Perceptual Evaluation of Focused Sources in a Concert Hall

Sebastià V. Amengual Garí, Malte Kob
Hochschule für Musik Detmold, 32756 Detmold, Deutschland, Email: amengual@hfm-detmold.de, kob@hfm-detmold.de

Introduction

Implementations of Wave Field Synthesis (WFS) in non-anechoic conditions - e.g. concert environments - for audio and music reproduction, are still in an early stage compared to the use of these systems for research purposes in ideal, anechoic conditions. Several studies about the perception of WFS in anechoic conditions have been published [1][2][3]. However, few studies about perceptions in concert halls or non-anechoic conditions have been published [5], and thus it is unknown how realistic the reconstructed sound fields in these conditions are.

This paper presents a pilot listening test on the perception of focused sources in the Detmold concert hall, which is equipped with a WFS system as arranged by Iosono GmbH. The goal of the listening test was to compare a real sound source (loudspeaker) versus a virtual sound source (focused source as emulated by the WFS system).

The test persons were trained listeners who were familiar with the specific room and setup; their task was to rate their perception of the two sound sources regarding size, localization precision, level, distance, sound coloration and clarity.

The goal of this study is to provide musicians and composers with knowledge on general perceptual differences between the mentioned reproduction methods and to provide clues and empirical recipes to improve the performance of focused sources in the Detmold Concert Hall.

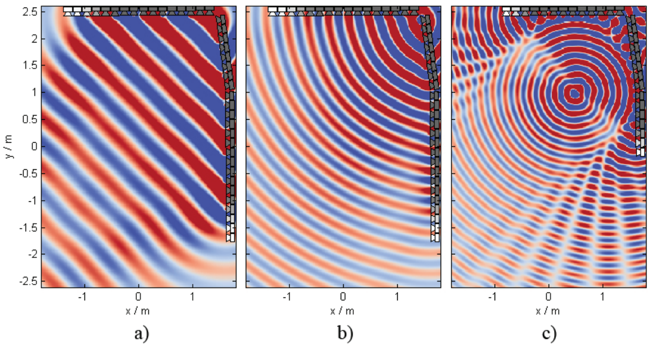

Figure 1: Wave field synthesis simulations performed with the Sound Field Synthesis Toolbox [8]. The loudspeaker set-up is the one present in the WFS Studio of the HfM Detmold. Subfigure a) shows a plane wave, b) a point source and c) a focused source

Figure 2: Image of the Detmold Konzerthaus. The linear MAPs array used by the WFS system is situated sorrounding the hall and can be seen on stage at middle height.

Wave Field Synthesis

Wave Field Synthesis (WFS) is an implementation of Sound Field Synthesis methods which aims at the reconstruction of arbitrary sound fields using loudspeaker arrays specifically driven with different amplitudes and phase values for every secondary source. The idea behind this implementation is based on Huygens-Fresnel principle [4], and modeled using the Kirchoff-Helmholtz integral [6].

A problem arises when spatial artifacts in the sound field appear due to the discrete nature of the loudspeakers and the distances between them [7]. To achieve a good reconstruction of a sound field at high frequencies, it is necessary to have a narrow spacing between loudspeakers [6].

The reconstructed sound fields are based mainly on the combination of two different types of synthesized sources - i.e. plane waves and spherical waves. Both types of sound sources can be synthesized either outside or inside the room. A spherical wave synthesized inside the room is usually called a focused source. Simulations of the mentioned synthesized sources are presented in Fig. 1.

Method

Set-up & listening conditions

The experiments carried out consisted of a set of sessions of listening tests using two different configurations:

- First experiment: The sources (both loudspeaker and focused source) were located on the left side of the stage while the listener was positioned in the middle of the audience area. The acoustic conditions of the room were not modified during the listening tests (see Fig. 3a).
- Second experiment: The sources were placed at the center of the stage and the listener position remained unchanged. In this case, the acoustic conditions

of the room were modified between the presentation of consecutive samples in the listening test (see Fig. 3b).

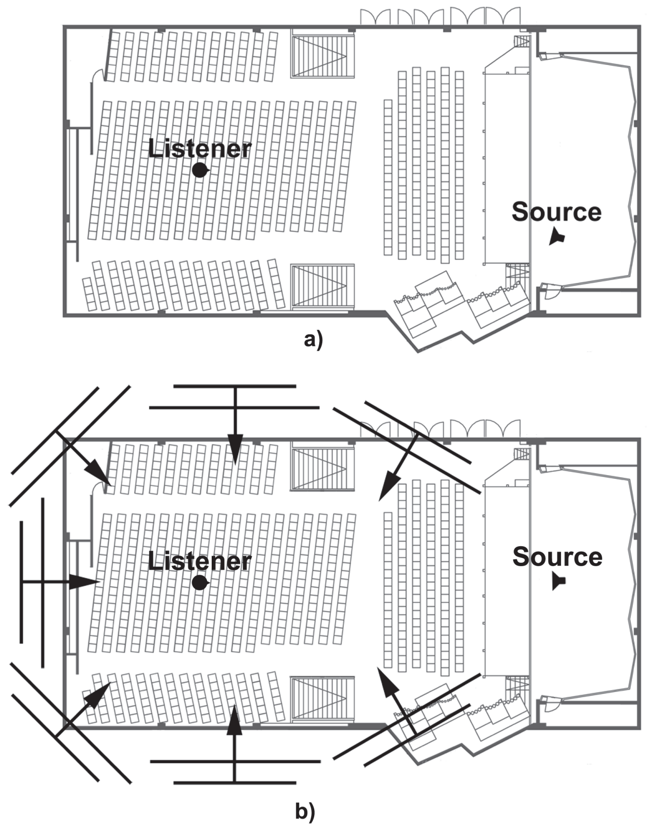

Figure 3: Schematic view of the acoustic scenes implemented for the listening tests. Subfigure a) shows the source and listener position for the first experiment, b) shows the positions of source and listener as well as the plane waves used to increase the reverberation time of the hall during the second experiment.

The point source used in all the experiments was a studio monitor Neumann KH 120A pointing towards the listener. The wave field synthesis system used is composed by a surrounding set-up of Multi Actuator Panel Loudspeakers (MAPs) [9] (see Fig. 2). In order to not interfere on the synthesized sound field, the loudspeaker was positioned 70 centimeters below the WFS array.

In the second, a set of 7 plane waves were used to increase the reverberation time of some of the presented samples following a similar approach as the one described by De Vries in [10]. The signals reproduced by plane waves were generated by convolving the presented audio samples with a set of measured impulse responses. The convolution was performed using the software Altiverb and, in order to adequate the impulse responses, all of them were modified removing the direct sound and part of the early reflections. Since only 3 impulse responses were used to generate 7 plane waves, all of them were slightly modified in order to decrease their correlation. Although correlated impulse responses could lead to comb filter effects [11], informal listening tests were done to ensure the absence of undesired artifacts. A schematic view of the signal flow implemented for the experiment is depicted in Fig. 4. In order to compensate the processing delay, the signal routed to the point source was appropriately delayed. The reverberation time of the different configurations was measured using both loudspeakers and focused sources as excitation sources (see Fig. 5).

The global level of the sources (and plane waves) has been calibrated using white noise to have the same Sound Pressure Level at the listening position.

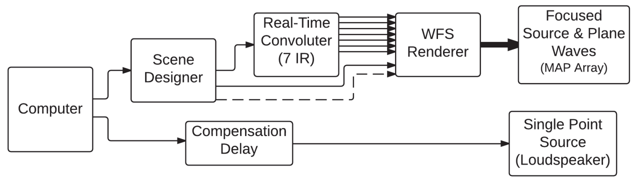

Figure 4: Implemented signal flow for simultaneous reproduction of single sources using increased reverberation. Thin continous lines, thick continuous lines, and dashed lines represent single audio signals, multiple audio signals, and control messages respectively

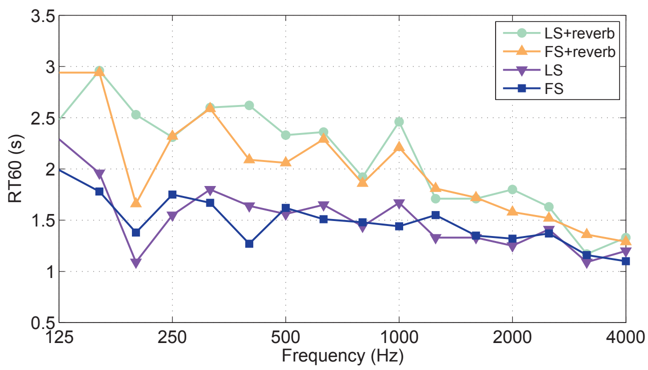

Figure 5: Measured RT60 of the different listening scenes implemented in the second experiment.

Participants

All the participants in the first experiment are expert listeners with extensive musical knowledge (Tonmeister and Music Acoustics students). None of the participants reported hearing problems. The group of participants was composed of 6 students (5 males, 1 female). Every listening test was performed in different individual sessions for every participant.

The second experiment was performed in only one session with a group of 4 participants distributed in a small area around the depicted listener position (see Fig. 3).

None of the participants were tested blind. This decision was made considering the impossibility to visually detect the position of a virtual source.

Stimuli

The stimuli used during the test were three anechoic audio samples:

- Violin: excerpt of a solo violin *partita* of 19 seconds of duration.

- Claps: sequence of claps of 4 seconds of duration.
- Speech: male speech sample in English of 5 seconds of duration.

Presentation of the stimuli
The stimuli were presented in pairs for direct comparison between samples.

In the second experiment, three different methods were used to generate the pairs: single loudspeaker source, focused source, and equalized focused source.

The presented sample pairs were evaluated including redundant combinations - e.g. A-B & B-A - in order to ensure consistency in the ratings. This set-up leads to six different pairs that had to be evaluated in direct comparison. These six pairs were presented in a random order.

In the experiment B, only the violin and speech samples were presented. Since the second listening test was performed after the evaluation of the first results, different equalization was chosen for the focused sources. In this case, the reproduction methods consisted on:

- Loudspeaker (LS)
- Focused source (FS)
- Loudspeaker and increased reverberation (LS+Reverb)
- Focused source and increased reverberation (FS+Reverb)

Rated attributes
Before starting with the evaluation, all the parameters were discussed with the participants in order to ensure uniformity of concepts. Due to the familiarity of the participants with the rated parameters it was not necessary to provide further instruction to the participants, and some technical parameters were used instead of psychoacoustic attributes. The parameters included in the evaluation and the respective questions are:

- Size/Width: Which of the sources is perceived as bigger or wider?
- Position: How different is the position of the sources?
- Distance: Which source is closer?
- Coloration: Is the frequency content of the sources equal?
- Level: Do both sources have the same loudness?
- Reverberation: Which source is more reverberant?
- Clarity/Presence of artifacts: Which of the sources present a clearer sound in terms of artifacts?
- Personal Preference: Which source do you prefer?

In addition to the mentioned parameters, a free personal description of every pair of samples was requested in order to obtain further information that could not be easily reported in the included parameters.

The scale used in the listening test was a discrete bidirectional scale composed by 11 steps using a combination of a same-different task approach together with a relative scale - i.e. the central point of the scale corresponds to equal samples. In some of the rated parameters - i.e. position and coloration - the scale was unidirectional in order to evaluate if the position and coloration of the sounds were the same or different.

Results

The results of the experiments are presented in Fig. 6 and Fig. 7. The first graph (Fig. 6) corresponds to the comparison of a single loudspeaker (LS) and a focused source (FS). The second graph shows only the results for the violin sample presented in different acoustic situations.

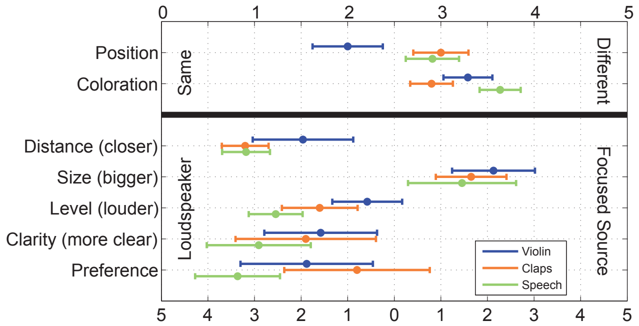

Figure 6: Results of the first experiment (LS vs FS). Points represent the mean of the rated attribute and the error bars represent the standard deviation (N=6).

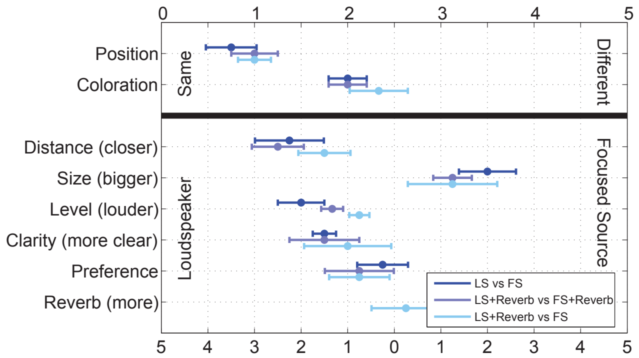

Figure 7: Results of the second experiment. The graph shows the different ratings of the Violin sample depending on the acoustic conditions. Points represent the mean of the rated attribute and the error bars represent the standard deviation (N=4).

Free elicitation
All participants could clearly distinguish loudspeaker and focused sources stimuli. The main information extracted from the free elicitation is that the loudspeaker sounds more "direct, full and warm" than the focused source.

In addition, for the clap stimulus, most of the participants reported difficulties to compare the stimulus due to the impulsive nature of the sound and presence of artifacts. Some participants reported the presence of pre-echoes in the clap sound presented by the focused sources. There was no report of pre-echoes in the other stimuli.

Discussion

As shown in the results, there is a considerable difference on the perception of focused sources and loudspeakers.

Having a focused source created by multiple loudspeakers contributes to a bigger excitation of the room resulting in more reflections and a perception of a bigger and more distant source. Moreover, having a synthesized sound field in non anechoic conditions may result in a non-focused converging point of the virtual sources contributing to the perception of a bigger source as well.

The frequency response of the MAPs loudspeakers and the point source present significant differences, especially in low frequencies in which the MAPs present a lack of low frequencies. This may cause an increase on the difference of coloration perceived when comparing loudspeakers and virtual sources. However, it is proven that equalization can reduce the coloration differences (see Fig. 6 and Fig. 7).

The loudspeaker is perceived as louder and closer in all the cases, possibly due to the non-omnidirectional radiation of the loudspeaker. However, adding extra reverberation to the point source tends to equalize the perceived loudness and distance.

In all the cases the loudspeaker is perceived as clearer and free of artifacts, while the clarity of the focused source strongly depends on the nature of the stimulus (see Fig. 6).

Conclusions and further work

This article presents results of a pilot test on the perception of virtual sources created by a Wave Field Synthesis system installed in the Detmold Concert Hall. The results show that the focused sources are here generally perceived as more distant and bigger. The presence of known problems such as pre-echoes are strongly dependent on the nature of the stimulus used in the rating. However, the acoustics of the room have a direct influence on the perception, and more reverberant situations lead to an increase on the similarity of the perception.

The results presented in this article are to be extended in further listening tests with larger populations and different configurations of sources. Further work includes microphone array measurements replicating the exposed listening conditions in order to analyze the properties of the synthesized sound field in relation to the acoustic characteristics of the room.

Finally, it is important to note that the results of the experiments presented in this article are only applicable to the set-up proposed in this article in the Detmold Concert Hall. Different implementations of the reproductions systems could lead to different results. Using different loudspeaker models with different radiation characteristics and a wider variety of rooms would provide a better understanding of the general behavior of focused sources in concert halls.

Acknowledgements

The authors wish to thank all the participants in the listening tests. Thanks also to Timo Grothe for the comprehensive feedback, Prof. Robert Mores for the recording of the violin excerpt and Dustin Eddy for the grammatical corrections. The research work presented in this article has been funded by the European Commission within the ITN Marie Curie Action project BATWOMAN under the 7th Framework Programme (EC grant agreement no. 605867).

References

[1] Geier, M.; Wierstof, H.; Ahrens, J.; Wechsung, I.; Raake, A.; and Spors, S.: Perceptual evaluation of focused sources in wave field synthesis. 128th Convention of the Audio Engineering Society, Berlin, 2012

[2] Spors, S.; Wierstof, H.; Geier, M.; Ahrens, J.: Physical and perceptual properties of focused sources in wave field synthesis, 127th Convention of the Audio Engineering Society, New York, 2009

[3] Wittek, H.: Perceptual differences between wave field synthesis and stereophony, PhD Thesis, University of Surrey, 2007

[4] Berkhout, A.J.: A Holographic Approach to Acoustic Control, Journal of the Audio Engineering Society 36 (1988), 977–995

[5] Start, E.: Direct sound enhancement by wave field synthesis, PhD Thesis, TU Delft, 1997

[6] Spors, S.; Rabenstein, R.; Ahrens, J.: The theory of wave field synthesis revisited, 124th Convention of the Audio Engineering Society, Amsterdam, 2008

[7] Spors, S.: Spatial sampling artifacts of focused sources in wave field synthesis, NAG-DAGA International Conference on Acoustics, Rotterdam, 2012

[8] Wierstof, H. and Spors, S.: Sound Field Synthesis Toolbox, 132nd Convention of the Audio Engineering Society, Budapest, 2012

[9] van Dorp Schuitman, J.: Wave field synthesis using multi actuator panel loudspeakers: design and application of various digital filtering algorithms, Master's Thesis, TU Delft, 2005

[10] de Vries, D.; Hulsebos, E.: Auralization of room acoustics by wave field synthesis based on array measurements of impulse responses, 12th European Signal Processing Conference, Vienna, 2004

[11] Hulsebos, E.: Auralization using wave field synthesis, PhD Thesis, TU Delft, 2004