Extension of a WFS loudspeaker configuration with a ceiling array for 3D auralization - perceptual aspects

Jasper van Dorp Schuitman[1], Diemer de Vries[2], Eugene Leung[3]

[1] *Delft University of Technology, P.O.B. 5046, 2600 GA Delft, The Netherlands, E-mail: j.vandorpschuitman@tudelft.nl.*

[2] *Delft University of Technology, P.O.B. 5046, 2600 GA Delft, The Netherlands, E-mail: d.devries@tudelft.nl.*

[3] *Delft University of Technology, P.O.B. 5046, 2600 GA Delft, The Netherlands, E-mail: K.Y.E.Leung-1@student.tudelft.nl.*

Introduction

Most operational Wave Field Synthesis [1] systems generate their sound fields through a loudspeaker array configuration in a horizontal plane, roughly matching the ear plane of the listeners. This way, 2D acoustic wave fields can very well be (re)produced. When, however, a 3D environment (e.g., a concert hall [2]) should be auralized, the lack of the vertical dimension leads to serious perceptual complaints.

In a hall occupied with a well-dressed audience, all reflections with non-zero elevation are highly damped, except the first-order ceiling reflections. Therefore, at TU Delft an attempt has been made to solve the problem by extending the rectangular array configuration in the horizontal plane with a ceiling array for reproduction of these first order reflections. Using a speech signal as a source, it was investigated in how far this solution is effective, and how critical it is, from a perceptual point of view, to reproduce the mirror image source which represents a ceiling reflection in a physically correct way.

The WFS setup

Figure 1 shows the current WFS test setup at TU Delft. The setup consists of a linear array of Multi Actuator Panels, or MAPs, all around the listening area. MAPs are flat panel loudspeakers with eight individual controlable actuators each. Two panels form the ceiling array, giving a total of 16 actuators at the ceiling.

Figure 1: The Wave Field Synthesis test setup at the TU Delft, including the ceiling array.

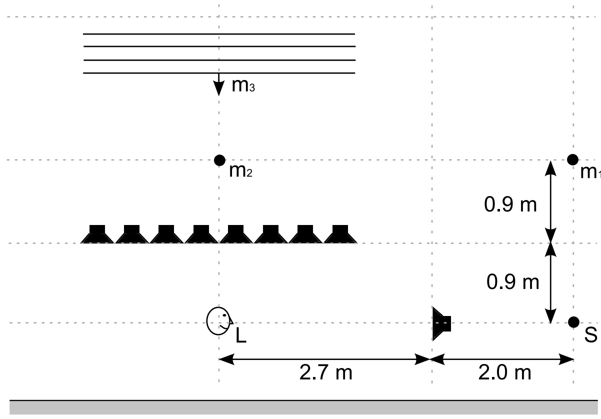

Figure 2: A schematic side view of the test setup, including the positions of the virtual (mirror) sources. Note that the ceiling array consists of 16 loudspeakers in reality.

The test setup

Figure 2 shows the test setup schematically. As can be seen, the primary source S is synthesized at 2 m behind the front loudspeaker array (which consists of 24 loudspeakers). The listener L is placed at 2.7 m distance. The ceiling array is located 0.9 m above the listener. Three mirror source positions are defined:

- m_1: The "true" mirror source position, mirrored in the line of ceiling array.
- m_2: A mirror source at the "true" height, but centrally above the ceiling array.
- m_3: A plane wave traveling perpendicular to the ceiling array.

Mirror sources m_2 and m_3 are included because they can be synthesized much more accurately than m_1 (which lies out of the aperture of the ceiling array). In the listening test, it will be tested if the subjects are able to hear a difference between those positions. The delay for all mirror sources was such, that the reflection arrives 0.98 ms after the direct sound. This is equal to the "true" delay of mirror source m_1.

The listening test

The test was conducted according to a double blind ABX test [3], with 22 subjects. Three parameters were tested:

- Ceiling reflection loudness

- Mirror source position
- Mirror source delay difference

The audio signal used in the test contained 10 seconds of anechoic male speech. It was chosen to accept the hypothesis "The subjects are able to distinguish samples A and B" when the probability of guessing $p_{val} < 0.01$. In this case that means that at least 17 correct answers are necessary: $p_{val}(17 \text{ out of } 22) = 0.0085$.

Ceiling reflection loudness

The relative loudness of the ceiling reflection at the listener's position was calibrated by simulating the wave fields with pink noise as a source, and calculating the RMS value. The second mirror source (m_2) was used in this test. For each loudness value, the subjects heard a sample "A" (direct sound only) and "B" (direct sound + reflection). Figure 3 shows the fraction of correct answers in the ABX test as a function of the reflection loudness.

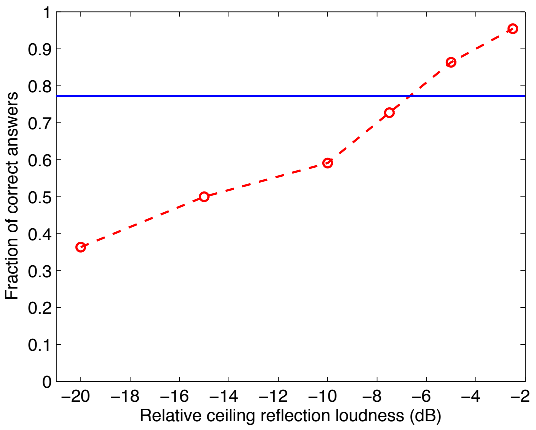

Figure 3: The fraction of correct answers as a function of the relative ceiling reflection loudness. The threshold above which $p_{val} < 0.01$ is drawn as a horizontal blue line.

Mirror source position

To test if the position of the mirror source is of importance, two sets of mirror source positions were included in the test. The delays were the same for all source positions (being equal to the "true" delay of source m_1). Also the direct sound was included in each sample and the relative ceiling reflection loudness was -5 dB. The results are:

- m_1 versus m_2: 17 out of 22 correct answers ($p_{val} = 0.0085$). This is significant since $p_{val} < 0.01$.
- m_3 versus m_2: 15 out of 22 correct answers ($p_{val} = 0.069$). This is not significant since $p_{val} > 0.01$.

Delay difference

In the delay test, the subjects heard the following two samples A and B:

- A: Direct sound + m_2 at -5 dB
- B: Direct sound + m_2 at -5 dB delayed with Δt

The results are shown in figure 4.

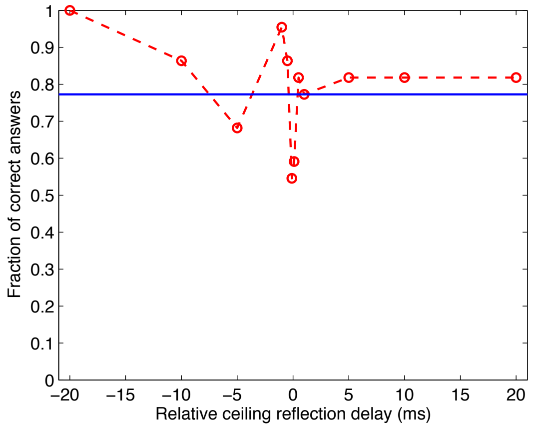

Figure 4: The fraction of correct answers as a function of the relative ceiling reflection delay Δt. The threshold above which $p_{val} < 0.01$ is drawn as a horizontal blue line.

Conclusions

Loudness: From figure 3 the general conclusion can be drawn, that in the case of a speech signal the ceiling reflection should have at least a relative loudness of approx. -7 dB. If the reflection was fully frontal, the predicted threshold would be [4]: $\Delta L \approx -0.6t_0 - 8 = -8.6$ dB, with $t_0 = 0.98$ ms.

Source position: The number of correct answers for "m_1 versus m_2" is statistically relevant, the result for "m_3 versus m_2" is not. In this case a difference in the *direction* of the sound field is noticable, a difference in the shape of the wave front is not (when the subject does not move).

Delay: From figure 4 the rough conclusion can be drawn that a delay difference is noticed for $\Delta t < -7.5$ ms and $\Delta t > 1$ ms. Please note that for $\Delta t < -1$ ms, the reflection will arrive at the listener *before* the direct sound.

For very small delays, the subjects could only distinguish the "normal" reflection from the "delayed" reflection by differences in timbre. Due to wave interferences these differences can be small or large, hence the large variation in the number of correct answers. For larger delays, the subjects could hear the delayed reflection as coming from a separate source.

References

[1] A.J. Berkhout. A holographic approach to acoustic control. *J. Audio Eng. Soc. (Engineering Reports)*, 36:977 – 995, 1988.

[2] E. Hulsebos. *Auralization using Wave Field Synthesis*. PhD thesis, Delft University of Technology, 2004.

[3] Hydrogenaudio. What is a blind abx test? http://www.hydrogenaudio.org, 2003.

[4] H. Kuttruff. *Room acoustics*. Elsevier Science Publishers Ltd, fourth edition, 2000.