

# Preliminary listening test results of the comparison of the ICSA 2D spatial Audio Systems

M. Otten<sup>1</sup>, M. Kob<sup>1</sup>, M. Sandner<sup>1</sup>, G. Theile<sup>2</sup>

<sup>1</sup> *Erich-Thienhaus Institut, Detmold University of Music, Germany, Email: [otten@hfm-detmold.de](mailto:otten@hfm-detmold.de)*

<sup>2</sup> *Institut für Rundfunktechnik (IRT), Munich, Germany*

## Abstract

During the ICSA 2011 in Detmold, Germany, a listening test was conducted to compare different spatial audio reproduction techniques. The object of the study was to gain insight on basic differences and strengths and weaknesses of the different approaches. Due to the limited time frame only reproduction in the horizontal plane was examined. Four different reproduction techniques were selected, stereo, 7.0, mixed order ambisonics and wave field synthesis (WFS) scenarios were presented. A questionnaire was filled in using polar-opposite attribute pairs in a five stage grading. A total of 11 attributes were inquired for each recording. The time frame for each subject was set to 60 minutes. The subject should be able to answer all questions in one minute, this led to eight different recordings, taking all four techniques into account and assuming a maximum duration of 20 seconds for each stimulus. To be able to test as many persons as possible during the conference it was decided to test 3 subjects simultaneously, one seated in the sweet spot and two behind. 68 persons participated. The overall results show only slight differences between WFS and 7.0, detailed examination of the answers revealed dependencies between the musical style and the favored technique. The excel sheets containing all gained data are included on the proceedings disc, some short remarks are provided in the following.

## Introduction

The international conference on spatial audio (ICSA) in Detmold, Germany, in November 2011 provided an excellent occasion to conduct an exhaustive comparison of commonly used reproduction techniques as more than 100 expert listeners were gathered. To be able to compare the different approaches, bias should be excluded as far as possible. For this purpose a recording session prior to the conference was carried out. Several musical pieces were recorded, the application of over 100 simultaneous recorded microphone channels provided sufficient material to be edited, mixed and played back in various audio formats. This way the musical performance was the same for every technique and could not impair the judgment.

Although the recorded material allowed to add height information to the mix, it was decided to dismiss this factor and thereby the third dimension for the listening tests because of the limited time frame and the playback situation. During the conference the 3D recordings were presented in various workshops.

## Preliminary work, the recording

It was the intention to study the performance characteristics of different spatial audio systems in different practical musical recording situations. For this reason a recording session prior to the conference was carried out. A number of musical pieces were recorded. More than 100 simultaneously microphone signals provided sufficient material to be edited, mixed and played back in various audio formats. This way the musical performance was the same for every spatial audio system. Still critical was the need to minimize the influence of the artistic intentions in the mixing process in order to ensure relevant comparative test material with respect

to performance characteristics of spatial audio systems rather than with respect to Tonmeister's performances. Against this background a binding recording and mixing ("production") guideline was given:

- Esthetical goal is a hearing impression close to the natural impression in the hall, giving special regard to the attribute "spatial imaging" while taking into account sound color and presence of musicians
- For the mixes only the system-specific microphone signals should be used, if absolutely necessary additional spot signals
- Documentation of the mixes should allow repeatability of mixing results

A three-day recording session took place in the Konzerthaus Detmold. Various microphone settings were installed, resulting in about 100 microphone channels to be recorded simultaneously. The recording was conducted from one of the control rooms of the Erich-Thienhaus Institute, one Tonmeister (Michael Sandner) and three Tonmeister students were responsible for the unobstructed realization. Various musical pieces performed by different ensembles were recorded.

After the recording the most beneficial pieces were selected, namely

1. La belle jeunesse by Francis Poulenc, performed by Harumi Saito (piano) and Axel Wolloscheck (vocals)
2. Quintett Nr. 3, G-Dur, op. 88 by Anton Reicha, performed by the Antares Quintett
3. Femkantgeneralen by Pust, performed by the Adlipps! Choir
4. Canzon septimi toni No. 1 by Giovanni Gabrieli, performed by the HfM Brass ensemble (one part of

the ensemble on stage, the other behind the listener)

5. Red and blue for six drummers by Schöllhorn, performed by students of the HfM arranged circular around the listener
6. Livre du Saint Sacrement pour orgue performed by Jan Croonenbroeck on concert organ
7. Audience applause around the listener

The selected items were mixed in four different techniques

1. wave field synthesis by Volker Werner
2. mixed order ambisonics by Jörn Nettingsmeier
3. 7.0 by Toygun Kirali
4. Stereo by Michael Sandner

As one test run should not exceed 45 minutes to one hour it was decided to limit the duration of the reproduction for every technique to approximately 20 seconds, taking into account that the subject would need about one minute to fill in the provided questionnaire.

To allow a direct comparison the same 20 seconds were taken to be reproduced with every technique. The cuts were made as smoothly as possible, they should not interrupt a musical phrase.

## Questionnaire

The questionnaire comprised eleven polar-opposite attribute pairs, sub divided into four groups. The participant was asked to grade the impression on a scale five-grade scale ("1" for positive vote, "5" for negative vote). Additionally they had the opportunity to give "no vote".

- Sound color  
satisfactory - unsatisfactory
- Spatial representation  
perfect spatial impression - imperfect spatial impression  
natural envelopment - unnatural envelopment  
no crosstalk artifacts in surround - strong crosstalk artifacts in surround
- Source imaging  
distant/deep - close/flat  
well balanced directions on stage - unbalanced directions on stage  
well balanced surround directions - unbalanced surround directions  
precise localization - blurred localization  
stable localization when moving head - unstable localization when moving head
- The recording  
sounds natural - sounds unnatural

sound good - sounds poor

## Reproduction room and setup

The test took place in a listening room measuring 5,6 m x 4 m, reverberation time ...s. The room was equipped with an 88 channel wave field synthesis system and a 9 loudspeaker play back setup for 7.0 Surround and Ambisonics reproduction. The signals of the other techniques were played back using 9 additional loudspeakers. To match the necessary frequency range 4 subwoofers were added.

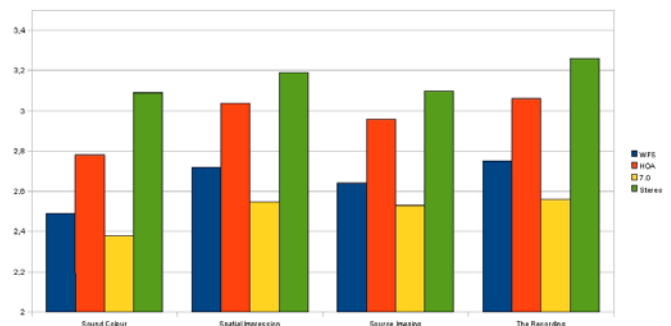
## Course of the test

During the test one person was seated in the common sweet spot (middle of the room), the other two subjects were placed behind. First the investigator explained the procedure and the questionnaire. After listening to the sound sample the questionnaires should be filled in, when everybody was done the person in the front seat pressed a button to trigger the next sample. This way the subjects could determine how fast or slow they proceeded did not have to wait for the next sample or hurry to get the questionnaire done before the next item. During the test the investigator left the room, but a talk-back was installed so that the subjects could ask questions or place remarks during the inquiry.

The sequence of the used techniques was randomized as was the sequence of the different pieces to avoid any influence on the results. The participants were not informed which techniques were used before the test started.

## Results

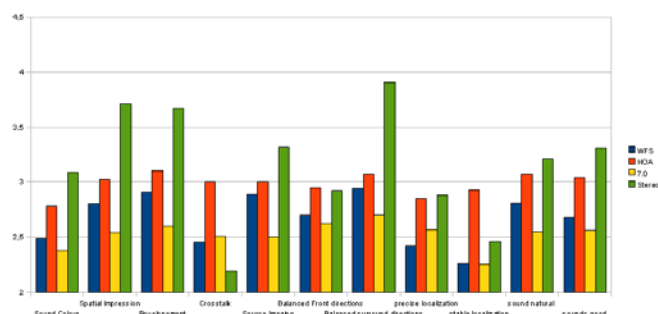
During the ICSA event 68 test persons took part. In the following graphs some results are shown. Please note that the best mark is "1", so the lower the bar, the more positive. It is inherent to tests with a grading scale that most subjects tend to avoid extreme marks ("1" or "5" in this case), so that on average the grading becomes compressed to a scale from "2" to "4".



**Figure 1:** Results averaged over all pieces, grouped into 4 main categories

Looking at the averages for the four categories in Figure 1 one gets a quite consistent pattern: The 7.0 surround recordings have been evaluated particularly positive in all categories while the 2.0 stereo recordings are in the end.

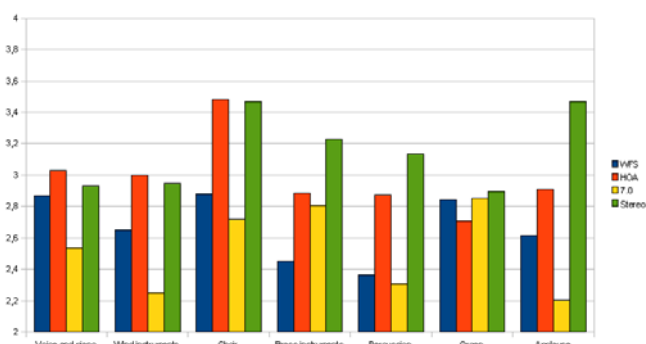
This might be due to the fact that some of the questions are related to the spatial impression of the mix which is obviously limited in the two channel stereophonic representation.



**Figure 2:** Results averaged over all pieces

More details are given in Figure 2. Not surprisingly, for Stereo the spatial attributes "spatial impression", "envelopment" and "balanced surround directions" were evaluated significantly negative.

The other spatial audio systems exhibit only slight variations in the ranking: For most parameters stereophonic 7.0 surround mixes favored over the sound field synthesis representations WFS and HOA. Only the precision of the localization seems to work better in the WFS representations. The rankings of the HOA mixes indicate disadvantages compared with WFS, in particular with respect to "stable localization" and "crosstalk".



**Figure 3:** Results averaged over all parameters, for every piece

Figure 3 shows the overall ratings across the musical pieces. The results found above are confirmed particularly clear in case of piece 2 "Wind instruments", piece 5 "Percussion" and piece 7 "Applause". Piece 6 "Organ" is a non-critical test sequence. An exceptional rating is found for piece 4 "Brass instruments", here the 7.0 Surround mix was rated less positive than the WFS mix.

## Conclusion and outlook

The results of the listening test of the different techniques are at close quarters. One possible reason could be found in the time constraints given during the test run. When asked after the test many participants complained about the shortness of the examples but at the same time proclaimed that 45 minutes were too long to listen carefully and to keep focus to the many different aspects

inquired in the questionnaire.

Additional tests with improved test design are recommended, providing shorter test sittings, longer test sequences and a reduced number of questions per test sitting. Furthermore, future investigations should concentrate on performance characteristics of the basic spatial audio systems, i.e. 7.0 Surround, WFS and HOA. Exclusion of 2.0 Stereo seems to be advantageous with respect to the time budget and to comparability.