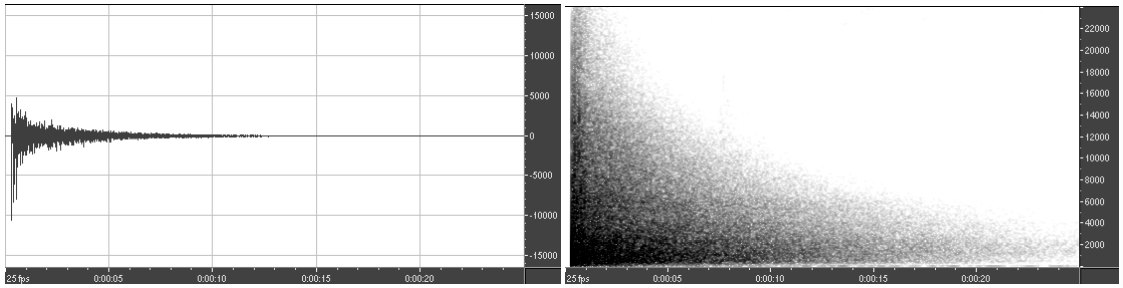


Fig. 8. Waveform and spectral views of the impulse response for the right front loudspeaker

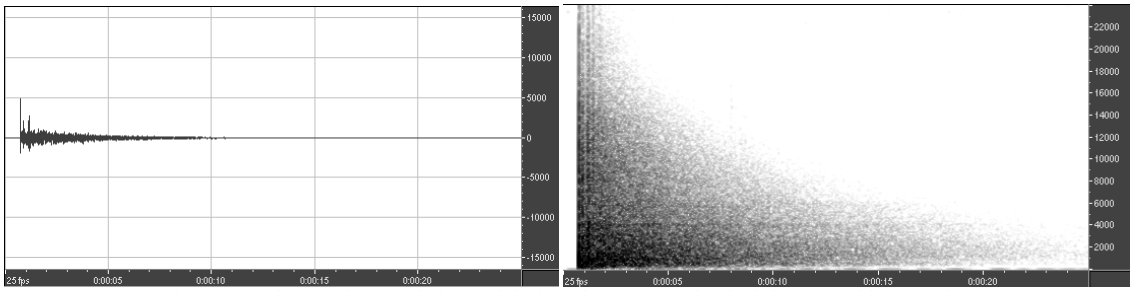


Fig. 9. Waveform and spectral views of the impulse response for the left surround loudspeaker

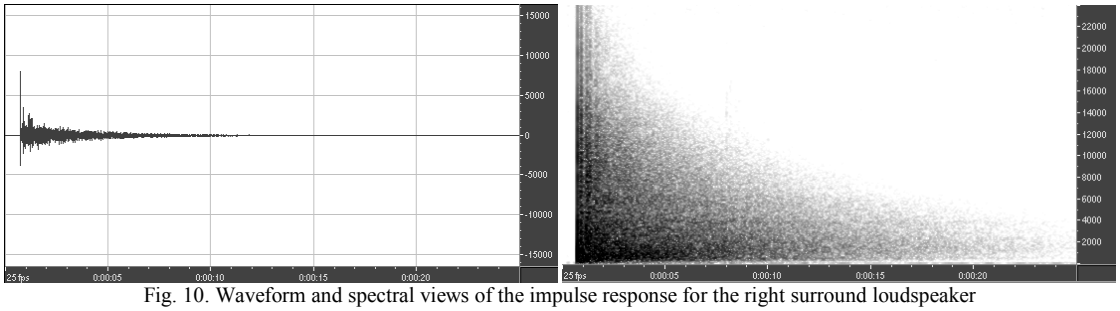


Fig. 10. Waveform and spectral views of the impulse response for the right surround loudspeaker

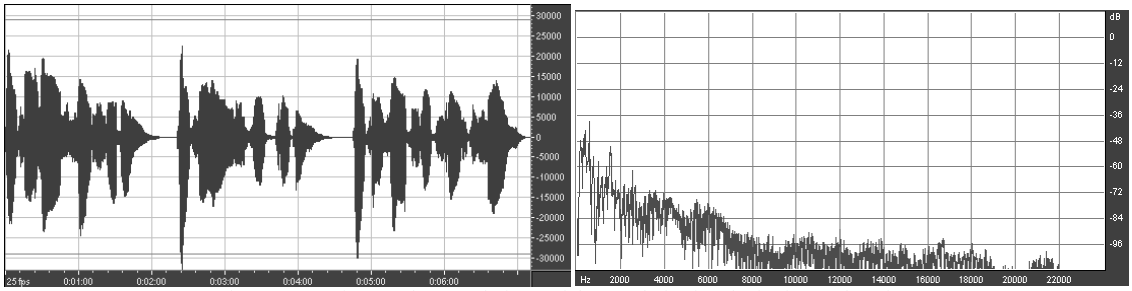


Fig. 11. Waveform and spectral views of the anechoic signal (singing male voice)

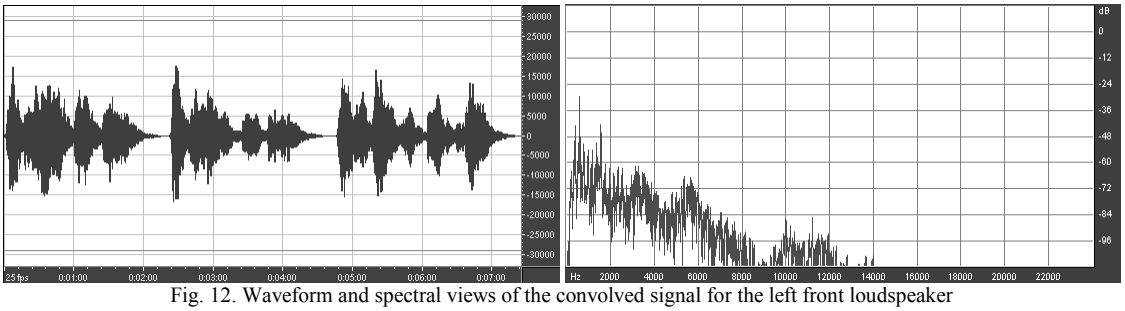


Fig. 12. Waveform and spectral views of the convolved signal for the left front loudspeaker

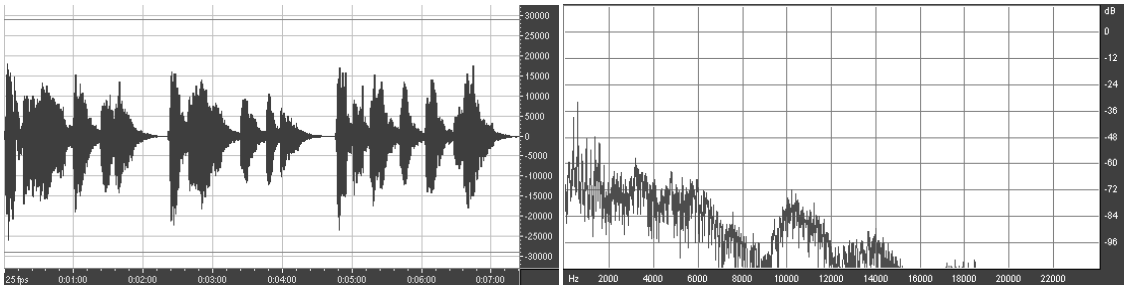


Fig. 13. Waveform and spectral views of the convolved signal for the central loudspeaker

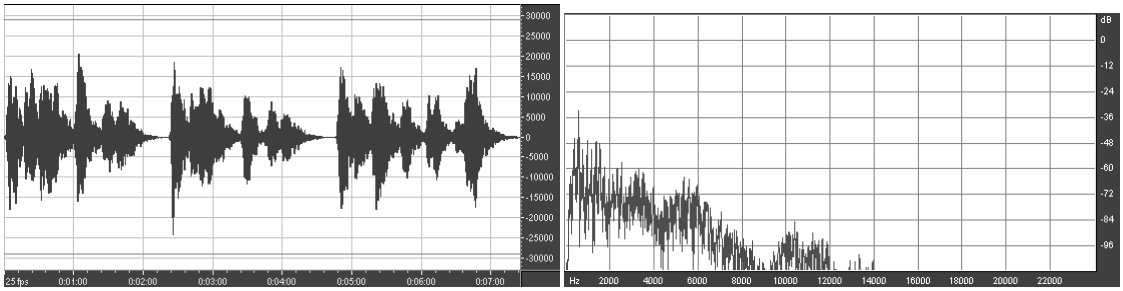


Fig. 14. Waveform and spectral views of the convolved signal for the right front loudspeaker

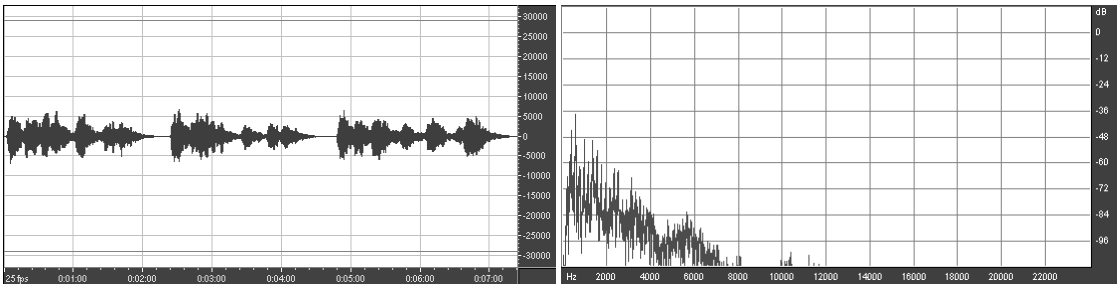


Fig. 15. Waveform and spectral views of the convolved signal for the left surround loudspeaker

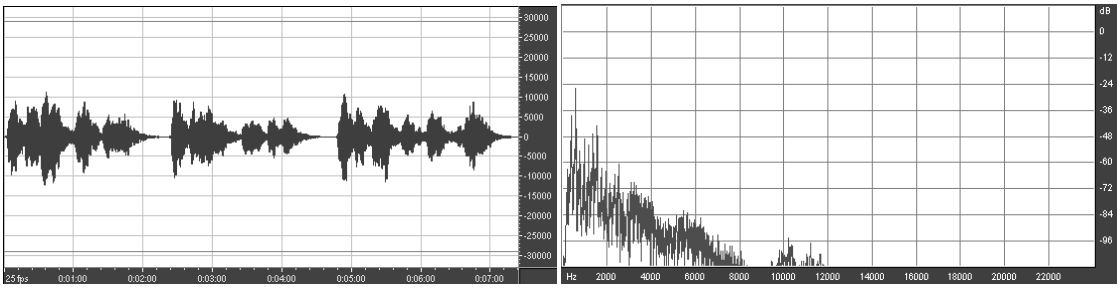


Fig. 16. Waveform and spectral views of the convolved signal for the right surround loudspeaker

Tab. 1. Average results of preliminary tests – subjective grades (sound samples ordered best-to-worse from the left to the right)

The Auditory Hall	J	T	K	W	D	L	I	B	S	H	F	P	G	U	M	E	N	O	R	A	C
Reverberation	0	0	0	0	0	0	0	0	0	1	0	0	0	1	-1	0	0	0	0	-1	-1
Clarity	0	0	0	-1	-1	-1	0	-1	-1	-1	-1	-2	2	-1	-2	2	-2	-2	2	2	2
Coloration	0	0	-0.5	0	0	0	1	0	0	0	-0.5	0	1	1	0	1	1	1	2	2	2
Tonal balance	0	0	0	-0.5	0	0	0	0	0	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1
Spaciousness	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	-1	-1	-1	-1	-1	-1
General impression	0	0	-0.5	-0.5	-1	-1	-1	-1.5	-1	-1	-1	-2	-2	-1	-2	-2	-2	-2	-2	-2	-2

Results

Eight experts took part in the subjective test session. The recordings were assessed in the parametric test. The experimental scenario was as follows. Subjects were asked to listen to the real and then to the prepared sound signal excerpts and were asked to try assigning grades to a provided list of parameters with reference to real recordings. Between each piece intervals of 5 seconds were made to enable jury to make notes in a questionnaire form assigned to this test. The following attributes were judged: reverberation, clarity, coloration, tonal balance, spaciousness, general impression. Values for each parameters (except the last one) could vary from –2 to +2 with 0.5 step. The general impression parameter varied from –2 to 0 with 0.5 step. A short interpretation is needed: values under zero meant that the expert's feeling was worse than while listening the original excerpt. Positive values were reserved for the recordings ranked higher than the real one. If there was no difference between the original and the convolved recording, the expert was supposed to select zero. Averaged results are presented in Tab. 1.

In listeners' opinion the best recordings were J and T. They were the most similar to the real recordings made in the Auditory Hall. The recording J was made using microphones placed at the corners of the room, about 180cm from walls, with cardioid polar patterns, main axis aimed at the center of the arrangement, and the shot also executed at the center. The recording T was made using the experimental microphone arrangement and the shot was at the place for the performer. As in J, cardioid polar patterns were switched on in microphones. The tendency of choosing the recording T by the experts might be caused by the fact that one of the real surround recordings was made using a similar microphone arrangement. By the term "real recording" we consider the recording employing the whole audio material (not only the monophonic source and impulse responses used in the convolution process).

It is noticeable that almost all recordings were assigned negative values for the tonal balance. Probably in this way some defects in firecracker explosion sharpness and smoothness influenced the sound. In practice, these explosions are not fully omnidirectional and reveal falling spectral characteristic.

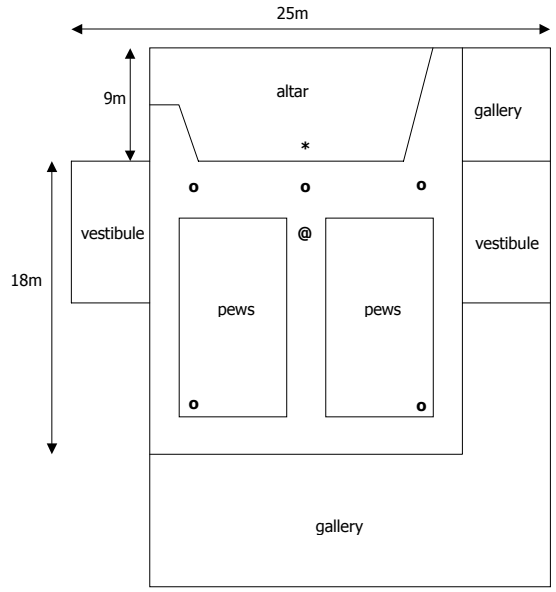
MAIN EXPERIMENTS

Main experiments were made in the Zmartwychwstancow Church in Gdansk. The lay-out of the church is presented in Fig. 17; its height is about 18 m. The reverberation time equals to 4.1s, thus it is quite long. In these conditions the ambient surround recording provides quite challenging task.

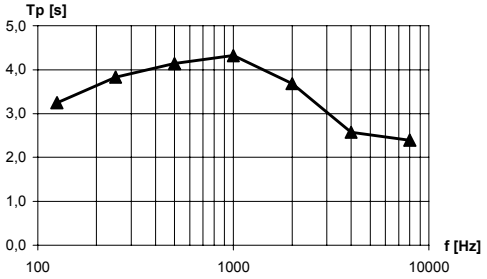
Material

The microphone arrangements chosen as the best one during the preliminary experiments were used, namely microphones placed at the corners of the room, about 180cm from the walls, microphones adopting directional polar patterns. Different shot places were selected and microphones were placed according to the experimental technique lay-out. Similarly to the preliminary tests, the Double ORTF technique also was used. The full list of recordings made is presented below:

A- arrangement o, omnis, the shot from the place for the performer
B- arrangement o, omnis, the shot from the center of the arrangement
C- arrangement o, cadioids, main axis aimed at the center, the shot from the performer
D- arrangement o, cardioids, main axis aimed at the center, the shot from the center of the arrangement
E- arrangement o, cardioids, main axis aimed at walls, the shot from the performer
F- arrangement o, cardioids, main axis aimed at walls, the shot from the center of the arrangement
G- experimental technique, cardioids, the shot from the place for the performer
H- Double ORTF, cardioids, the shot from the place for the performer



a)



b)

Fig. 17. Lay-out of St. Church (a) and its reverberation time characteristics (b)
o - microphones placed at the corners of the room, about 180 cm from walls; * - loudspeaker
@- experimental and Double ORTF technique

Similarly to the preliminary tests, some observations were made employing waveforms and spectral analyses of the impulse responses. Conclusions are identical as in case of the preliminary tests. That is why, only few figures are enclosed (Fig. 18-20). They are related to the recording A.

Results

The experiments scenario was similar to the preliminary tests. The same eight experts took part and they assessed the same attributes. Obtained results are presented in Tab. 2.
The results do not differ much from the obtained previously. Experts decided that the best recordings were D and G. It means that experts have chosen the same two arrangements as in the

Tab. 2. Average results of the main tests – subjective grades

The church	D	G	H	C	B	F	A	E
Reverberation	0	0	0	0	-1	0	-1	-1
Clarity	1	0	-1	1	0	-1	2	2
Coloration	0	0	0	0	0	0	2	2
Tonal balance	0	-1	0	0	-1	0	-1	-1
Spaciousness	0	0	0	0	-1	-1	-1	-2
General impression	0	0	-1	-1	-1	-1	-2	-2

CONCLUSIONS

Basing on the experiments described in this paper and opinions of experts taking part in them, it can be stated that the studied approach is appropriate for creating surround sound recordings.
The subjective listening tests proved that the convolution process allows one to create almost “real” surround sound recordings. Furthermore, experts selected only two the best arrangements in their opinion form among many different ones. Their opinion was the same in case of some quite different kinds of recorded material and in case of really different acoustical conditions of recording rooms. The presented method of picking-up ambient sound leads to an effective data compression, because after transmitting 5 impulse responses (providing short signals) only monophonic or stereophonic signal is to be transmitted to the receiving site. The remaining part of the surround sound retrieval (multi-convolution) could be done during the play-back of the audio material.
Both chosen arrangements use cardioid microphones, which is a bit surprising, because in majority of multichannel recording techniques an application of omnidirectional microphones is generally recommended. However, directional microphones can receive sound reflections more selectively with regard to different areas of the recording hall.
There are also some limitations of the introduced sound recording technique. First of all, the listening room must have reverberation time much shorter than the simulated one. Otherwise recordings will be perceived as unnaturally sounding. However, in typical home theater systems this condition is always fulfilled. There is still no clear answer as how to convolve stereo sound files (some basic techniques are discussed in the next chapter). The biggest concern is related to the fact that the presented method should perform in real-time, however it is currently hardly achievable. Hopefully, future computers would be fast enough to convolve five audio streams simultaneously without any audible delays or a specialized signal processing hardware will be employed to this task.

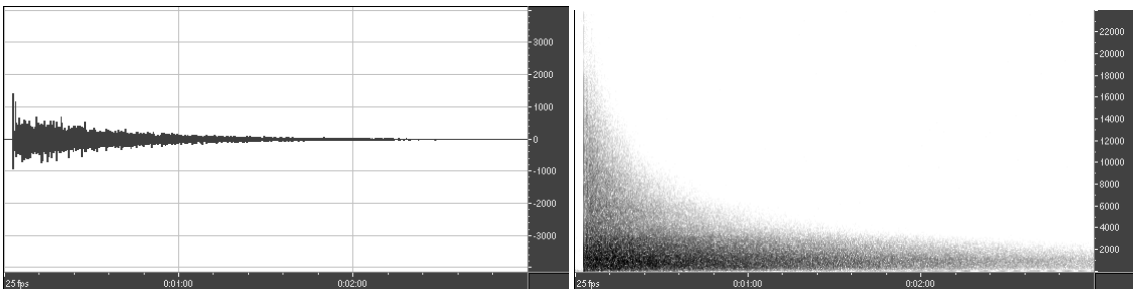


Fig. 18. Waveform and spectral analyzes of the impulse response for the right surround loudspeaker

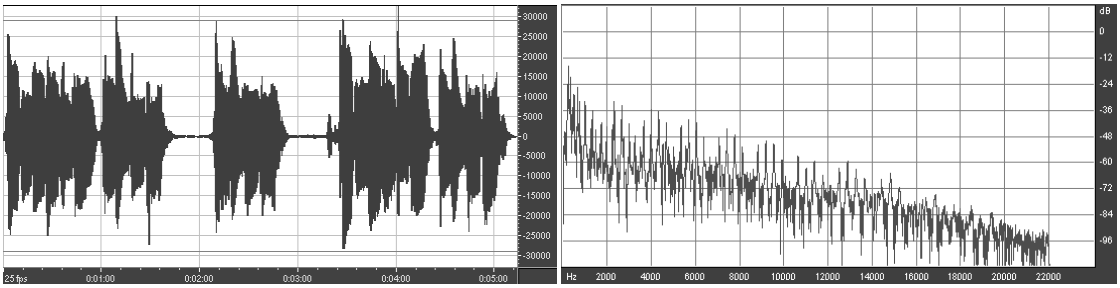


Fig. 19. Waveform and spectral analysis of the anechoic signal (guitar)

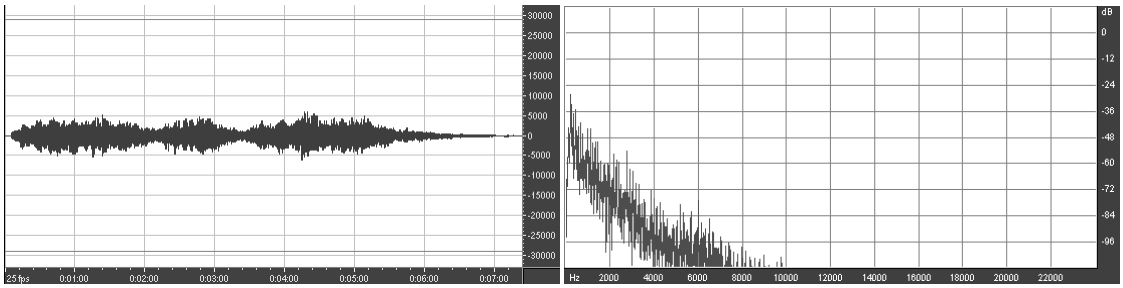


Fig. 20. Waveform and spectral analysis of the convolved signal for the right surround loudspeaker

FUTURE EXPERIMENTS

At the current stage of experiments, only mono sound files were convolved. Described method should allow convolving also stereo files. In this way, it could be possible to create 5.1 recordings using traditional sound source like CD's, TV, radio. Furthermore, the recording would adopt the acoustics of optionally selected room. This remains the subject of future experiments, however some first tests were already made by the authors. The left channel of the stereo recording was convolved with the impulse responses of the left loudspeakers (front and surround). Similarly signals of the right channels were obtained. The sum of two stereo channels was convolved with the impulse response of the central loudspeaker. This experiment showed that assumptions made were correct, but some improvements should be introduced. For example, from signals for the left and right front loudspeakers should be removed the "central channel" information. This "central channel" signal feeds (after adequate convolution process) the central channel. In this way some disturbances in simulated acoustical space continuity might be removed. More results related to this issue will be presented in future papers.

Since the role of the directional characteristics of microphones in the proposed recording technique is important, it is planned to introduce to the signal processing chain the previously engineered algorithms for the spatial filtration of sound based on neural networks [1] [2] [8]. More information on this issue will be included in the oral presentation of this paper.

References

[1] Czyżewski A., Królikowski R., Kostek B., "Neural Networks Applied to Sound Source Localization", 110th Audio Eng. Soc. Con., Preprint No. 5375, Amsterdam, May 2001.

[2] Czyżewski A., Królikowski R., Kostek B., "Encoding Spatial Information for Advanced Teleconferencing", Proceedings of the AES 19th International Conference, pp. 309-322, Shloss Elmau, Germany, 21-24 June 2001.

[3] Farina A., Glasgal R., Armelloni E., Torger A., "Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music", Proceedings of the AES 19th International Conference, pp. 26-46, Shloss Elmau, Germany, 21-24 June 2001.

[4] Farina A., Righini F., "Software Implementation of an MLS Analyzer, with Tools for Convolution, Auralization and Inverse Filtering", 103rd Audio Eng. Soc. Conv., Preprint No. 4605, New York, 1997.

[5] Fukada A., Tsujimoto K., Akita S., "Microphone Techniques for Ambient Sound on a Music Recording". 103rd Audio Eng. Soc. Conv., Preprint No. 4540, New York, 1997.

[6] Glasgal R., "The Ambiohone Deriveration of a Recording Methodology Optimized for Ambiophonic Reproduction". Proceedings of the AES 19th International Conference, pp. 13-25, Shloss Elmau, Germany, 21-24 June 2001.

[7] Kornacki A., Kostek B., Odya P., Czyżewski A., "Problems Related to Surround Sound Production". 110th Audio Eng. Soc. Conv., Amsterdam, Preprint No. 5374, May 2001.

[8] Królikowski R., Czyżewski A., Kostek B., "Localization of Sound Sources by Means of Recurrent Neural Networks". Series: Lecture Notes in Computer Science, vol. 2005, Springer-Verlag, pp. 603 - 610, 2001.

[9] Mason R., Rumsey F, "An Investigation of Microphone Techniques for Ambient Sound in Surround Sound Systems". Preprint No. 4912, Monachium, 106th AES Conv., May 1999.

[10] Theile G., "Multichannel Natural Music Recording Based on Psychoacoustic Principles", 108th AES Convention. Preprint No. 5156, Paris, 2000.

Acknowledgments

The research is sponsored by the Committee for Scientific Research, Warsaw, Poland. Grant No. 8 T11D 00218 and by the Foundation for Polish Science.