



Audio Engineering Society Convention Paper

Presented at the 123rd Convention
2007 October 5–8 New York, NY

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Localization in Spatial Audio - from Wave Field Synthesis to 22.2

Judith Liebetrau¹, Thomas Sporer¹, Thomas Korn¹, Kristina Kunze², Christoph Mank², Daniel Marquard², Timo Matheja², Stephan Mauer², Thomas Mayenfels², Robert Möller², Michael-Andreas Schnabel², Benjamin Slobbe², and Andreas Ueberschaer²

¹*Fraunhofer IDMT, 98693 Ilmenau, Germany*

²*Technische Universität Ilmenau, 98693 Ilmenau, Germany*

Correspondence should be addressed to Thomas Sporer (spo@idmt.fraunhofer.de)

ABSTRACT

Spatial audio reproduction used to concentrate on systems with a low number of loudspeakers arranged in the horizontal plane. Wave Field Synthesis (WFS) and NHK's 22.2 two systems promise better localisation and envelopment. Comparisons of 22.2 with 5.1 concerning spatial attributes on one hand, and evaluation of spatial properties of WFS on the other hand have been published in the past, but different methods have been used.

In this paper a listening test method is presented which is tailored on the evaluation of localisation of 3D audio formats at different listener positions. Two experiments have been conducted: In the first experiment the localisation precision of 22.2 reproduction was evaluated. In a second experiment the localisation precision in the horizontal plane as a function of spatial sampling was studied.

An important factor for perceived audio quality is the localization of sound sources. A lot of research in spatial audio has been concentrated on an improvement of sound localization in the horizontal plane and on improvement of envelopment. New approaches promise more realistic sound including proper perception of direction, distance and eleva-

tion. Applications of such systems will include cinemas, theaters and homes.

For a realistic interaction of the visual and hearing sense it is important that the localization of the audio signal is equal to the position of the audio object in the movie scene or on the stage. Other-

wise the audio quality will decrease. To achieve realistic surround sound, several approaches like Ambisonic [2], which uses the theory of spherical harmonics for a reproduction of the sound field, Vector Based Amplitude Panning [3], amplitude panning in all dimensions by using vectors to describe the virtual sound position, or the Wave-Field-Synthesis [4], based on the Huygens Principle, are used. All these approaches have a variable number of loudspeakers and so the complexity of the setup is variable. A channel based approach using 22 loudspeakers in a fixed three dimensional arrangement plus two low-frequency effect channels (therefore called 22.2) has been proposed by NHK [1].

An infinite number of loudspeakers and the use of a complex software model for reproducing the propagation of sound waves would create the best 3D impression of a sound field. Unfortunately it is necessary to make a compromise between the localization resolution and the complexity of the system which is often combined with large costs and implementation complexity. To analyze the correlation of the system complexity (for example the number of loudspeakers) and the localization resolution it is necessary to conduct listening tests with different setups. This paper describes the accomplishment and results of 2 listening tests.

The first test is about a three dimensional 22.2 surround system with an loudspeaker adjustment proposed by NHK [1]. In the test we combined this setup with a panning algorithm currently under development. The focus is on the ability of humans to localize sound sources played by the system on different positions in all three dimensions.

The second test is about localization in the horizontal plane. A Wave Field Synthesis ring, which provides very good localization, is compared with a system using fewer loudspeaker in the number of loudspeakers and panning. Note, that this system is not exactly the same as NHK's 22.2 but has similar properties concerning horizontal localization.

1. EXPERIMENTAL DESIGN

The goal of the test is to evaluate whether a sound reproduction system is able to reproduce spatial audio signals in a way that listeners at different positions in the listening room have the same or at least

plausible spatial perception. The most important issue in such listening test is how to obtain data about perceived direction from the listeners. In contrast to previous work, where the general ability of humans had been in focus, we wanted to use a method able to assess the 3D properties of reproduction systems which enables to repeat such a test with manageable efforts. At the same time we wanted to include as many source positions as possible in such a test. Many tests in the past have been done using artificial sound signals like pink noise only. In application scenarios people listen to music, speech or noise, and such signals might have different properties concerning localisation. Due to these reasons the following design decisions have been made:

- up to three listeners can do the test in parallel
- the coordinate system is printed on the wall
- the listeners have to fill their scores in a questionnaire
- to avoid listeners fatigue different test material is used within each test session
- an essential part of the test preparation is the selection of test material

The test items are randomized in position and sequence for each test session. Each test subject usually has to participate in three test sessions to cover all three listening positions. In addition to the position of the sound source the subjects are also asked to indicate how "well defined" the position is.

1.1. Listening room

A test setup was built up in a fabric hall with inadequate acoustic qualities. To obtain good acoustical conditions, acoustical treatment of the room had to be done. The boundaries of the listening room are constructed with sound absorbing elements. The floor is covered with a carpet. The reverberation time of the room is given Table 1 An acoustic transparent canvas is attached about one meter in front of the acoustical walls. The loudspeakers are placed behind the textile, so the subject is not able to see them. This has been done to avoid any hint about setup information avoiding bias of the listener. The covering of loudspeakers leads into a visible rectangular listening room with a floor space of 3.36 m by 5.30 m and a height of 3 m.

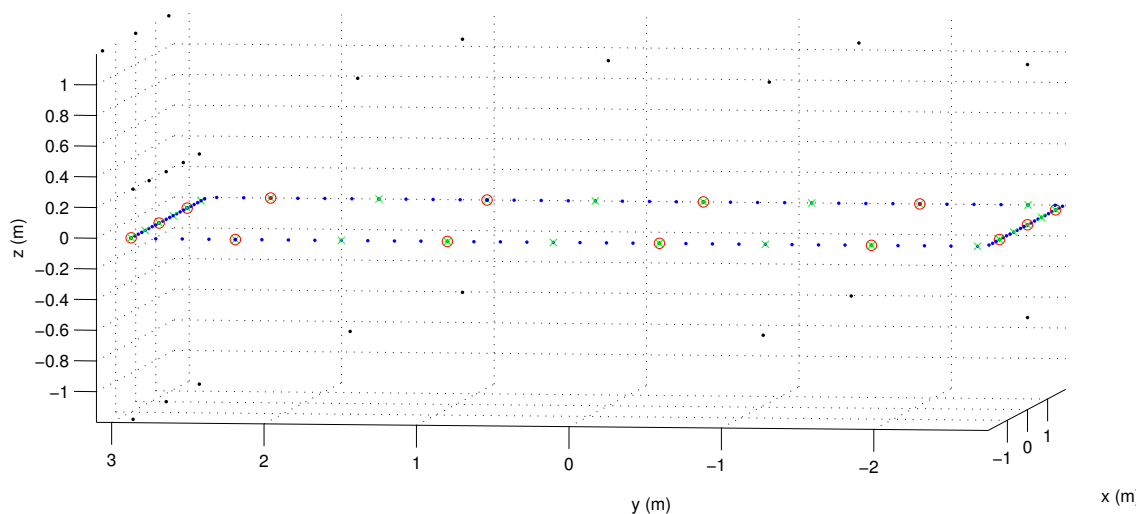


Fig. 1: Loudspeaker arrangement for experiments A and B.

octave band [Hz]	RT [s]
125	0.46
250	0.39
500	0.32
1000	0.24
2000	0.16
4000	0.14
8000	0.12

Table 1: Reverberation time of the listening room

1.2. Loudspeaker Arrangement

To enable comparability of the results the test systems for both experiments use loudspeakers of the same type (K&H CA 106). For the WFS ring 108 loudspeakers are used. The dimensions of the WFS ring are 3.85 m in width and 5.60 m in length. The 3D sound system is built up with 22 loudspeakers. The single loudspeakers are arranged due to the system's objectives and is derivate from the 22.2 setup proposed from NHK [1]: There are three layers of loudspeakers. The loudspeaker concentration is highest in the front and declines toward the back of the room because of the aspired usage in cinema applications. The arrangement as a matter of principle is shown in Figure 1.

Note, that in the setup in total 6 subwoofers (4 for

WFS and 2 for 22.2) are present but had been disabled during the listening tests.

1.2.1. Signal Flow

The optical digital audio output of the playout PC was linked to rendering PCs which do the signal processing of the WFS and panning algorithm. In case of experiment A, only one Rendering PC was used (Panning). In case of experiment B, two rendering units were necessary (Panning and WFS). The rendered signals were transmitted via MADI to the Lawo Dallis. Here the D/A conversions were done. The analog signals were sent to a Powersoft Q4004 amplifier array. Finally the signals were transmitted via common speaker cable to the loudspeakers.

For both tests Control PCs were used. Each of them is controlling the signal playout and the different rendering algorithms. To enable an easy handling of the system, a Graphical Control Interface was implemented in Pure Data.

To inform the listeners about which trial is currently running, an additional PC with a screen in the listening room has been installed. This computer is connected to the Control PC via Ethernet.

1.3. Test data

Three different audio samples are used in the first experiment to study the influence on the results of different program material. The first is a female voice

and has been selected because it is known that localization is very sensitive with speech and singing voice and that the human voice is very critical in sense of audio quality. The second is an acoustic guitar that has been chosen to incorporate a more transient signal. Note, that both voice item and guitar item are not recorded in the anechoic room and that both signals contain some spatial information about the recording room. The third sample is pink noise, a signal frequently used in spatial listening experiments. No room simulation is applied. For experiment 2 only pink noise and female voice are used.

1.4. Training of subjects

Before the real test starts, the subjects had to attend a training session. This is done to equal the state of knowledge of all subjects. The listener should become familiar with all conditions and items under test. In the trainings session all three test signals are presented at all positions to be used later in the test.

1.5. Test procedure in detail

The test is designed as multiple-subject test. That means that three listeners can attend the test at the same time. The subjects are placed on positions which are relevant for the evaluation. These are the sweet spot and additional critical positions. The sweet spot position, called position 2 is placed 0.08 meter in front of the listening room's center, the point of origin. Position 1 is set one meter left and one meter in front of position 2. Equal to this, position 3 is set each one meter towards the back and right respective position 2. The resultant dimensions with reference to the point of origin are shown in Figure 2.

The diagonal positioning is chosen to reduce acoustic shadowing caused by the subjects. First of all, the subject will be introduced to the test. The instruction includes information about the background of the test, the test procedure and the usability. The sitting position of each listener is adjusted to the same height of ear (0.9 m above the floor). This height is equal to the height of the tweeter of the loudspeaker ring in the middle. Afterwards there is the training as described in section 1.4. The test consists of three subtests. Every subtest has its own randomized playlist regarding stimulus and source position to avoid sequence effects. It is supposed

that the number of subtests is equal to the number of listening positions. In that case, every listening position is occupied by one subject. Every subtest consists of 39 trials in experiment A and 48 trials in experiment B. Within every trail, the source position to be scored is presented twice for 7 s separated by a short pause of 1 s. After the second presentation a pause of 10 s is provided for scoring. After that, the next source position is played back. Due to the automatic test procedure, subjects can not listen to the condition as often as they want. The whole room is covered with acoustic transparent canvas to hide the loudspeakers. Furthermore, the raster (resolution 10 cm by 10 cm), displayed in Figure 3, is drawn on the canvas.

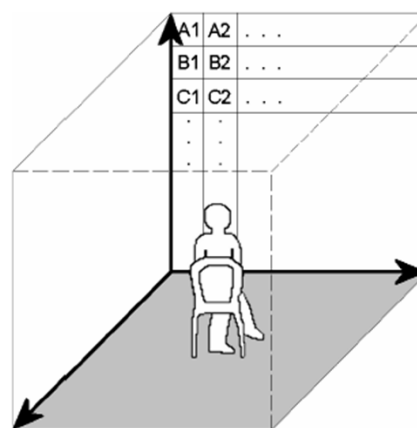


Fig. 3: Coordinate system drawn on the canvas hiding loudspeakers: Columns 1 ... 33 indicate positions on the frontal wall, while columns 34 ... 85 are at the right wall. Rows indicated by letters A ... Z.

The description of the localized position is done via this coordinate system. In addition to the coordinates, the character of the sound source is requested. The subject should decide if it is diffuse or accurately defined according to a four point quality scale. "Very diffuse" accords to 4 and "well defined" is defined to 1. The subjects have to fill in a form with the needed information as shown in Table 2.

When the first session is done, there is a short break. After the break the test subjects change to another listening positions. After three sessions each listener has been listening at all three listening positions.

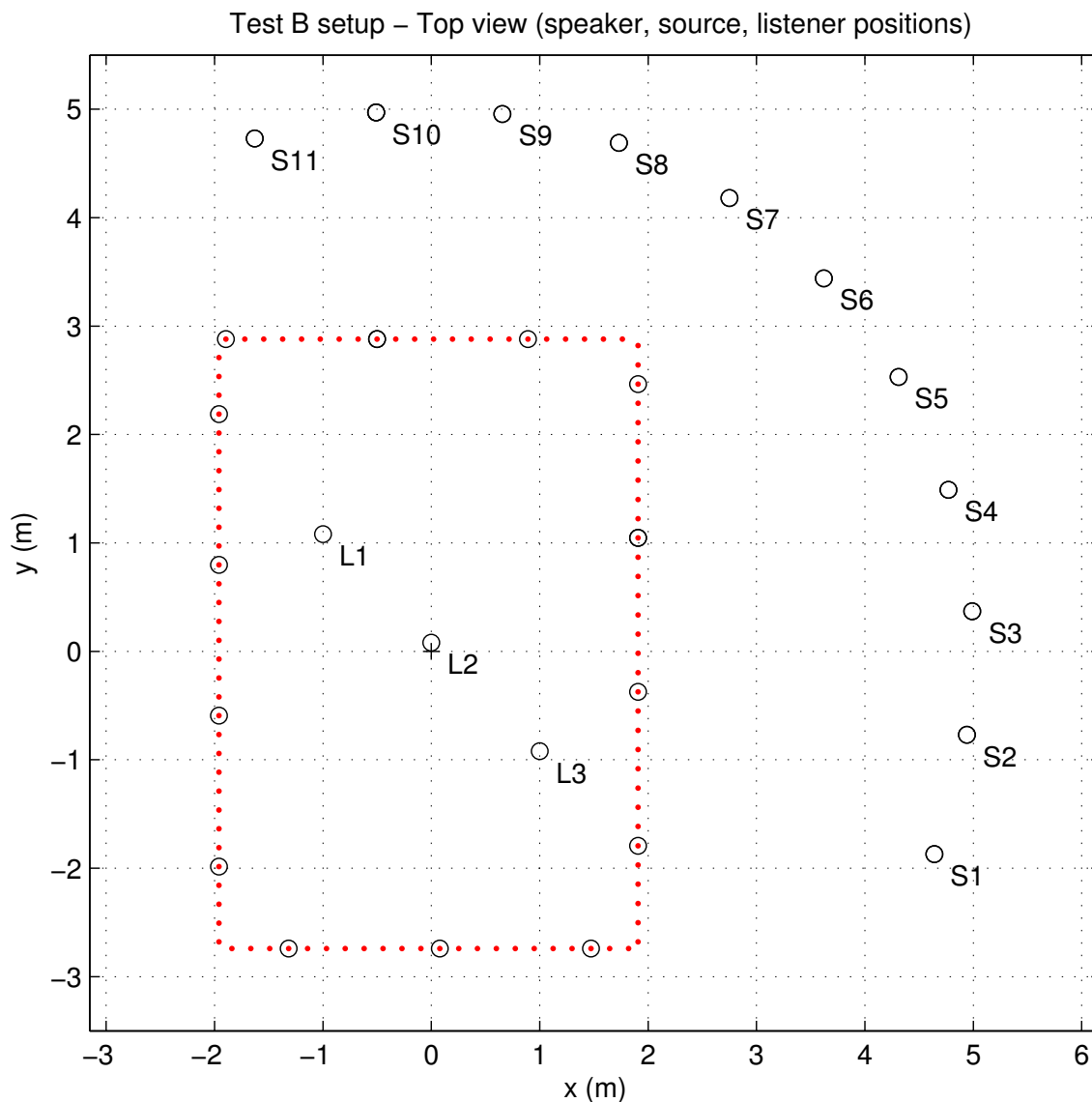


Fig. 2: Position of listeners (L1-L3) and loudspeakers. Also shown is the position of the virtual sound sources (S1 to S11) used in experiment B. 108 loudspeakers indicated by \cdot , subset of 14 loudspeakers for experiment B indicated by \odot .

Stimulus	1
Coordinates	
Spatial Image	1 - 2 - 3 - 4

Table 2: Excerpt from evaluation sheet

The alternation of the changing order helps to avoid sequence effects. The test conditions comprise three anchors and one repetition of one source position. In a post-screening process, an additional analysis is done to reject unreliable listeners. This analysis is based on the anchors and the repetition of one source position.

1.6. Test panel

The test panels consist of staff members of Fraunhofer Institut für Digitale Medientechnologie IDMT, the Technische Universität Ilmenau and of students. All of them are trained for listening tests and some of them are active musicians or are doing recordings.

2. EXPERIMENTS

2.1. Experiment A

Three different test signals are presented. They are each played randomly on 12 virtual source positions and 3 anchor positions. The order of the source positions is given by a randomized playlist. The subjects are asked to note the source's position and its definition. In this experiment 34 subjects participated.

2.2. Experiment B

Although in this experiment the sound source reproduction of two different audio systems is compared, we decided not to use a paired comparison method described in ITU-R BS.1284 [6]. The stimuli are played back in random order using one of the two playback devices. The subjects are told to evaluate and score the actual source position, disregarding changes in sound coloration which might be caused by the different reproduction principles. 112 virtual source positions, 2 anchors and 2 repetitions were used. 41 subjects participated in this second experiment.

3. RESULTS

None of the subjects has shown unreliable performance and therefore all subjects have been used in the statistical analysis. There are two alternatives to compare the results of spatial listening tests from different listening positions:

- The direction of each loudspeaker, each virtual (panned) sound source and each perceived sound source is transformed to a spherical coordinate system centered around the listener position. This method can be called a “user-centric” approach. Using this approach it is difficult to compare the results of different positions directly. However it is possible to compare the mismatches between virtual sound direction and perceived direction.

- All directions are projected on the wall of the room, like using a projection screen. This method is adequate to evaluate the audio-visual coherence especially in the context of sound for cinemas.

In this paper we only present the result of the first evaluation method. The data collected in this listening tests contain much more information which will be subject of further studies.

3.1. Experiment A

Observations in Figure 4: If the sound source is

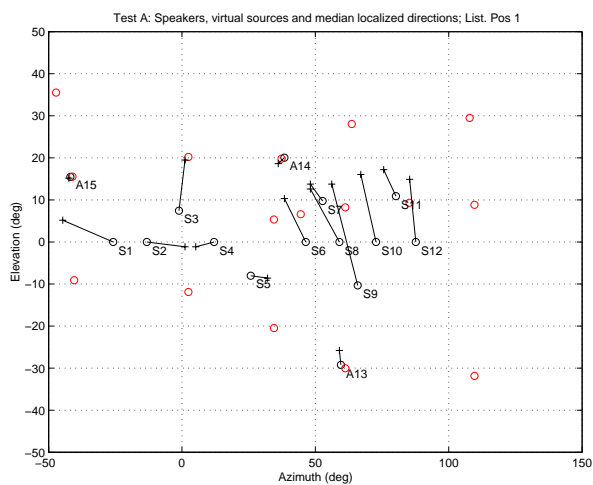


Fig. 4: Test A, listener position 1, loudspeaker positions \circ , virtual sound direction \circ and median of perceived direction $+$ connected with line

between speakers with a large angular distance between each other, then the perceived position is moved towards one of the loudspeakers. There is a tendency to overestimate the elevation of sound sources. The horizontal error is smaller if the loudspeakers are closer together. The perceived position of anchors (A13-A15) is very close to real position.

Observations in Figures 5 and 6: If the sound source is between speakers with a large angular distance between each other, then the perceived position again is moved towards one of the loudspeakers. Again there is a tendency to overestimate the elevation of sound sources. The horizontal error is smaller if the

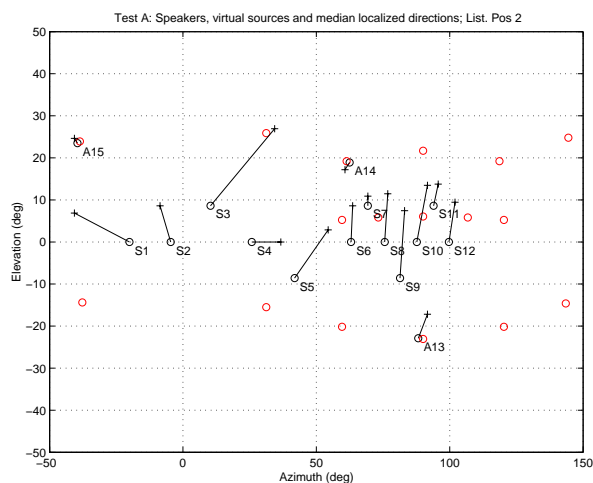


Fig. 5: Test A, listener position 2, loudspeaker positions \circ , virtual sound direction \circ and median of perceived direction $+$ connected with line

loudspeakers are closer together (see especially S6 to S12). The perceived position of anchors (A13-A15) is very close to real position.

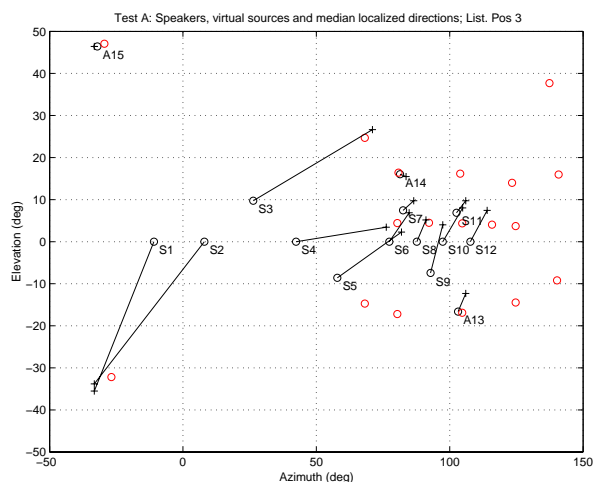


Fig. 6: Test A, listener position 3, loudspeaker positions \circ , virtual sound direction \circ and median of perceived direction $+$ connected with line

Figure 7 gives an indication about the inter-subject

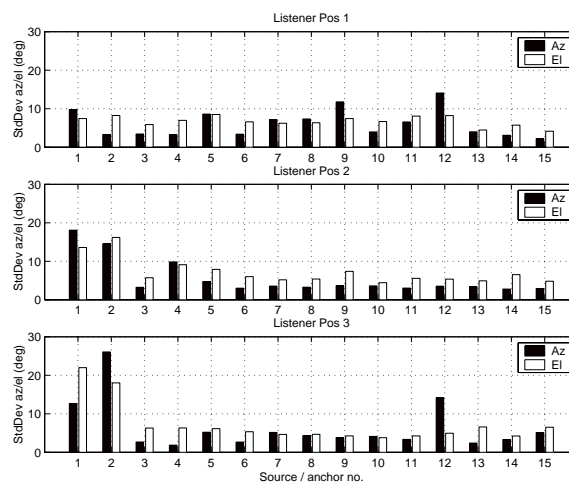


Fig. 7: Test A, coherence of results of different listeners: std.dev. of scores given by listeners, separated in azimuthal (black) and elevation (white)

differences at the different listening positions. Scores of listeners are most coherent at position 2 ("sweet-spot"). There are large differences between listeners for source positions S1 and S2 at all listener positions and both in elevation and azimuth. There are large difference of azimuthal scores given by listeners at listener positions 1 and 2. For the source position S12 Listeners in position 1 (closest to the front) have less coherent results than listeners at the other two positions.

3.2. Experiment B

Figure 8 shows that the reduced number of loudspeakers increases the probability that listeners perceive sound source close to one of the neighboring loudspeakers. Most of the sources are perceived as elevated. This might be a systematic error of the coordinate system, a property of the reproduction room or a perceptual effect, and needs further analysis.

Figure 9 shows that the scores of listeners are most coherent at position 1. Both reproduction schemes have about the same coherence of scores. Source position S11, which is in the front left, seems to be difficult to localize for all listener positions. This effect needs further analysis, too.

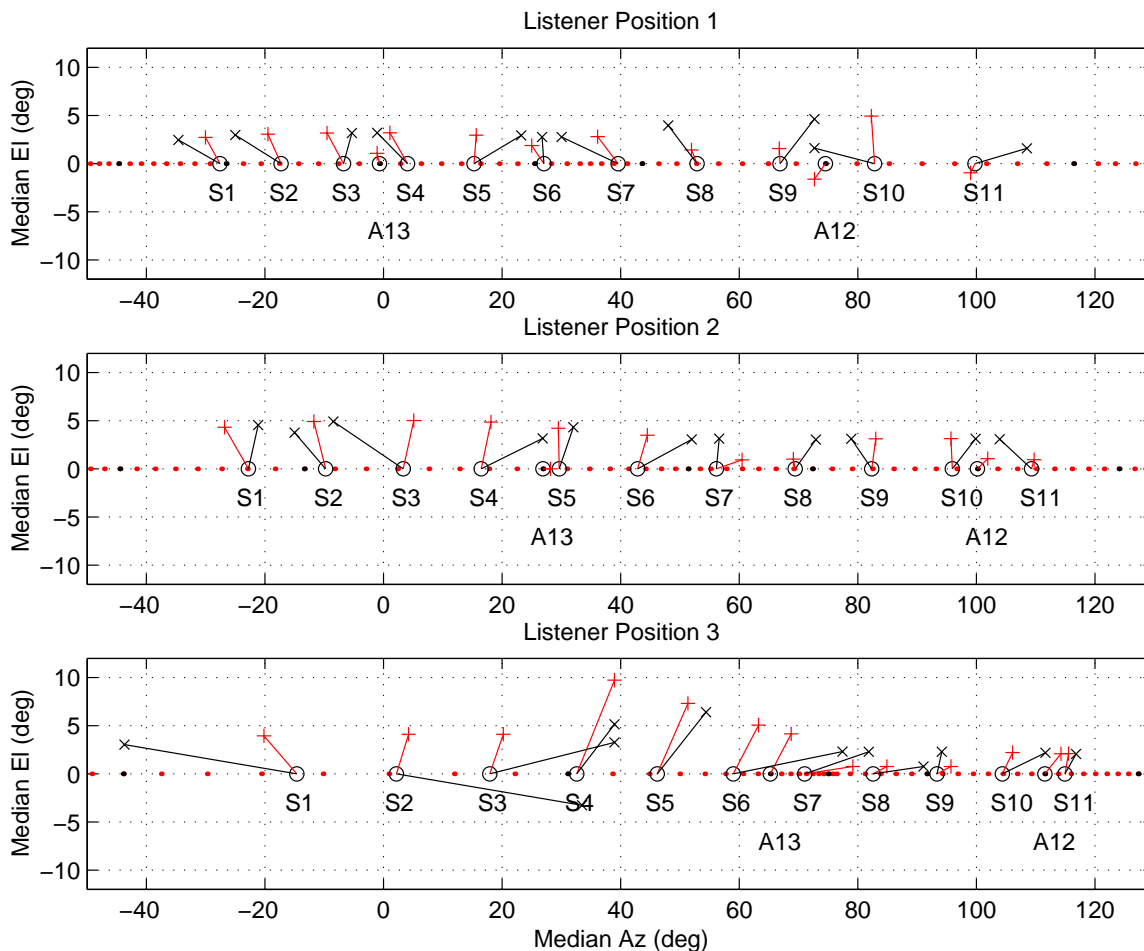


Fig. 8: Test B, all listener positions, direction of virtual source \circ connected by lines with median of perceived direction of WFS $+$ and of panned reproduction \times . A12 and A13: anchors (only one loudspeaker playing)

4. CONCLUSIONS

A test method to evaluate the location properties of spatial audio reproduction systems was presented. The test method was applied to NHK’s 22.2 and WFS. It was shown that the method is able to evaluate the location properties of the systems. This includes the perceived resolution depending on the density of loudspeakers at different walls. It can be seen that the large distance between loudspeakers in 22.2 very often leads to significant offsets between the intended and the perceived direction of loud-

speakers. In our setup we also observed with both systems an offset of the elevation, which might be a real effect or an artifact of our setup.

5. ACKNOWLEDGEMENTS

The presented work is part of the project EDcine, supported by the IST 6th framework program of the European Commission (<http://www.edcine.org>). The authors also want to thank the many patient listeners. Many thanks also to Felix Richter and Christian Rose, for their help while constructing and bulding the test setup.

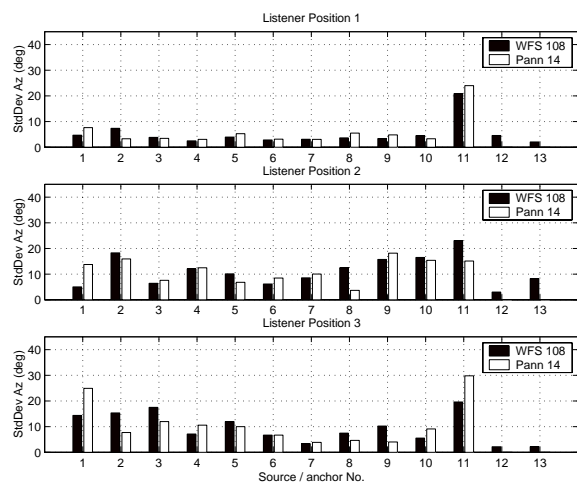


Fig. 9: Test B, coherence of results of different listeners: std.dev. of azimuthal scores given by listeners, for WFS (black) and panned reproduction (white)

6. REFERENCES

- [1] Hamasaki, K.; Hatano, W.; Hiyama, K.; Komiyama, S.; Okubo, H.: “5.1 and 22.2 Multichannel Sound Productions Using an Integrated Surround Sound Panning System” AES preprint 6226, 117th Convention, San Francisco, 2004 October
- [2] D.G. Malham, A. Myatt; “3-D Sound Spatialization Using Ambisonic Techniques,” *Computer Music J.*, vol. 19, no. 4, pp. 58-70, 1995
- [3] Pulkki, Ville; “Virtual Sound Positioning Using Vector Base Amplitude Panning,” In: *J. Audio Eng. Soc.*, Vol. 45, No. 6, 1997 June
- [4] Berkhout, A. J.; de Vries, D.: “Acoustic Holography for Sound Control” AES preprint 2801, 86th Convention, 1989 February
- [5] Klehs, B.; Sporer, T.: “Wave Field Synthesis in the Real World: Part 1 - In the Living Room”. AES preprint 5727, 114th Convention, Amsterdam, 2003
- [6] Recommendation ITU-R BS.1284-1 (12/2003) General methods for the subjective assessment of sound quality. International Telecommunication Union, Radiocommunication Assembly
- [7] Recommendation ITU-R BS.1116-1 (10/1997) Methods for the subjective assessment of small impairments in audio systems including Multichannel Sound Systems. International Telecommunication Union, Radiocommunication Assembly
- [8] Melchior, F.; Brix, S.; Sporer, T.; Roder, T.; Klehs, B.: “Wave Field Syntheses in Combination with 2D Video Projection”, 24th International AES Conference (May 2003)