# Practical 3 dimensional sound reproduction using Wave Field Synthesis, theory and perceptual validation

E. Corteel[a], L. Rohr[b], X. Falourd[b], K.-V. Nguyen[a] and H. Lissek[b]

[a]sonic emotion labs, 42 bis rue de Lourmel, 75015 Paris, France
[b]Ecole Polytechnique Fédérale de Lausanne, EPFL STI IEL LEMA, 1015 Lausanne, Switzerland
etienne.corteel@sonicemotion.com

Sound field reproduction using Wave Field Synthesis has been so far limited to the positioning of virtual sources and listeners in the horizontal plane only although the underlying formulation (Kirchhoff-Helmholtz) describes the reproduction of 3 dimensional sound fields in a 3 dimensional subspace. However, a strict use of this formulation would require a surface loudspeaker array with an impractical number of loudspeakers. The authors propose here an optimized formulation of Wave Field Synthesis in 3 dimensions that account both for the limitation of localization accuracy of elevated sources and the target listening area size. In contrast to other 3 dimensional sound reproduction techniques such as Higher Order Ambisonics, the proposed approach allows for irregular and incomplete loudspeaker layouts for targeting specific areas for virtual positioning and accounting for practical limitations in loudspeaker positioning. The paper also proposes a subjective evaluation of the proposed approach in an extended listening area. The experiment relies on elevated physical sources (loudspeakers) to be matched in localization with virtual sources reproduced with the proposed approach with a 24 channels loudspeaker array that covers the frontal quarter of the upper half of a rectangular room.

# 1 Introduction

Wave Field Synthesis (WFS) is a sound field reproduction technique that enables to reproduce correct spatio-temporal properties of target sound sources in an extended listening area [1]. The classical formulation of WFS, often referred to as 2 $\frac{1}{2}$ D WFS [12], considers that virtual sources, loudspeakers and listeners are all located in the same horizontal plane, thus limiting WFS to 2D reproduction.
A 3D formulation of WFS has been proposed in the literature [12, 10]. However, this formulation does not face any practical constraints as the 2 $\frac{1}{2}$ D WFS does. One particular aspect concerns the use of a finite number of loudspeakers. It is often recommended in the literature to employ a loudspeaker spacing of 10 to 20 cm for optimum results for 2 $\frac{1}{2}$ D WFS. Therefore, a classical 2 $\frac{1}{2}$ D WFS installation may comprise tens to hundreds of loudspeakers. Applying the same sampling rule to 3D WFS would require to square the number of loudspeakers. Therefore, it is highly important to propose methods for reducing the number of loudspeakers so as to offer 3D WFS rendering in an extended listen area in practical applications.
In this paper, we propose a general formulation of 3D WFS that applies to arbitrary loudspeaker distributions on an open loudspeaker surface. We propose an optimization technique for improving localization accuracy with 3D WFS that targets reduction of apparent source extension (source "width"). Evaluation results are provided, involving an elevation localization comparison task using individual loudspeakers as targets and 3D WFS, with or without source width control, as pointer. Results are analyzed and discussed against available studies in the literature for real and virtual sound sources.

# 2 Wave Field Synthesis for three dimensional reproduction of sound

In this section, we propose a new approach for 3D reproduction using WFS that is introduced in more details in [9]. In the following, bold letters refer to vectors, $\omega$ is the angular frequency.

**Kirchhoff Helmholtz integral - 3D WFS** Wave Field Synthesis, as a boundary based sound field reproduction technique, relies on approximations of the Kirchhoff-Helmholtz integral [1, 14]. The Kirchhoff-Helmholtz integral based sound field reproduction requires a continuous distribution of both omnidirectional and dipolar secondary sources located on the boundary $\partial V$. The so-called 3D formulation of Wave Field Synthesis [12, 10] realizes a first simplification by selecting only omnidirectional sources. The driving filter $U_{3D}(\mathbf{x_0}, \omega)$ can be expressed as:

$$U_{3D}(\mathbf{x_0}, \omega) = W(\mathbf{x_S}, \mathbf{x_0}) F_{3D}(\omega) e^{-j\frac{\omega}{c}|\mathbf{x_S}-\mathbf{x_0}|)}, \quad (1)$$

where $W(\mathbf{x_S}, \mathbf{x_0})$ is a gain factor, $F_{3D}(\omega)$ is a secondary source location independent filter and the last term corresponds to a delay, expressed in the frequency domain, that depends on the distance between the primary source and the considered secondary source. The proposed formulation is thus very similar to 2 $\frac{1}{2}$ D WFS, except that the filter $F_{3D}(\omega)$ exhibits a 6 dB per octave curve in contrast to the 3 dB per octave curve of the filter $F_{2D}(\omega)$ of 2 $\frac{1}{2}$ D WFS [8].
The authors would like to outline that the formulation of 3D WFS from the literature is only valid for a continuous distribution of omnidirectional sources and cannot be used directly as a practical formulation with a finite number of loudspeakers.

**Use of discrete loudspeakers - spatial sampling** Any practical formulation of WFS must include a step of spatial sampling of the secondary source distribution. In 2 $\frac{1}{2}$ D WFS, this step is simply realized by considering that loudspeakers are regularly spaced and by applying a compensation gain that equals the loudspeaker spacing [15].
We propose here to perform a decomposition of the boundary $\partial V$ into smaller surfaces $\partial V_i$ such that each surface is associated to one loudspeaker only. The equivalent driving filter for loudspeaker $i$ is thus expressed as:

$$U_{3D}(\mathbf{x_i}, \omega) = \frac{S_i}{S} W(\mathbf{x_S}, \mathbf{x_i}) \hat{F}_{3D}(\mathbf{x_i}, \omega) e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x_i}|)}, \quad (2)$$

where $S_i$ is the surface of $\partial V_i$, $S$ is the surface of $\partial V$, and $\hat{F}_{3D}(x_i, \omega)$ is a modified version of the filter $F_{3D}(\omega)$ to account for the spatial sampling. The exact definitions of surface calculation and of the modified filter $\hat{F}_{3D}$ are beyond the scope of this paper. The decomposition of the surface into smaller surfaces that are attached to a given loudspeaker may be done using triangulation methods for arbitrary surfaces or using simple sampling rules for regular loudspeakers setups and simple shapes (sphere, shoe box, ...).
The effect of spatial sampling on perceived sound quality has been already addressed in 2 $\frac{1}{2}$ D WFS. Spatial sampling creates physical inaccuracies in the synthesized sound field that may lead to perceptual artifacts such as localization bias [15, 11], increase of source width [13], sound coloration for fixed

[16] and moving listeners [5]. The audibility of these artifacts for a given loudspeaker configuration mostly depends on the frequency content of the sound material [11, 13, 5].

# 3 Reduction of loudspeaker number

**Sampling strategy** Localization in humans is known to be very different for sources located in the horizontal plane or in elevation [3]. Therefore, we propose to take into account this limitation by using a higher density of loudspeakers in the horizontal plane rather than for elevated positions.

**Reducing loudspeaker surface** The total number of loudspeakers can be further reduced by limiting the size of the loudspeaker surface. Such incomplete loudspeaker arrays are often used in 2 $\frac{1}{2}$ D WFS (finite-length linear arrays, U-shaped, ...). There are two main consequences to such a reduction:

- diffraction artifacts occur but are known to cause limited perceptual artifacts [15],
- the positioning of virtual sources has to be limited such that they remain visible within an extended listening area through the opening of the limited loudspeaker array. The corresponding source visibility area can be easily defined using simple geometric criteria [6].

It is therefore possible to limit the size of the loudspeaker array for 3D WFS in a similar way by considering an open surface that may span the locations in which it is physically possible to put loudspeakers in the installation. The loudspeaker surface can be further defined by considering the subspace where virtual source positioning is required, according to the application.

In most applications, it is not possible to put loudspeakers at low elevations because they are either masked by other people in the audience or because it is simply not possible to do so. Therefore, we mostly focus on loudspeaker distributions that target the reproduction of virtual sources above and around the listener. This is not a limitation of the proposed method, but rather a practical choice for reducing the number of required loudspeakers.

**Reduction of spatial sampling artifacts** Various methods for the reduction of spatial sampling artifacts have been proposed in the literature using either the so-called spatial bandwidth reduction [15], partial de-correlation of loudspeakers at high frequencies [7], stereophonic reproduction at high frequencies [16], or reducing the number of active speakers for increasing the spatial aliasing frequency in a preferred listening area [8]. All these techniques have been defined for horizontal reproduction only.

We propose here to extend to 3D WFS the technique proposed by Corteel *et al.* in [8] for 2 $\frac{1}{2}$ D WFS. A simple modified loudspeaker driving filter $\widehat{U}_{3D}$ can be expressed as:

$$U_{3D}(s_w, \mathbf{x_S}, \mathbf{x_i}, \omega) = \frac{S_i}{S} W(s_w, \mathbf{x_S}, \mathbf{x_i}) \times \hat{F}_{3D}(s_w, \mathbf{x_S}, \mathbf{x_i}, \omega) e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x_i}|)}. \tag{3}$$

In this simple formulation, we consider that the origin of the coordinate system corresponds to a reference listening position located within the preferred listening area. The parameter $s_w$ can be used to control the preferred area size around the reference position. We propose here to call this parameter "source width control", since this parameter affects source width as will be seen in the following experiments.

High values of $s_w$ tend to use all loudspeakers of the original 3D WFS driving function of equation . This setting is referred to as "Large" width in the experimental part. Lower values of this parameter can be used for concentrating the rendering on a lower number of loudspeakers located around the direction of the virtual sound source. This setting is referred to as "Small" width in the experimental part.

# 4 Experimental setup

**WFS system** In order to validate the proposed method, we ran a vertical localization experiment. To do so, an experimental WFS setup was mounted in a listening chamber at EPFL. The mean reverberation time of the room was measured to be 0.2 $s$, which is similar to studio conditions.

The WFS system was composed of 24 ELAC 301.2 loudspeakers, which were distributed over two horizontal rows (9 and 7 loudspeakers at heights 0 $m$ and 1.20 $m$ respectively relative to the position of a listener's head (blue and green rows on figure 1) and a ceiling over which the remaining eight loudspeakers were distributed in two other rows (olive and yellow rows on figure 1). The loudspeaker setup therefore covered an azimuthal range of roughly 90° ($-45° \leq \theta \leq 45°$) and an elevation range of 90° ($0° \leq \phi \leq 90°$) in front of the listener ($(\theta, \phi, r)$ being spherical coordinates).

The proposed method was implemented on a Wave 1 3D sound processor[1], which delivered the loudspeaker driving signals to 4 sonic emotion M3S amplifiers through a RME ADI-648 MADI to ADAT converter. All software components, commands and stimuli were generated with MATLAB® on a PC connected to a MOTU HD-896 soundcard.

The set of possible virtual sources was located at a constant distance of 5.4 $m$ and could be controlled in elevation with a precision of ∼1.5°. In this study, we consider that all sources are located on the median plane, at an azimuth of 0°.

Since the implemented method allows different source width parameters, we chose to use two settings: In the first setting, there was no restriction in source width ("large" width), resulting in spatially broad virtual sources, whereas in the second setting ("small" width), spatially precise rendering was targeted.

**Reporting** Since visual or motional reporting of perceived location is subject to sensory bias, we made use of an auditory pointer as employed by Bertet *et al.* [2]. The task of the participant therefore consisted in matching the perceived location of a pointer source (rendered with 3D WFS) with the perceived location of a target source (physical reference loudspeaker). Eight additional loudspeakers, not contributing to the WFS were mounted on the setup to serve as potential targets (grey on figure 1). However, all WFS loudspeakers could be commanded separately and serve as target

[1] http://www.sonicemotion.com

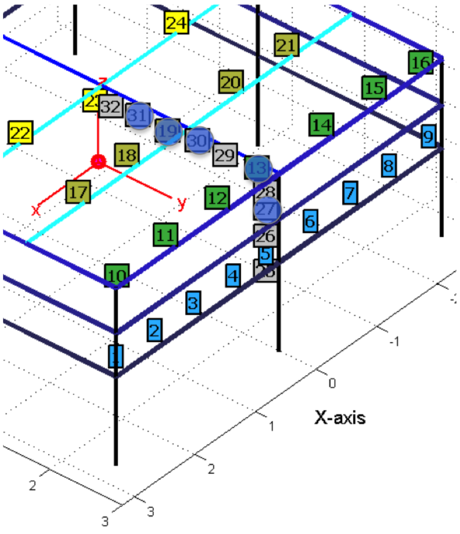

Figure 1: Loudspeaker setup at EPFL - Colored squares are loudspeaker positions whereas lines are aluminium tubes of the rack stand. The red spot represents the position of a participant's head at a centered position.

as well. To avoid edge effects (*i.e.* bias in the rendered location), we chose to test 5 central loudspeaker positions as targets, defined by their elevation: $\phi_{target} = \{14°, 26°, 36°, 43°, 58°\}$ corresponding to loudspeaker numbers {27, 13, 30, 19, 31} on figure 1. Two of the target sources therefore were part of the WFS system (numbers below 25) and the three others weren't.

The pointer source could be moved in elevation with the arrow keys of a computer keyboard by increments of 1.67° between $\phi = 0°$ and $\phi = 90°$.

**Stimuli** Amplitude-modulated pink noise was used as stimuli both for target and pointer sources. By employing broadband noise, we wanted to provide maximum binaural cues to the participant to minimize confusion in localization. The target signal was modulated at $f_{mod,target} = 15\ Hz$ whereas the pointer signal was modulated at $f_{mod,pointer} = 20\ Hz$. The amplitude modulation depth was $d_{mod} = 50\%$ in both cases. In order to minimize the influence of timbre during the matching task, in addition to equalizing the loudspeakers, the target signal was high-pass filtered using a second-order Butterworth filter with $f_{3dB} = 500\ Hz$. The two stimuli therefore could be easily distinguished and the participants could not rely on timbre to match the locations. To avoid any additional bias, the two stimuli were adjusted to present equal loudness.

**Participants** 11 participants, 2 women and 9 men between ages 22 and 38, took part in the study. They all reported normal hearing but no audiometric measurement was made.

**Protocol** The loudspeakers setup was hidden by acoustically transparent curtains and the participant was placed at one out of two listening positions.

The initial elevation of the pointer source was randomly set for each trial (*i.e.* each target/pointer pair). The participant was free to switch between the target and the pointer sources and had no time limit to fulfill the matching task. He could store the pointer elevation by pressing "Enter" on the keyboard as a confirmation of his estimate.

The participant was instructed to feel free to move his head. At the beginning of the experiment, each participant had to complete at least one training trial to confirm their understanding of the task.

The experiment was split into two parts, differing by the listening position of the participant. In the first part, the listener was seated at the origin of the coordinate system (center of the complete installation), facing the loudspeaker setup. For the second part, the listening position was translated 1 meter to the left, but the listener's orientation was kept constant. Each part was composed of 5 runs. In each run, 10 trials (5 target elevations x 2 source width settings) were presented in random order. The two parts did take place at different times (3-4 months apart). Within each part, there was no break between runs, but the participant was free to pause the experiment once. Each part of the experiment took approximately 30 minutes per participant and the participants needed 25.9 seconds per trial on average to complete the elevation matching task.

# 5 Results

**Reported localization** The pointer source elevation at the end of each trial was recorded. Additionally, the history of pointer source movement over time as well as the history of switching between pointer and target source by the participant was recorded for each trial. Measurements where the participant didn't even listen to the target were discarded. This happened for 3 measurements out of 1100 in the available data set and is due to participants pressing the "Enter" key twice and therefore omitting one trial.

An analysis of variance (ANOVA) of the localization data is conducted to test the data for influences of the following factors: listening position ("centered", "1m to the left"), target source number (27, 13, 30, 19, 31) and source width ("large", "small").

The analysis reveals that the source number (*i.e.* the target source elevation) has a significant effect on the mean reported source elevation ($p < .001$). This means that different elevations were globally reported for different reference loudspeakers, confirming the good functioning of the rendering method. The estimated marginal means and the corresponding 95% confidence intervals are shown on figures 2 and 3 for the centered and the left listening positions respectively.

A significant effect of the listening position is also reported ($p < .05$). This means participants globally set a 2.1° higher elevation when they were seated at the left listening position compared to the centered position. Additionally, there seems to be a significant interaction between the loudspeaker number and the source width parameter ($p < .01$).

To better understand the meaning of these effects, pairwise *t*-tests were made between reference loudspeakers to test if levels are well distinguished one from another. Details are not reported here, but the analysis shows that all 5 levels show significantly different mean estimated elevations in all situations (listening position / source width combination) with 4 exceptions: for the centered listening position and a large source width, the difference between mean estimated elevations of sources 13 and 27 reveals only weak evidence for

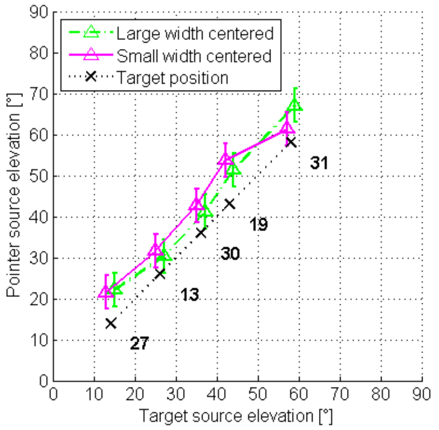

Figure 2: Estimated marginal means and 95% confidence intervals of the estimated elevation data (centered listening position).

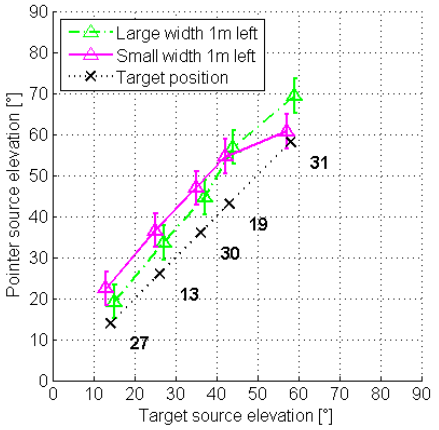

Figure 3: Estimated marginal means and 95% confidence intervals of the estimated elevation data (listening position 1m to the left).

proper separation ($p$ = .058). The same applies to sources 19 and 31 at the centered listening position ($p$ = .098) as well as for sources 19 and 30 at the left listening position ($p$ = .088), both at small source widths. However, mean estimated source elevations did not differ significantly for reference sources 19 and 31 for a small source width at the left listening position ($p$ = .337), leaving only 4 levels for this condition. All of the cited exceptions concern neighboring reference sources.

**Localization blur** Localization blur is traditionally defined as the minimum audible angle (MAA) for a given target source position. A classical estimate is given by the standard deviation for each participant / target / width combination. However, we compute here the half inter-quartile range (HIQR) for every combination to take into account the fact that we had not many values per position (only 5 per participant and condition) and that their distribution may not be symmetric since participants could only manipulate the elevation between 0 and 90 degrees.
To evaluate the impact of the source width parameter on the localization blur, we perform an ANOVA on the HIQR data with target source number (27, 13, 30, 19, 31), listening position (“centered”, “1m to the left”) and source width (”large”, ”small”) as factors.
The analysis reveals two significant effects at the .05 level: Firstly, the source width significantly contributes to reducing the localization blur ($p$ < .001). When the source width is set to “small” (mean HIQR: 4.2°), HIQR is 1.9° smaller on average than for a “large” setting (mean HIQR: 6.1°). The second effect is an interaction between the reference source number and the source width ($p$ < .01). A more detailed analysis shows that this is due to a significant effect of the source number factor at the left listening position with a large source width ($p$ < .05). For all other combinations of listening position and source width, the interaction is statistically not significant. It is also worth noting that the listening position does not influence localization blur in a significant manner.

## 6 Discussion

**Observed bias** The first observation that can be made on the results is that the proposed 3D WFS method allows to properly discriminate 4-5 target elevations between 14° and 58°, even for inter-elevation differences as small 7° (between 36° and 43°). This confirms the resolving ability of the method in a first approximation. However, there seems to be a systematic bias between the estimated and the real target positions. On average, median values of the reported virtual source positions are 7° higher than the real target positions for the centered listening position and 9° higher for the left listening position. We must therefore conclude that the WFS system introduces this error, but this could be easily compensated for. The difference in elevation between listening positions can be ignored in practical implementations, because a localization error of 2° when moving over a distance of roughly one fourth of the total system width seems more than reasonable and barely noticeable in practice.

**Comparison with real source localization** Blauert summarized the results of some evaluation of the MAA for vertical localization in free field [3, p. 44]: The localization blur for continuous speech by a unfamiliar person is about 17°, about 9° for continuous speech by a familiar person and about 4° for white noise (all measurements made at 0° elevation). For increasing elevation, the localization blur tends to increase and attains 13° at 90° elevation for familiar continuous speech.
In our study, we did not restrict head movement, which means that the observed localization blur may be optimistic when comparing to measurement without head movement. The results however show that the proposed 3D WFS method seems to offer a good resolution in elevation and the source width parameter allows to further reduce localization blur (4.2° against 6.1°).

**Comparison with virtual source localization** The experiments cited above were all free-field listening experiments. An interesting comparison can also be made with other WFS implementations or more generally with other virtual source synthesis techniques.
De Bruijn [5] studied vertical localization using a visual pointing task, comparing vertical localization accuracy using a

dense WFS vertical array (12.5 *cm* spacing) against phantom source imaging (lower and upper loudspeakers of his WFS array) with speech stimuli for his study. A standard deviation of ~7° is reported when employing the dense WFS array. Phantom source imaging was shown to be little robust for vertical localization, being close to random for small listening distance where loudspeakers appear to be spaced by more than 60 degrees in elevation.

Another technique that may be worth comparing is binaural synthesis. Bronkhorst obtained about 13° of localization blur when presenting virtual sources producing harmonic signals to participants [4].

The data mentioned in these studies report only standard deviation, which corresponds to the $84.1^{th}$ percentile for a normal distribution. The HIQR used in our experiment should therefore under-estimate localization blur compared to standard deviation. However, the reported localization blur values are significantly lower than the ones reported in the literature for virtual source positioning.

# 7 Conclusion

In this paper, we have proposed a practical formulation of 3D WFS using a practical number of loudspeakers. The formulation allows for arbitrary loudspeaker positioning in 3 dimensions distributed over possibly open 3 dimensional loudspeaker surfaces. We have also proposed a source width parameter that allows to control the precision of reproduction using 3D WFS.

The proposed technique is evaluated using a 24 loudspeaker setup in an elevation localization comparison task using individual loudspeakers as targets and 3D WFS, with or without source width control, as pointer. It is shown that 3D WFS enables to accurately discriminate 4-5 different elevation levels between 14° and 58° with a spacing as small as 7° even for listening positions that are not in the center of the system. It is also shown that the localization blur in elevation can be reduced using the source width control parameter leading to similar results to free-field listening and WFS with very dense vertical loudspeaker arrays.

# Acknowledgements

This work was supported by the EUREKA Eurostars i3Dmusic project [2]. It is also cofunded by the Federal Office for Professional Education and Technology and the European Union.

# References

[1] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. Journal of the Acoustical Society of America, 93:2764-2778, 1993.

[2] S. Bertet, J. Daniel, E. Parizet, L. Gros, and O. Warusfel. Investigation of the perceived spatial resolution of higher order ambisonics sound fields: a subjective evaluation involving virtual and real 3d microphones. In 30th International Conference of the Audio Engineering Society, Saariselka, Finland, March 2007.

[3] J. Blauert. Spatial Hearing, The Psychophysics of Human Sound Localization. MIT Press, 1999.

[4] A. W. Bronkhorst. Localization of real and virtual sound sources. Journal of the Acoustical Society Of America, 98(5):2542-2553, November 1995.

[5] W. de Bruijn. Application of Wave Field Synthesis in Videoconferencing. PhD thesis, TU Delft, Delft, the Netherlands, 2004.

[6] E. Corteel. Equalization in extended area using multichannel inversion and wave field synthesis. Journal of the Audio Engineering Society, 54(12), December 2006.

[7] E. Corteel, K.-V. NGuyen, O. Warusfel, T. Caulkins, and R. S. Pellegrini. Objective and subjective comparison of electrodynamic and map loudspeakers for wave field synthesis. In 30th conference of the Audio Engineering Society, 2007.

[8] E. Corteel, R. S. Pellegrini, and C. Kuhn-Rahloff. Wave field synthesis with increased aliasing frequency. In 124th conference of the Audio Engineering Society, 2008.

[9] E. Corteel, L. Rohr, X. Falourd, K. NGuyen, and H. Lissek. A practical formulation of 3 dimensional sound reproduction using wave field synthesis. In International Conference on Spatial Audio, 2011.

[10] M. Naoe, T. Kimura, Y. Yamakata, and M. Katsumoto. Performance evaluation of 3d sound field reproduction system using a few loudspeakers and wave field synthesis. In Second International Symposium on Universal Communication, 2008.

[11] J. Sanson, E. Corteel, and O. Warusfel. Objective and subjective analysis of localization accuracy in wave field synthesis. In AES 124th Convention, Amsterdam, The Netherlands, May 2008. Audio Engineering Society.

[12] S. Spors, R. Rabenstein, and J. Ahrens. The theory of wave field synthesis revisited. In 124th conference of the Audio Engineering Society, 2008.

[13] E. W. Start. Direct Sound Enhancement by Wave Field Synthesis. PhD thesis, TU Delft, Delft, Pays Bas, 1997.

[14] P. Vogel. Application of Wave Field Synthesis in room acoustics. PhD thesis, TU Delft, Delft, Pays Bas, 1993.

[15] E. N. G. Verheijen. Sound Reproduction by Wave Field Synthesis. PhD thesis, TU Delft, Delft, Pays Bas, 1997.

[16] H. Wittek, F. Rumsey, and G. Theile. Perceptual enhancement of wave field synthesis by stereophonic means. Journal of the Audio Engineering Society, 55:723-751, 2007.

[2] http://i3dmusic.audionamix.com/