Vertical Localization Performance in a Practical 3-D WFS Formulation

LUKAS ROHR,[1] *AES Student Member*, **ETIENNE CORTEEL,**[2] *AES Member*, **KHOA-VAN NGUYEN**[2], **AND**
(lukas.rohr@epfl.ch) (etienne.corteel@sonicemotion.com) (van.nguyen@sonicemotion.com)
HERVÉ LISSEK,[1] *AES Member*
(herve.lissek@epfl.ch)

[1] *Laboratoire d'Electromagnétisme et d'Acoustique, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*
[2] *Sonic Emotion Labs, 42 bis rue de Lourmel, 75015 Paris, France*

Vertical localization performance in a practical wave field synthesis formulation is investigated. The implemented 3-D rendering method allows precise sound source reproduction while taking into account practical constraints such as the required number of loudspeakers, arbitrary open loudspeaker surfaces, and required localization accuracy. A vertical localization experiment is carried out on an experimental system. Estimated source elevation is reported during an elevation matching task using an auditory pointer. Vertical localization accuracy is shown to be good with five elevation levels being discriminated. Localization precision remains as good as 6° – 9° with only 24 loudspeakers contributing to the wave field synthesis system covering the frontal quarter of the upper half sphere of the listening space. The response time is used as an additional performance index that further supports the localization results.

0 INTRODUCTION

This article presents an experimental study that investigates sound source localization performance in a virtual audio environment rendered with Wave Field Synthesis (WFS).

Virtual audio environments are generated through different techniques that generally aim at reproducing a target sound field at the listener's ears as accurately as possible. Well known techniques include WFS [1], Ambisonics [2], and vector base amplitude panning (VBAP) [3] as well as other derived sound field control techniques (see, e.g., [4] or [5]). Every method presents its advantages and drawbacks in terms of localization accuracy of the reproduced sources, bandwidth, listening area, number of loudspeakers, etc.

The study presented here investigates the performance of an innovative WFS implementation in terms of a listener's ability to localize sound sources in the median plane. Performance is measured by means of localization accuracy, localization precision, and response time for two different seating positions.

0.1 3-D Wave Field Synthesis

WFS is a sound field reproduction technique that enables the accurate reproduction of 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 [6], considers that virtual sources, loudspeakers, and listeners are all located in the same horizontal plane, thus limiting WFS to 2-D reproduction.

While a 3-D formulation of WFS has been proposed in the literature [6,7], it does not account for any practical constraints as the $2\ \frac{1}{2}$ -D WFS does. Usual $2\ \frac{1}{2}$ -D implementations use a loudspeaker spacing of 10 to 20 cm, which implies thousands of loudspeakers when extended to 3-D.

In the following, bold letters refer to vectors and ω is the angular frequency.

0.1.1 Kirchhoff Helmholtz Integral

Wave Field Synthesis, as a boundary-based sound field reproduction technique, is based upon approximations of the Kirchhoff-Helmholtz integral [1,8]. The Kirchhoff-Helmholtz integral provides a direct solution for reproducing arbitrary sound fields in a source free subspace V such that the pressure $P(\mathbf{x})$ at any point $\mathbf{x}$ of V can be expressed as:

$$P(\mathbf{x},\omega) = -\oint_{\partial V} P(\mathbf{x_0},\omega)\frac{\partial G(\mathbf{x}|\mathbf{x_0},\omega)}{\partial \mathbf{n}} - G(\mathbf{x}|\mathbf{x_0},\omega)\frac{\partial P(\mathbf{x_0},\omega)}{\partial \mathbf{n}} dS_0, \tag{1}$$

where $P(\mathbf{x_0},\omega)$ is the acoustic pressure at the boundary ∂V, the complementary subspace of Ω_R, on $\partial\Omega$; $\mathbf{n}$ is the inward normal vector to ∂V, and G is the free field Green's function in three dimensions:

$$G(\mathbf{x}|\mathbf{x_0},\omega) = \frac{e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x_0}|}}{4\pi|\mathbf{x}-\mathbf{x_0}|}. \tag{2}$$

According to Eq. (1), Kirchhoff-Helmholtz integral based sound field reproduction requires a continuous distribution of both omnidirectional and dipolar secondary sources located on the boundary ∂V. The so-called 3-D formulation of Wave Field Synthesis [6,7] realizes a first simplification by selecting only omnidirectional sources:

$$P(\mathbf{x},\omega) \approx -2\oint_{\partial V} a(\mathbf{x_S},\mathbf{x_0})G(\mathbf{x}|\mathbf{x_0},\omega)\frac{\partial P(\mathbf{x_0},\omega)}{\partial \mathbf{n}}dS_0, \tag{3}$$

where $a(\mathbf{x_0})$ is a rectangular windowing function that selects only a subset of omnidirectional sources. Spors et al. specify that ∂V must be convex to prevent the unwanted components from re-entering the reproduction volume V [6].

In Wave Field Synthesis, the target sound field is often described as emitted by a so-called primary point source located at $\mathbf{x_0}$. In this case, the windowing function is expressed as:

$$a(\mathbf{x_S},\mathbf{x_0}) = \begin{cases} 1 & \text{if } \langle \mathbf{x_0}-\mathbf{x_S}, \mathbf{n}(\mathbf{x_0})\rangle > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

The driving function $D_{3D}(\mathbf{x_0},\omega)$ of the omnidirectional secondary sound source located at $\mathbf{x_s}$ is thus given as:

$$D_{3D}(\mathbf{x_0},\omega) = -2a(\mathbf{x_S},\mathbf{x_0})\frac{(\mathbf{x_0}-\mathbf{x_S})^T\mathbf{n}(\mathbf{x_0})}{4\pi|\mathbf{x_0}-\mathbf{x_S}|^2} \times \left(\frac{1}{|\mathbf{x_0}-\mathbf{x_S}|}+\frac{j\omega}{c}\right)e^{-j\frac{\omega}{c}|\mathbf{x_S}-\mathbf{x_0}|}\hat{S}_{sw}(\omega), \tag{5}$$

where $\hat{S}_{sw}(\omega)$ is the source signal. Assuming the primary source is located in the far field of all secondary sources $(\frac{1}{|\mathbf{x_0}-\mathbf{x_S}}| \ll \frac{j\omega}{c})$ and neglecting the dependency to the source signal, the driving filter $U_{3D}(\mathbf{x_0},\omega)$ can be expressed as:

$$U_{3D}(\mathbf{x_0},\omega) = W(\mathbf{x_S},\mathbf{x_0})F_{3D}(\omega)e^{-j\frac{\omega}{c}|\mathbf{x_S}-\mathbf{x_0}|)}, \tag{6}$$

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 known formulations of WFS for horizontal reproduction, so-called 2 $\frac{1}{2}$ -D WFS, except that the filter $F_{3D}(\omega)$ exhibits a 6 dB per octave slope in contrast to the 3 dB per octave slope of the filter $F_{2D}(\omega)$ used for 2 $\frac{1}{2}$ -D WFS [9].

This approach uses two approximations of the original Kirchhoff-Helmholtz integral that limit the rendering quality of the target sound field. First, the restriction to omnidirectional sources imposes a windowing of the secondary source distribution, thus introducing artifacts due to diffraction. These artifacts affect the sound field in a similar way to 2 $\frac{1}{2}$ -D WFS. Second, the far field approximation used for the derivation of Eq. (6) is valid mostly at high frequencies or if the primary source is located far from all secondary sources.

It is worth mentioning that this formulation of three-dimensional WFS issued from the literature is only valid for a continuous distribution of omnidirectional sources (i.e., the entire surface ∂V acts as a continuum of monopolar sources). This cannot be achieved in real world conditions where sound sources (loudspeakers) are discrete and present in a finite number. A practical formulation that accounts for these constraints is therefore needed. The following sections show approximations that are made to enable practical WFS.

0.1.2 Spatial Sampling

Any practical formulation of WFS in either two or three dimensions 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 in meters [10].

We propose here to perform a decomposition of the boundary ∂V into smaller surfaces ∂V_i such that each surface is associated to one loudspeaker only. The surface integral in Eq. (3) can then be approximated as a finite sum. The equivalent driving filter for loudspeaker i is thus expressed as:

$$U_{3D}(\mathbf{x_i},\omega) = \frac{S_i}{S}W(\mathbf{x_S},\mathbf{x_i})\hat{F}_{3D}(\mathbf{x_i},\omega)e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x_i}|)}, \tag{7}$$

where S_i is the surface of ∂V_i, S is the surface of ∂V, and $\hat{F}_{3D}(x_i,\omega)$ is a modified version of the filter $F_{3D}(\omega)$ accounting for the spatial sampling. Above the so-called spatial aliasing frequency (Nyquist frequency of the spatial sampling process), the loudspeakers are not interacting in the same way as at lower frequencies and the compensation filter should be modified. This is also true for 2 $\frac{1}{2}$ -D WFS [11].

The exact definition of the modified filter $\hat{F}_{3D}$ is 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, etc.). However, the exact calculation is not detailed in this paper.

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 [10,12], increase of source width [13], sound coloration for fixed [14] and moving listeners [15]. The audibility of these artifacts for a given loudspeaker configuration mostly depends on the frequency content of the sound material [12,13, 15]. This paper aims at investigating the audibility of these artifacts in terms of localization performance.

0.2 Simplification Strategies for 3-D WFS

The previous section has introduced general driving filters for 3-D WFS that can be used with arbitrary loudspeaker distributions over a closed surface. The following section now proposes methods that enable the reduction of the number of loudspeakers so as to achieve 3-D WFS in a practical manner. These methods have been the subject of two conference presentations by the authors [16,17].

0.2.1 Sampling Strategy

Methods for 3-D sound reproduction such as Vector Base Amplitude Panning [3] and Higher Order Ambisonics (HOA, [18]) often consider loudspeaker distributions that have similar density over the horizontal and the vertical dimension. In particular, HOA is best reproduced with a spherical distribution of loudspeakers with a regular sampling.

However, the localization capabilities of humans are known to be very different for sources located in the horizontal plane compared with sources located in elevation [19]. Therefore, we propose to account for this limitation by using a higher density of loudspeakers in the horizontal plane than in the vertical plane.

0.2.2 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, etc.). There are two main consequences of such a reduction:

- Diffraction artifacts may occur but are known to cause limited perceptual artifacts [10],
- The positioning of virtual sources has to be limited in such a way 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 [4].

It is therefore possible to limit the size of the loudspeaker array for 3-D 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 practical 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 choice for reducing the number of required loudspeakers.

0.2.3 Reduction of Spatial Sampling Artifacts

Various methods for the reduction of spatial sampling artifacts have been proposed in the literature using either spatial bandwidth reduction [10], partial de-correlation of loudspeakers at high frequencies [20], stereophonic reproduction at high frequencies [14], or reducing the number of active speakers for increasing the spatial aliasing frequency in a preferred listening area [9]. All these techniques have been defined for horizontal reproduction only.

We propose here to extend to 3-D WFS the technique proposed by Corteel et al. in [9] for 2 $\frac{1}{2}$ -D WFS. It targets the improvement of reproduction accuracy in a preferred listening area. A simple modified loudspeaker driving filter $\widehat{U}_{3D}$ can be expressed as:

$$U_{3D}(s_p, \mathbf{x_S}, \mathbf{x_i}, \omega) = \frac{S_i}{S} W(s_p, \mathbf{x_S}, \mathbf{x_i}) \times \hat{F}_{3D}(s_p, \mathbf{x_S}, \mathbf{x_i}, \omega) e^{-j \frac{\omega}{c} |\mathbf{x}-\mathbf{x_i}|)}. \quad (8)$$

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_p can be used to control the size of the preferred listening area around the reference position reducing the number of active loudspeakers as can be seen in Eq. (8) and [9]. We propose here to denote this parameter "spatial precision control," since this parameter affects spatial precision as will be seen in the following experiments. This parameter may be expressed in percentages for practical implementation. It can either be a design choice of the system installer or a parameter offered to the user of the system.

For $s_p = 0\%$, all loudspeakers of the original 3-D WFS driving function in Eq. (7) are used. This setting is referred to as "Low" spatial precision in the experimental part. Higher percentages 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. The "High" spatial precision setting of the experimental part corresponds to a value of $s_p = 70\%$ where a large number of loudspeakers remain active (see Section 1.1).

0.3 Sound Source Localization Evaluation

Sound source localization performance can be measured in different ways. When using an absolute localization protocol (i.e., pointing at the perceived location of a source) localization judgment data can be modeled as normal distribution, as explained in [21]. Two types of performance indices can then be distinguished as for all Gaussian processes: accuracy and precision. Accuracy describes the location of the distribution relative to a reference, corresponding to the constant error component [21]. When applied to localization error data, it corresponds to a localization bias, i.e., a difference between the mean of the distribution and the actual location of the sound source.

Precision, however, describes the spread of the distribution, corresponding to the random error component. Although other quantities may be considered, the standard deviation of the distribution gives a good measure of precision [21].

If a relative localization task is used (i.e., compare the locations of two sources that are presented sequentially based on a forced-choice protocol), one may also consider a threshold based on a psychometric function to measure localization acuity. Blauert, for example, defines localization blur to be the minimum audible angle (MAA). For a given initial source position, this corresponds to the angular distance for which 50% of the participants noticed a change in source location when moving the source away from its initial position [19]. It is however not clear if localization acuity and localization precision are related [22], making a comparison of results difficult.

Response time was also used to measure localization performance in [23]. Participants were asked to turn their nose towards the perceived location of a sound source while the position of their head was tracked. The time until a stable position was reached was measured and analyzed as performance index. The measurement of this cue should therefore allow to gain an additional insight on the localization performance of human listeners in 3-D WFS.

Note that all of the evaluation methods cited above refer to the human ability to locate one or several sound sources. In any situation where localization performance is measured, an audio rendering system that may introduce additional errors (i.e., with its own accuracy and precision) is involved (see, e.g., the limitations found by [24]). When measuring localization performance, the measured error components are therefore always the results of two phenomena: human localization uncertainty and rendering acuity of the audio system. Depending on the experiment (e.g., when using a single point-like sound source), the audio system may play only a minor role and be ignored. In the context of WFS however, the phenomena are not separable and sum up in the measured error components.

In this paper we focus on vertical localization, since the novelty of the employed spatialization technique is to enable 3-D rendering in WFS. Such a system will however never be employed in an environment where all listeners are at fixed positions without moving their head. We did not, therefore, restrict head movement. The participants could rely on a full set of cues, potentially giving them an increased localization accuracy when comparing to studies with a fixed head position (see [19] for a review). Given the localization task and the chosen protocol, localization accuracy and precision as well as the response time are analyzed.

0.3.1 Reporting Method

Sound source localization experiments may be biased depending on the chosen reporting method. Different methods have been applied in the literature. Oral reporting, such as the "absolute judgment" technique used by Wightman and Kistler [25] have the disadvantage of necessitating extensive training. Another solution may be head-tracking and asking the participants to turn their head towards the location at which the sound source is perceived such as employed by Makous and Middlebrooks [26]. This may, however, introduce errors due to the lag that is introduced by headtracking devices, and it may be quite uncomfortable for locations in the median plane. Listeners also have no way of knowing if they are actually pointing to the desired location since no feedback whatsoever is provided. Oldfield and Parker previously had used a special gun to point at perceived locations [27] and photography to record the answers of the participants. Besides being unpractical for the rear quadrant (listeners had to aim through their head), such visual pointing methods have the disadvantage of possible mismatch between the visual and auditory modalities [28].

More recently, Pulkki and Hirvonen used an auditory pointer in the form of a loudspeaker mounted on an arm that could be moved [29]. This method has the advantage of providing immediate feedback and not being multimodal. This method was later extended to virtual audio sources by Bertet et al. who used a virtual source as an auditory pointer that can be controlled by the participant [30]. The task was then to align the pointer location to the perceived location a physical sound source (loudspeaker). Given the advantages of this method and the similarity of the task at hand, the present study uses this reporting method, submitting the participants to a source location matching task with a physical reference loudspeaker used as target (see Section 1.2).

0.4 Objectives

The present study aims at gaining an insight into the spatialization performance of a practical implementation of a 3-D WFS algorithm. Three-dimensional WFS systems provide a mean of placing virtual sources all around and especially above a target listening area. The vertical localization performance in a 3-D WFS setup is investigated by conducting a localization experiment. Based on WFS theory that states that ideally any virtual source position may be synthesized accurately and based on the fact that the proposed simplifications will introduce some degradations while conserving a certain degree of precision and accuracy, we expect that the participants are able to discriminate between several source elevations and that there is no influence of a participant's position on localization performance. Moreover, the chosen rendering algorithm introduces a spatial precision parameter, which should additionally increase localization performance by reducing the number of active loudspeakers as long as several loudspeakers remain active (see Section 1.1 for the choices made). Performance is also measured by means of the response time of the participants, giving an additional cue on the difficulty of the task at hand. It is expected that the more accurately a source location is perceived, the quicker the localization task will be accomplished. Localization precision is also expected to increase when the spatial precision is decreased.

1 METHOD

1.1 WFS System

A WFS setup was installed in a listening room of 6.70 × 6.80 × 2.60 m. The mean reverberation time of the room was measured to be about 0.25 s and flat below 5.3 kHz and decaying for higher frequencies to reach 0.18 s at 16 kHz, which is similar to studio conditions. The background noise level of the room was measured to be approximately 23 dB(A) (1 second integration period, averaged

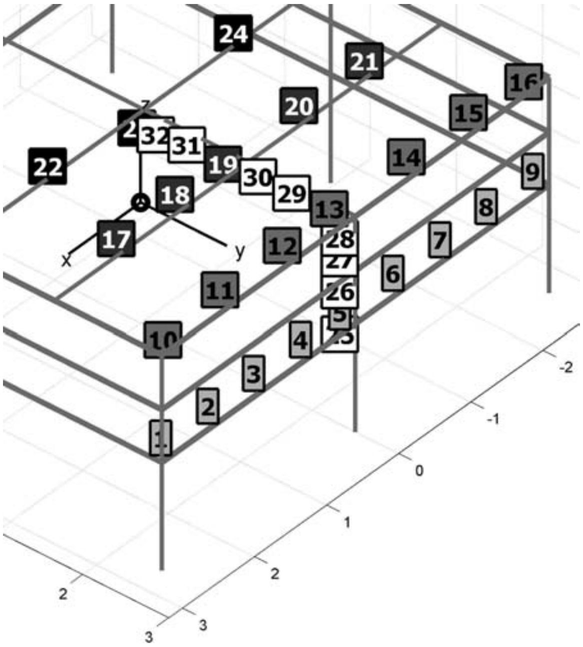


Fig. 1. Loudspeaker setup at EPFL - Squares are loudspeaker positions whereas lines are aluminium tubes of the rack stand. The black spot represents the position of a participant's head at a centered position.

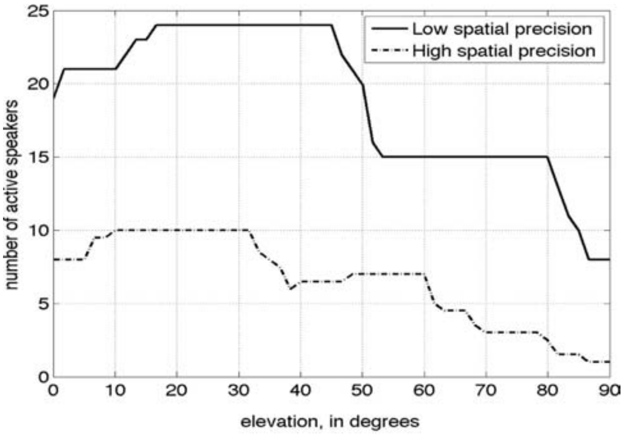


Fig. 2. Number of active speakers depending on target elevation for high and low spatial precision settings. The number of active speakers corresponds to the number of speakers having a driving signal level between that of the loudspeaker receiving the maximum level for a given source position and 15 dB below that.

over 2 × 10 minutes of measurement). Three of the walls are coated with absorbing materials (mineral wool covered with tissue), the floor is entirely covered with carpet, and the ceiling is acoustically treated. The fourth wall contains windows. Even though no measurements for early reflections that potentially influence perceived location [31] were made, the configuration of the room and the covering of the windows with heavy curtains should minimize the influence of the room.

The WFS rendering system was composed of 24 ELAC 301.2 loudspeakers, which were distributed as illustrated on Fig. 1: two horizontal rows of nine and seven loudspeakers at heights 0 m and 1.20 m respectively relative to the position of a listener's head (brightest grey rows) and a ceiling over which the remaining eight loudspeakers were distributed in two other rows (darker grey rows). 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 ((θ, φ, r) being spherical coordinates).

Eight additional loudspeakers, not contributing to the WFS were mounted on the setup to serve as potential targets (white squares on Fig. 1). That said, all WFS loudspeakers could also be used separately and serve as target as well. To avoid any visual influence, the setup was hidden by acoustically transparent curtains.

The 3-D WFS algorithm was implemented on a Sonic Wave 1 3-D sound processor,[1] which delivered the loudspeaker driving signals to four 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 sound processor allowed for a low shelf, a high shelf, and three parametric equalizers on every output (i.e., every loudspeaker). Manual measurement of the output spectrum at the center of the setup for each loudspeaker in combination with these equalizers was used to compensate for room coloration.

The set of possible virtual sources was located at a constant distance of $r = 5.4$ m with respect to the center of the system and could be controlled in elevation with a precision of ∼1.67°. 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 values of the spatial precision parameter, we chose to use two settings: in a first setting, there is no restriction in spatial precision ("low" precision), resulting in spatially broad perceived virtual sources, whereas in the second setting ("high" precision), spatially precise rendering is targeted. The "high" precision setting has been determined considering a preferred listening area of 2 m diameter around the center point of the installation.

Fig. 2 provides the number of "active" speakers depending on the virtual source elevation and spatial precision setting. It can be seen that in the "low" spatial precision setting, 15 or more loudspeakers contribute to the virtual source rendering for nearly all source elevations. The number of active speakers is only related to the source visibility criterion. For the "high" spatial precision setting, the number of active speakers remains large: around 10 between 0 and 35 degrees, around 7 up to 60 degrees, and gradually reducing to 1 around 90 degrees (the "voice of god" speaker). It should be noted that the number of active speakers is given as the number of speakers having a relative level between the level of the loudest speaker at the given source elevation and 15 dB below that. It can be regarded as the number of speakers that significantly contribute to the sound field reproduction.

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

1.2 Experimental Task

Since visual or motional reporting of perceived location is subject to sensory bias, we used an auditory pointer as employed by Bertet et al. [30]. The task of the participant, therefore, consisted in matching the perceived location of a pointer source (rendered with 3-D WFS) with the perceived location of one of the target sources (physical reference loudspeaker). The pointer source could be moved in elevation with the arrow keys of a computer keyboard by increments of 1.67° between $\phi = 0^\circ$ and $\phi = 90^\circ$. 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.

1.3 Stimuli

Amplitude-modulated pink noise was used as stimuli for both target and pointer sources. By employing time-varying broadband noise, we wanted to provide maximum localization cues to the participant to minimize confusion, since localization has been shown to improve with increasing bandwidth (see, e.g., [32]) and different modulation frequencies enable the participant to distinguish between the two stimuli. 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 subjectively adjusted to present equal loudness.

1.4 Experimental Design

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 setup, see Fig. 1), 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 spatial precision settings) were presented in random order. The initial elevation of the pointer source was randomly set for each trial (i.e., each target/pointer pair).

To prevent edge effects (i.e., bias in the perceived location due to the sound field not being rendered completely when the virtual source is on the edge of the valid rendering domain), we chose to test five central loudspeaker positions as targets, defined by their elevation: $\phi_{target} = \{14^\circ, 26^\circ, 36^\circ, 43^\circ, 58^\circ\}$ corresponding to loudspeaker numbers $\{27, 13, 30, 19, 31\}$ on Fig. 1. Two of the target sources therefore were part of the WFS system (numbers below 25) and the three others weren't.

Each 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 understand the task.

The two parts took place at different times (3–4 months apart), but with the same panel of participants. Within each part there was no break between runs, but the participant was free to have a break during 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. Eleven participants, 2 women and 9 men between ages 22 and 38 (M = 28.4, SD = 5.6), took part in the study. The panel was composed of master and Ph.D. students, as well as post-doc researchers at EPFL, including two of the authors. Seven of them had already heard the spatialization system in a different context. None of the participants was compensated in any way for the experiment. They all reported normal hearing although no audiometric measurement was made.

The experimental design in this case was a repeated measures design with four factors: the target elevation (5 levels), the spatial precision setting (2 levels), the seating position (2 levels), and the repetition (5 levels).

2 RESULTS

2.1 Analysis

The pointer source elevation at the end of each trial was recorded. Additionally, the history of pointer source movement over time was recorded for each trial. Three measurements out of 1100 in the available data set were discarded during post-screening because participants pressed the "Enter" key twice and therefore skipped one trial.

In the first part, an analysis of variance (ANOVA) of the localization data is conducted to test the data for influences of the following factors: listening position "pos" ("centered," "1 m to the left"), target source number "refS" (27, 13, 30, 19, 31), spatial precision "prec" ("low," "high"), and repetition "time" (1, 2, 3, 4, 5). The dependent variable is the matched source elevation.

To perform the ANOVA, a mixed model is considered with all factors being modeled as fixed effects. Since the order of presentation of the different runs is random, the covariance matrix is assumed to have a compound symmetry structure. Covariance is therefore assumed being the same between any two measurements and variance being the same for each measurement. A model with a covariance matrix having a heterogeneous compound symmetry structure was also considered. Such a model would allow for different variances for every measurement while keeping the constant covariance assumption. However, it showed no improvement over the previous model. The value of the Akaike information criterion (AIC) was slightly smaller in the second case, but the value of the bayesian information criterion (BIC) was larger. Since the BIC penalizes the estimation of a too large number of parameters, the simpler model with the compound symmetry covariance matrix

Table 1. Statistical analysis of effects on matched source position

Effect	F	p
refS	619.2	<.001
prec	0.3	.571
pos	10.9	<.005
time	2.5	<.05
refS*prec	9.7	<.001
refS*pos	3.0	<.05
prec*pos	0.1	.758
refS*time	1.0	.431
prec*time	0.1	.981
pos*time	1.0	.421

structure was kept. For all pairwise comparisons that are made during the analysis, the Sidak correction is used to account for multiple comparisons. The significance level is set to $\alpha = 0.05$.

Main effects and two-by-two interactions are tested. The results of this analysis are given in Table 1.

2.2 Localization Accuracy

The analysis reveals that the source number (i.e., the target source elevation) "refS" 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. Pairwise comparisons using the Sidak correction to account for multiple comparisons are made between reference loudspeaker levels to test if matched source location levels are well distinguished one from each other in every situation (seating position and spatial precision combination). The differences prove to be statistically significant in all situations. Significance levels are $p < .001$ for all differences, except for the difference between sources 19 and 31 with the "high" spatial precision at the left seating position, where significance level is $p < .05$. Five levels of target elevations between 14° and 58°, even for inter-elevation differences as small as 7° (between 36° and 43°) could therefore be distinguished in any situation. The estimated marginal means and the corresponding 95% confidence intervals are shown on Fig. 3 for the centered and the left listening positions.

A significant effect of the listening position "pos" is also reported ($p < .005$). Participants globally set a 2.0° higher elevation ($p < .005$) when they are seated at the left listening position compared to the centered position. No main effects are reported for the spatial precision "prec" ($p = .571$). The analysis also shows a significant effect of the repetition number factor ("time," $p < .05$). Participants globally report lower matched source locations as the repetition number increases. Pairwise comparisons, however, show that the difference between matched source locations is significant only between repetitions 3 and 5 ($p < .05$). Participants set the virtual source 2.86° lower during the 5th repetition than during the 3rd repetition. All other comparisons are not statistically significant ($p > .15$).

Table 2. Statistical analysis of effects on mean standard deviation of matched source position

Effect	F	p
refS	0.6	.664
prec	20.7	<.001
pos	0.8	.379
refS*prec	3.5	<.01
refS*pos	3.1	<.05
prec*pos	1.7	.191
refS*prec*pos	1.4	.248

The analysis also reveals that two interactions are statistically significant. The first one is between the reference source number and the spatial precision ("refS*prec," $p < .001$). A quick inspection of Fig. 3 shows that the slope of the curves is getting smaller for source 31 when using the high spatial precision, whereas it seems to remain constant when using the low precision. This is confirmed by the fact that there is a significant difference between the matched elevations for target source 31 when using the "high" spatial precision, as compared to using the "low" spatial precision. The difference in matched elevation is 7.7° for that case ($p < .001$). Other pairwise comparisons were not statistically significant ($p > .05$).

The second significant interaction is between reference source number and seating position ("refS*pos," $p < .05$). This would translate into different slopes of the response curves between seating positions if the spatial precision parameter was ignored in Fig. 3. Pairwise comparisons of matched source locations between left and centered seating position were not all significant, which prevents any further comment. All other two-by-two interactions were not statistically significant.

Another fact that may be worth mentioning is that there is a systematic bias in the matched source elevation with respect to the target elevation. The average matched elevation being 8.2° higher than the corresponding reference source. If broken down by seating position, the bias is 7.2° for the centered listening position and 9.2° for the left listening position.

2.3 Localization Precision

Localization precision is given by the standard deviation (SD) for each participant / target / precision combination, which was computed as a new dependent variable. To evaluate the impact of the spatial precision parameter on the localization precision, we perform an ANOVA on this new dependent variable with target source number "refS," listening position "pos," and spatial precision "prec" as factors. Main effects are computed as well as all possible interactions. The parameters of the model remain the same as for the analysis on the reported localization.

The results reported in Table 2 show that there are three significant fixed effects at the .05 level. First, the spatial precision significantly contributes to enhance localization precision ($p < .001$). Estimated marginal means reveal that when the spatial precision is set to "high" (estimated mean

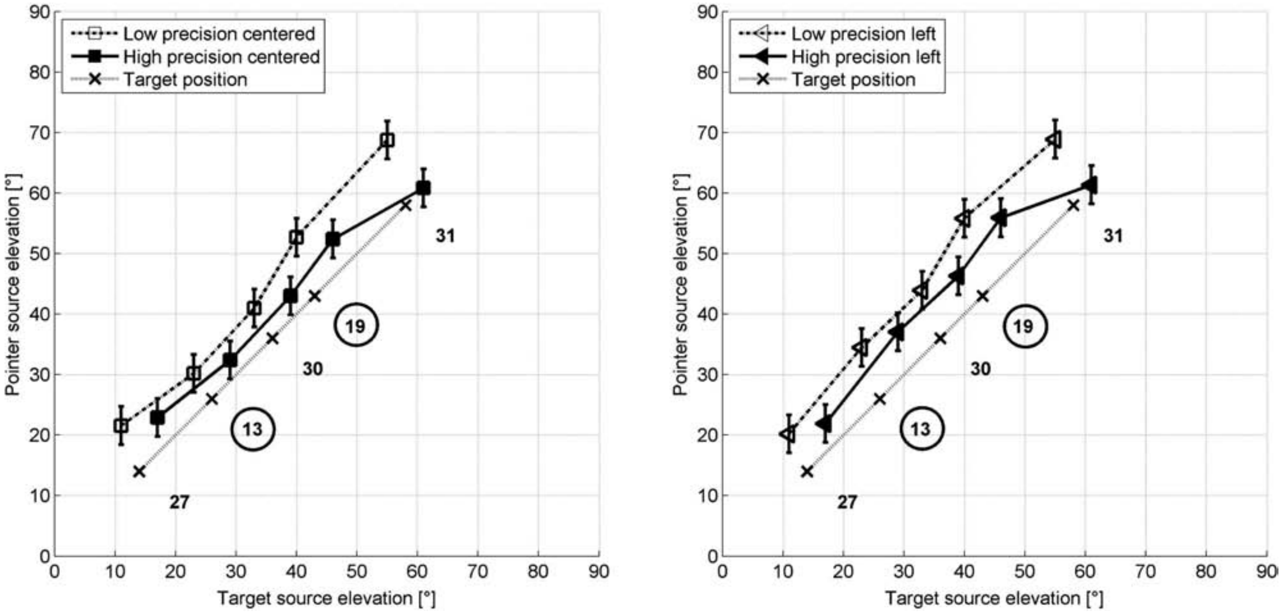


Fig. 3. Localization accuracy: Estimated marginal means and 95% confidence intervals of the matched elevation data at the centered listening position (left) and at the listening position 1 m to the left (right). Circled numbers correspond to loudspeaker positions that are part of the WFS system.

SD: 6.3°), the SD is 2.3° smaller on average than for a "low" setting (estimated mean SD: 8.7°), and the difference between both is significant at the .001 level.

The two other statistically significant effects are two interaction effects: reference source number with spatial precision ("refS*prec," $p < .01$) and reference source number with listening position ("refS*pos," $p < .05$). For the first of these two interaction effects, a comparison by pairs (splitting up the effect for each reference source number and comparing between "low" and "high" spatial precision) reveals that the difference in SD is strong for the highest three reference sources ($p < .05$) and not significant for the lowest two reference sources. This interaction can also be seen on Fig. 4, where the curves following the results of the "low" spatial precision setting globally have a different slope than those who follow the results of the "high" setting.

On the other hand, even though the interaction effect between listening position and reference source number is statistically significant, a comparison by pairs (splitting up the effect for each reference source number and comparing between left and centered listening positions) shows only a significant difference in SD for the highest reference source ($p < .05$). All other pairs do not show statistically significant SD differences.

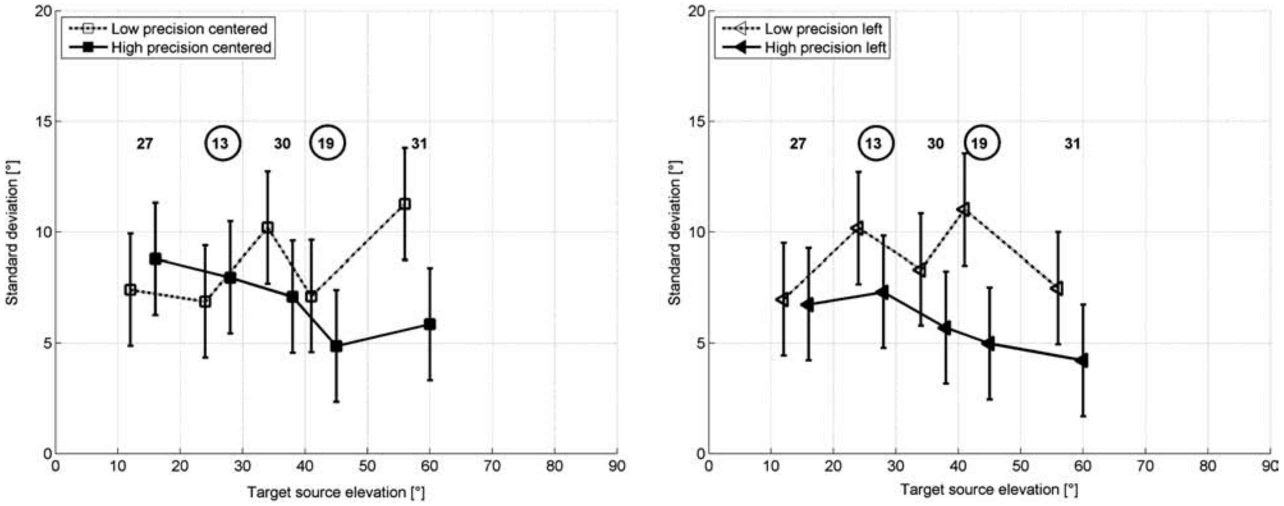


Fig. 4. Localization precision: Estimated marginal means and 95% confidence intervals of the standard deviation data at the centered listening position (left) and the listening position 1 m to the left (right). Circled numbers correspond to loudspeaker positions that are part of the WFS system.

Table 3. Statistical analysis of effects on response time

Effect	F	p
refS	1.2	.289
prec	37.9	<.001
pos	0.8	.367
time	3.3	<.05
refS*prec	3.3	<.05
refS*pos	0.7	.591
prec*pos	0.0	.842
refS*time	1.0	.391
prec*time	0.6	.684
pos*time	1.0	.425

2.4 Response Time

The same analysis that was run on the reported matched locations (Section 2.2) was run on the response time "resp-Time" dependent variable. Table 3 reports the main effects and two-by-two interactions.

Three effects are shown to have a statistically significant influence on the average response time: the spatial precision ("prec," $p < .001$), the repetition ("time," $p < .05$), and the interaction between the reference source number and the spatial precision("refS*prec," $p < .05$).

In this study a low spatial precision setting is potentially detrimental to accurate localization, which is confirmed by the variation of the participants' response times as a function of the spatial precision setting. A pairwise comparison shows that the mean response time decreases by 4.3 s ($p < .001$) from 28.0 s to 23.7 s when using the "high" rather than the "low" spatial precision setting. This is further illustrated by Fig. 5 where estimated marginal means and 95% confidence intervals are shown for both settings. The curves with the high spatial precision setting are globally below the curves with the low spatial precision setting.

The effect of the repetition number is also statistically significant ($p < .05$). The response time therefore globally decreases with the number of repetitions. However, the mean response time difference is significant only between repetition times 1 and 3 ($p < .01$).

The last significant effect is the interaction between the reference source number and the spatial precision ("refS*prec," $p < .05$). This is also illustrated on Fig. 5, where the slopes of the curves between the "low" and the "high" spatial precision differ.

All other effects are not significant. There is, therefore, no influence of the reference source number ($p = .289$) or of the seating position ($p = .367$) on the mean response time.

3 DISCUSSION

3.1 General Discussion and Localization Accuracy

The first observation that can be made on the results is that the implemented 3-D WFS method allows to properly discriminate five target elevations between 14° and 58°, even for inter-elevation differences as small as 7° (between 36° and 43°). This confirms the spatial resolution of the method in a first approximation. However, there seems to be a systematic bias between the matched and the actual target positions. On average, mean values of the reported virtual source positions are 7.2° higher than the real target positions for the centered listening position and 9.2° higher for the left listening position. This bias cannot be explained in terms of the positions of the loudspeakers that contribute to the WFS array. If reported source locations were biased towards the nearest loudspeaker for example, the bias would disappear for virtual source positions that correspond to the position of a loudspeaker contributing to the WFS. This is, however, not the case for loudspeakers 13 and 19, which are circled on Fig. 3 where the "zero bias" line corresponds to the dotted diagonal line. We must therefore conclude

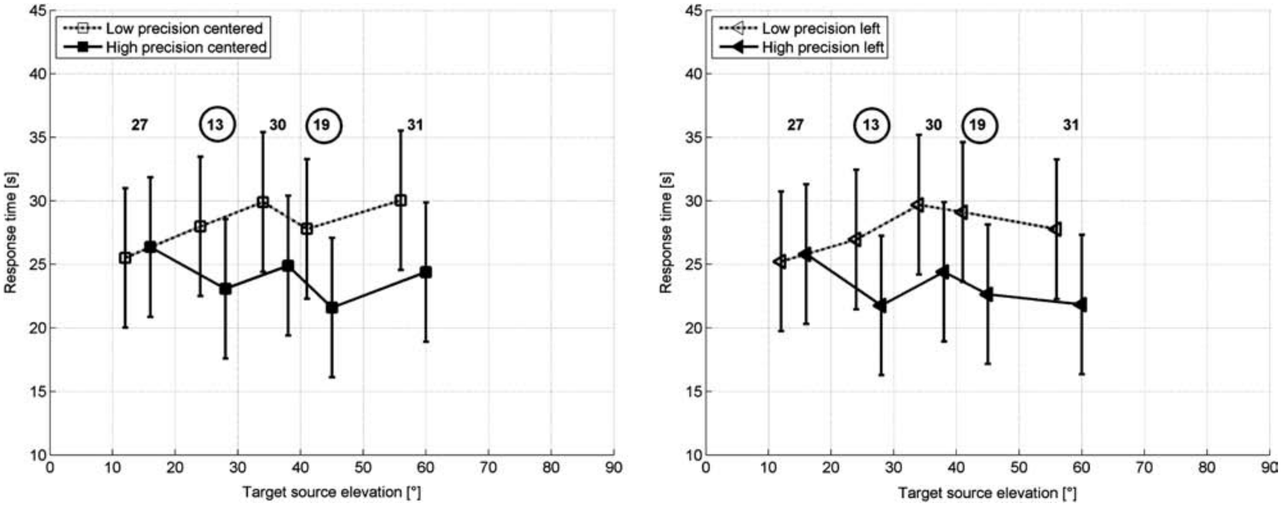


Fig. 5. Response time: Estimated marginal means and 95% confidence intervals of the response time data at the centered listening position (left) and the listening position 1 m to the left (right). Circled numbers correspond to loudspeaker positions that are part of the WFS system.

that the WFS system introduces this constant error, but this could be easily compensated for.

It has to be noted that the five levels of elevation are discriminated at both listening positions. The difference in elevation between listening positions can be ignored in practical implementations, because a localization accuracy shift 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.

A second observation is that the interaction between the reference source number and the spatial precision parameter is always statistically significant. This can be readily explained by the setup geometry and the spatial precision parameter definition. The WFS setup is constructed such that there are fewer loudspeakers on the uppermost layers than for the lower layers. The spatial precision parameter further reduces the number of active loudspeakers. The combination of both results in few loudspeakers that are active when a high elevation is to be rendered with a high spatial precision. The matched source locations should therefore be quite precise and present almost no bias at high elevations, whereas at low elevations, the effect is less present. This is also expected to enhance localization precision and reduce response time. This is measured by said interaction and can be seen across the results. It is noticeable however, that even though only a small number of loudspeakers may be active at high elevations with high spatial precision settings, localization results do not vary significantly even if the virtual source location does not match the positions of a physical loudspeaker of the rendering system (e.g., source #19).

There also seems to be a small learning effect, since the response time and the systematic bias are both slightly reduced with increasing number of repetitions. However, the analysis shows that the performance increase is not important.

3.2 Virtual Sound Source Localization Performance

When comparing the results to other studies, a difference has to be made between free-field localization with physical sound sources and localization of virtual sound sources. For physical sound source localization, former studies report best accuracy in the frontal quadrant. Oldfield and Parker, e.g., report 6° or less azimuthal error in the horizontal plane and 8° or less elevation error in the median plane in the frontal quadrant when using broadband white noise and a manual pointing reporting method (pointing with a gun while blindfolded) [27]. The limitations of a rendering system, however, will influence the resolving ability of human audition and therefore the results of localization experiments. Virtual sound source synthesis can be implemented using different techniques, such as WFS, amplitude panning (vector-based (VBAP) or simple stereo phantom source imaging), Ambisonics, or even binaural synthesis. Vertical localization performance varies depending on the proposed technique and the experimental setup. Moreover, since virtual sound scenes are rarely directly compared to the original sound scene (when it exists), localization accuracy may not be the most relevant performance index. Localization precision on the other hand does not depend on direct comparison of two sound scenes and may therefore be a more meaningful performance index when comparing the presented results with other studies.

De Bruijn [15] studied vertical localization using a visual pointing task, comparing vertical localization accuracy using a dense vertical WFS array (12.5 cm spacing) against phantom source imaging (lower- and uppermost 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 non-robust for vertical localization, results being close to random for small listening distances where loudspeakers appear to be spaced by more than 60 degrees in elevation. We obtain similar degrees of localization precision but with distances between loudspeakers that are much greater (smallest distance is ∼54 cm in the horizontal dimension and ∼105 cm in the vertical dimension).

Chung et al. proposed a combination of WFS and vertical amplitude panning in [33]. Two horizontal WFS arrays were used to generate a third virtual WFS array by vertical amplitude panning, which was intended to generate the targeted sound field. This approach seems interesting since the number of loudspeakers is greatly reduced as it is with the proposed method. However, the results suggest that vertical localization is very poor with vertical panning between the two horizontal WFS arrays even though the stimuli were pink noise bursts presenting all necessary cues.

Pieleanu conducted an extensive study about horizontal and vertical localization for first- and second-order Ambisonics in her masters thesis [34]. She reports a mean localization error of up to 13.5° and a SD of around 10° in the median plane depending on experimental conditions when using pink noise bursts. She found no dependency of the localization accuracy on the Ambisonics order when comparing first and second order Ambisonics.

For HOA, attempts have been made in reducing the number of loudspeakers and in tackling other practical constraints as well. One of the proposals is mixed-order Ambisonic (MOA) systems, as for example in [35]. While no localization experiments were conducted in the cited paper, the subjective tests focusing on spatial resolution, clarity, and distance perception seem to show good results. Travis reviewed the basics of the MOA technique regarding elevation rendering in [36], but the simulated systems present elevation localization errors that may easily surpass 10°.

With a mean localization accuracy of $7^\circ - 9^\circ$ and a mean localization precision of $6^\circ - 9^\circ$, the results of our study tend towards the performance that has been shown for physical sound source localization. When compared to other spatialization systems, our study shows similar performance to the dense WFS loudspeaker array and outperforms reported localization precision of other WFS implementations and Ambisonics systems as well as phantom source imaging. Moreover, our study reports results at two listening positions that prove to show similar performance where most studies in the literature only provide results at an ideal listening position (“sweet-spot”).

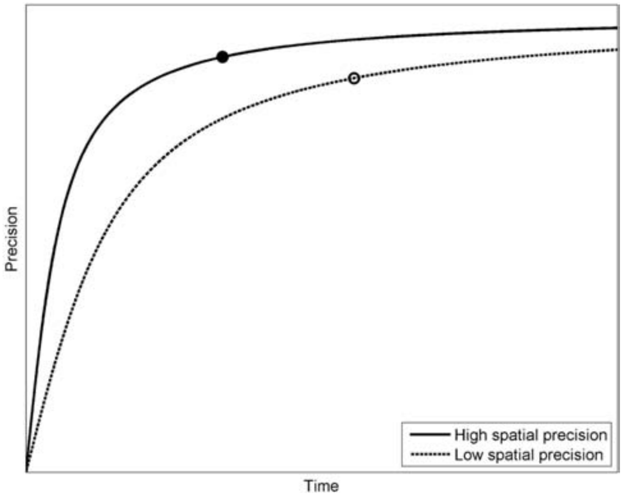


Fig. 6. Hypothesized time-precision tradeoff curves corresponding to the two spatial precision settings with the two points representing the reported results.

3.3 Response Time as Localization Performance Measurement

Response time measurements are a quite recent development in the field of localization performance assessment. In fact, most if not all cognitive tasks are subject to a so-called speed-accuracy tradeoff [37]. The more time a participant is given to accomplish a task, the more accurate a participant's response will be and inversely, the quicker the response has to be given, the less accurate it is. Even if no instructions are given about the time in which a task has to be accomplished, a participant will make such a tradeoff that can be hypothesized to be just below optimal accuracy. So the assumption can be made that the combination between achieved accuracy and response time can give important information about the underlying difficulty of the task. Previous studies showed that factors that are potentially detrimental to accurate localization (such as nonindividualized head-related transfer functions in [38] and high system latencies in [39]) increase the localization response time.

In the presented study an interesting observation is made when comparing the results while using the two different spatial precision settings. Not only a quicker response time is achieved when using a high spatial precision setting (4.3 s decrease), but a better localization precision is also reported (better by 2.3°). In terms of a speed-accuracy tradeoff (i.e., a time-precision tradeoff in this case), these results are expressed as two points lying on different curves as illustrated in Fig. 6. No units are given, since the actual shape of the curves has not been measured during the experiment. The general form of the curves is, however, inferred from the theory and the results found in [37]. No time limit is given for the task, so each point is situated just below optimal precision, but since the attained precision is different, two different curves have to be hypothesized. The optimal tradeoff would be located in the upper left corner, attaining optimal precision in a very small time. Since a high spatial precision gives better source localization precision and smaller response times at the same time, the localization task can be assumed to be accomplished with more ease than with a low spatial precision.

Last, it must be noted once again that head movement was not restricted during the experiment, allowing the participants to move their head freely when listening to the stimuli. Dynamic binaural cues were therefore exploitable by the participants, but this was the case for both spatial precision settings and there should therefore be no bias coming from that fact. It may influence the global mean response time across all settings but not the measured difference in mean response times when alternating between the two settings of the precision parameter.

3.4 Is it Still Wave Field Synthesis?

The proposed technique relies on two fundamental properties of Wave Field Synthesis although it is using a significantly smaller number of loudspeakers than in conventional WFS for the same installation size.

First, it is derived from the Kirchhoff-Hemholtz integral, using a description of the target sound field at the boundaries of a reproduction subspace. As illustrated in the first section of this article, the proposed technique follows similar approximations:

1. Selection of a reduced portion of the surface using a 3-D source visibility criterion,
2. Selection of omnidirectional sources only,
3. Discretization of the line/surface.

The proposed method offers a more general discretization of the surface allowing for irregular loudspeaker distributions. It also proposes an additional weighting of the loudspeakers so as to improve the rendering in a target listening area using an extension of the technique proposed in [9] for 2 $\frac{1}{2}$ -D WFS.

Second, it could be shown in this article that the proposed method preserves localization accuracy within an extended listening area. The proposed method does not realize a perfectly valid physical reproduction. However, the restriction to a horizontal linear array in 2 $\frac{1}{2}$ -D WFS does not preserve the attenuation of the natural sound field and the sound field is not accurately reproduced above the aliasing frequency either. A large portion of the audible bandwidth therefore remains, where the sound field is only reproduced in a plausible way with limited localization artifacts even in conventional WFS. The proposed method goes only one step further but proves to provide reliable localization cues in height at two distinct listening positions separated by one meter.

4 CONCLUSION

The implemented formulation for 3-D WFS enables precise spatial rendering of sound sources while addressing known problems of this reproduction technique. This is confirmed by a source location matching experiment in the median plane. Participants are asked to match the perceived vertical location of a reference loudspeaker with a WFS virtual source pointer. Localization performance is investigated using localization accuracy, localization precision, and response time as performance cues.

Localization accuracy is shown to be good with five levels of elevation being discriminated for two listening positions and two spatial precision settings. A systematic bias of 8.2° is found but this can be easily compensated if good absolute localization is required. Even though not expected, the listening position is shown to influence the perceived location of a source. However, the change is only about 2° that can be neglected.

Localization precision is shown to be about 6° – 9° even though only 24 loudspeakers are employed for the WFS system. As expected, the implemented spatial precision parameter increases the localization precision by 2.3°. The benefits of this parameter are also shown in the response time analysis, where quicker response times were achieved with a higher spatial precision setting, implying a simpler localization process.

Localization performance is therefore judged to be good when compared to other studies with denser loudspeaker arrays or other spatial reproduction techniques. This supports the implemented 3-D WFS technique as a serious alternative to other state-of-the-art spatialization methods.

5 ACKNOWLEDGMENT

This work was supported by the EUREKA Eurostars i3Dmusic project. It is co-funded by the Swiss Federal Office for Professional Education and Technology under grant agreement E!5582 and the European Union under grant agreement INT.2010.0023.[2]

6 REFERENCES

[1] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic Control by Wave Field Synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 11, pp. 2764–2778 (1993).

[2] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, pp. 2–10 (1973 Feb.).

[3] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, pp. 456–466 (1997 June).

[4] E. Corteel, "Equalization in Extended Area Using Multichannel Inversion and Wave Field Synthesis," *J. Audio Eng. Soc.*, vol. 54, pp. 1140–1161 (2006 Dec.).

[5] M. Kolundžija, C. Faller, and M. Vetterli, "Reproducing Sound Fields Using MIMO Acoustic Channel Inversion," *J. Audio Eng. Soc.*, vol. 59, pp. 721–734 (2011 Oct.).

[6] S. Spors, R. Rabenstein, and J. Ahrens, "The Theory of Wave Field Synthesis Revisited," presented at *the 124th Convention of the Audio Engineering Society* (2008 May), convention paper 7358.

[7] M. Naoe, T. Kimura, Y. Yamakata, and M. Katsumoto, "Performance Evaluation of 3-D Sound Field Reproduction System Using a Few Loudspeakers and Wave Field Synthesis," presented at *Second International Symposium on Universal Communication* (2008).

[8] P. Vogel, "Application of Wave Field Synthesis in Room Acoustics," Ph.D. thesis, TU Delft, Delft, The Netherlands (1993).

[9] E. Corteel, R. S. Pellegrini, and C. Kuhn-Rahloff, "Wave Field Synthesis with Increased Aliasing Frequency," presented at *the 124th Convention of the Audio Engineering Society* (2008 May), convention paper 7362.

[10] E. N. G. Verheijen, "Sound Reproduction by Wave Field Synthesis," *Ph.D. thesis*, TU Delft, Delft, The Netherlands (1997).

[11] S. Spors and J. Ahrens, "Analysis and Improvement of Pre-Equalization in 2.5-Dimensional Wave Field Synthesis," presented at *the 128th Convention of the Audio Engineering Society* (2010 May), convention paper 8121.

[12] J. Sanson, E. Corteel, and O. Warusfel, "Objective and Subjective Analysis of Localization Accuracy in Wave Field Synthesis," presented at *the 124th Convention of the Audio Engineering Society* (2008 May), convention paper 7361.

[13] E. W. Start, "Direct Sound Enhancement by Wave Field Synthesis, Ph.D. thesis, TU Delft, Delft, The Netherlands (1997).

[14] H. Wittek, F. Rumsey, and G. Theile, "Perceptual Enhancement of Wave Field Synthesis by Stereophonic Means," *J. Audio Eng. Soc.*, vol. 55, pp. 723–751 (2007 Sept.).

[15] W. de Bruijn, " Application of Wave Field Synthesis in Videoconferencing," Ph.D. thesis, TU Delft, Delft, The Netherlands (2004).

[16] E. Corteel, L. Rohr, X. Falourd, K.-V. NGuyen and H. Lissek, "A Practical Formulation of 3 Dimensional Sound Reproduction Using Wave Field Synthesis," presented at *International Conference on Spatial Audio 2011* (2011 Nov.).

[17] E. Corteel, L. Rohr, X. Falourd, K.- V. NGuyen and H. Lissek, "Practical 3 Dimensional Sound Reproduction Using Wave Field Synthesis, Theory and Perceptual Validation," presented at *Acoustics 2012* (2012 Apr.).

[18] J. Daniel, "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format," presented at *the 23rd International Conference of the Audio Engineering Society: Signal Processing in Audio Recording and Reproduction* (May 2003), conference paper 16.

[19] J. Blauert, "Spatial Hearing, The Psychophysics of Human Sound Localization," revised ed., MIT Press, Cambridge, MA, USA (1999).

[20] 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," presented at *the 30th International Conference of the Audio Engineering Society*: Intelligent Audio Environments (2007 Mar.), conference paper 15.

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

[21] T. Letowski and S. Letowski, "Localization Error: Accuracy and Precision of Auditory Localization," in P. Strumillo (ed.), *Advances in Sound Localization* (InTech, 2011).

[22] J. M. Moore, D. J. Tollin and T. C.T. Yin, "Can Measures of Sound Localization Acuity Be Related to the Precision of Absolute Location Estimates?" *Hearing Research*, vol. 238, no. 1–2, pp. 94–109 (2008 Apr.).

[23] P. Guillon, "Individualisation des Indices Spectraux pour la Synthèse Binaurale : Recherche et Exploitation des Similarités Inter-Individuelles pour l'Adaptation ou la Reconstruction de HRTF," Ph.D. Thesis, Université du Maine, Le Mans, France (2009).

[24] E. Blanco-Martin, F. J. Casajús-Quirós, J. J. Gómez-Alfageme, and L. I. Ortiz-Berenguer, "Objective Measurement of Sound Event Localization in Horizontal and Median Planes," *J. Audio Eng. Soc.*, vol. 59, pp. 124–136 (2011 Mar.).

[25] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 868–878 (1989 Feb.).

[26] J. C. Makous and J. C. Middlebrooks, "Two-Dimensional Sound Localization by Human Listeners," *J. Acoust. Soc. Am.*, vol. 87, no. 5, pp. 2188–2200 (1990 May).

[27] S. R. Oldfield and P. A. Parker, "Acuity of Sound Localization: A Topography of Auditory Space. I. Normal Hearing Conditions," *Perception*, vol. 13, pp. 581–600 (1984).

[28] M. I. Knudsen and M. S. Brainard, "Creating a Unified Representation of Visual and Auditory Space in the Brain," *Ann. Rev. Neurosci.*, vol. 18, pp. 19–43 (1995).

[29] V. Pulkki and T. Hirvonen, "Localization of Virtual Sources in Multichannel Audio Reproduction," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 105–119 (2005 Jan.).

[30] 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 3-D Microphones," presented at *the 30th International Conference of the Audio Engineering Society: Intelligent Audio Environments* (2007 Mar.), conference paper 26.

[31] B. Rakerd and W. M. Hartmann, "Localization of Sound in Rooms, II: The Effect of a Single Reflecting Surface," *J. Acoust. Soc. Am.*, vol. 78, no. 2, pp. 524–533 (1985 Aug.).

[32] R. B. King and S. R. Oldfield, "The Impact of Signal Bandwidth on Auditory Localization: Implications for the Design of Three-Dimensional Audio Displays," *Human Factors*, vol. 39, no. 2, pp. 287–295 (1997 June).

[33] H. Chung, S. B. Chon, J.-h. Yoo and K.-M. Sung, "Analysis of Frontal Localization in Double Layered Loudspeaker Array System," *Proceedings of 20th International Congress on Acoustics* (2010 Aug.).

[34] I. N. Pieleanu, " Localization Performance with Low-Order Ambisonics Auralization," Masters Thesis, Rensselaer Polytechnic Institute, Troy, NY, USA (2004).

[35] J. Käsbach, S. Favrot and J. Buchholz, "Evaluation of a Mixed-Order Planar and Periphonic Ambisonics Playback Implementation," presented at *Forum Acusticum 2011* (2011 June).

[36] C. Travis, "A New Mixed-Order Scheme for Ambisonics Signals," presented at *Ambisonics Symposium 2009* (2009 June).

[37] R. G. Pachella, "The Interpretation of Reaction Time in Information Processing Research," in B. Kantowitz (ed.), *Human Information Processing: Tutorials in Performance and Cognition* (Lawrence Erlbaum Assoc., New York, NY, USA, 1974).

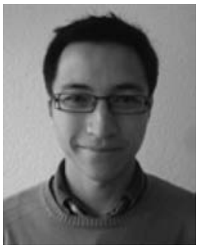
[38] F. Chen, "The Reaction Time for Subjects to Localize 3-D Sounds Via Headphones," presented at *the 22nd International Conference of the Audio Engineering Society*: Virtual, Synthetic, and Entertainment Audio (2002 June), conference paper 257.

[39] S. Yairi, Y. Iwaya and Y. Suzuki, "Influence of Large System Latency of Virtual Auditory Display on Behavior of Head Movement in Sound Localization Task," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 1016–1023 (2008 Nov./Dec.).

THE AUTHORS

Lukas Rohr

Etienne Corteel



Khoa-Van NGuyen

Hervé Lissek

Lukas Rohr was born in Basel, Switzerland, in 1985. He studied electronic engineering at Ecole Polytechnique Fédérale de Lausanne (EPFL) in Lausanne, Switzerland, and received a M.Sc. degree in 2010. The presented study is part of his work towards a Ph.D. degree on the audibility of artifacts from audio source separation in the context of 3-D Wave Field Synthesis. Having joined the Laboratory of Electromagnetics and Acoustics at EPFL in 2010, he is currently involved the European project i3Dmusic, working on sound localization, sound perception, and psychoacoustic modeling. Mr. Rohr is a student member of the Audio Engineering Society and the Swiss Acoustical Society.

●

Etienne Corteel was born in Vernon, France, in 1978. He received a Ph. D. degree in acoustics and signal processing from Paris 6 University, France, in 2004. He joined Studer Professional Audio AG in 2001 in the context of the European Carrouso IST project #1999-20993. He followed up this research at IRCAM, Paris, France, between 2002 and 2004. Between 2005 and 2007, he has shared his time between IRCAM and sonic emotion, Oberglatt, Switzerland. He has joined sonic emotion in 2008. Since 2011, he is the Chief Science Officer of sonic emotion labs in Paris, France. His research interests include the design and evaluation of spatial sound rendering techniques for virtual or augmented reality applications

Khoa-Van Nguyen was born in Rennes, France, in 1983. He obtained a Ph.D. in acoustics and signal processing from Paris Université Pierre et Marie Curie (UPMC) in 2012 for his work on binaural technology. His research was done between 2006 and 2010 at Institut de Recherche Coordination Acoustique Music (IRCAM), Paris, in the Acoustic and Cognitive Spaces (former Room Acoustic) team. In September 2010, he joined sonic emotion labs as researcher in audio, acoustics, and signal processing. His research deals with sound spatialization techniques for consumer and professional markets, spatial auditory perception, and multi-sensory interactions for augmented or virtual reality applications.

●

Hervé Lissek was born in Strasbourg, France, in 1974. He graduated in fundamental physics from Université Paris XI, Orsay, France, in 1998, and received the Ph.D. degree from Université du Maine, Le Mans, France, in July 2002, with a specialty in acoustics. From 2003 to 2005, he was a research assistant at Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, with a specialization in electroacoustics and active noise control. Since 2006, he heads the Acoustic Group of the Laboratoire d'Electromagnétisme et d'Acoustique at EPFL, working on numerous applicative fields of electroacoustics and audio engineering.