

Analysis of Frontal Localization in Double Layered Loudspeaker Array System

Hyunjoo Chung (1), Sang Bae Chon (1), Jae-hyoun Yoo (2) and Koeng-Mo Sung (1)

(1) INMC, School of EECS, Seoul National University, Seoul, Korea

(2) Electronics and Telecommunications Research Institute, Daejeon, Korea

PACS: 43.10.Ce, 60.Fg, 66.Qp

ABSTRACT

For the reproduction of 3-D acoustic images, the horizontal plane of loudspeakers should be extended to vertical directions using additional loudspeakers. In this paper, 32 channel double layered loudspeaker arrays were proposed for the reproduction of sound images in front of a listener. For the localization of virtual sources in both azimuth and elevation, WFS based rendering method is used. First of all, 2-D wave field is synthesized by virtual loudspeaker array, and then each column of upper and lower loudspeaker pair generates virtual loudspeakers by vertical amplitude panning using calculated elevation vectors. Subjective listening test comparing proposed method to 3-D VBAP were conducted for the evaluation of frontal localization quality in this system.

INTRODUCTION

In spatial audio reproduction, loudspeakers are arranged in the horizontal plane in most instances. In these ways, acoustic images are restricted to be placed on the projected plane. For the reproduction of 3-D acoustic images, the horizontal plane of loudspeakers should be extended to vertical directions using additional loudspeakers. Such as 10.2 or 22.2 channel surround formats have upper layer of loudspeakers to represent elevated images of sound sources [4]. Although binaural technique could be used by head-related transfer functions (HRTFs), it has a constraint that is reproduced only by headphones or earphones for one person.

Wave Field Synthesis (WFS) is a holographic approach to reproduce sound field based on Huygens' Principle. The main idea of WFS is that any wave front can be represented by superposition of elementary spherical waves. For driving loudspeakers in WFS system, the individual loudspeaker driving signals must be calculated. Its mathematical basis is the Kirchhoff-Helmholtz integral. It can be simplified into the Rayleigh I integral for monopoles. By Rayleigh's representation theorem and discretization, the loudspeaker functions can be achieved. In many WFS systems, like the conventional 5.1 channel surround system, the listening area was only interested in the horizontal plane at the listener's ear level. So the planar array of loudspeakers could be reduced to line arrays because it does not affect the shape of the wave fronts in the listener's horizontal plane. Therefore such as circular or square arrays surrounding entire listening area (horizontal plane) can be ideal loudspeaker arrangements for the reproduction of sound field by WFS. [2, 3]

However these loudspeaker arrangements could reproduce sound sources only projected on the horizontal plane as mentioned earlier. Though the elevated image source using HRTF elevation cues was researched in [5], it was focused on rear and side reproduction. In this paper double layered loudspeaker arrays were used to reproduce virtual sources expanded to vertical plane in front of a listener. For the localization of virtual sources in both azimuth and elevation, spatial sound rendering technique

is proposed base on WFS. Normal WFS loudspeaker array was replaced by vertical panning image array of two loudspeaker layers above and below listener's ear level.

Subjective listening tests were conducted for the evaluation of frontal localization in this system. Though localization experiments about various spatial reproduction systems had been researched in [1, 8], the used WFS system remained in horizontal plane. In the first experiment the accuracy of listening test system were carried out. In a second experiment the frontal localization test using 3-D vector base amplitude panning (VBAP) [7] was carried out for the comparison with proposed rendering method. In the last experiment the frontal localization of proposed WFS rendering method was evaluated.

RENDERING METHODS

In 3-D sound impression, the most important cue is elevation of sound source. For the localization of virtual sources in both azimuth and elevation, spatial sound rendering technique is proposed base on WFS. Vector Base Amplitude Panning (VBAP) was also used to compare the localization quality between two techniques.

3-D Vector Base Panning

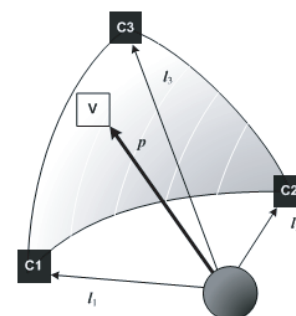


Figure 1: An example of 3-D VBAP.

3-D VBAP is generalization of VBAP to three dimension using

three loudspeakers. The virtual sound source is positioned into a triangle formed by three loudspeakers. In figure 1, \mathbf{l} is the unit vector from the listener to loudspeaker; \mathbf{p} is the unit vector from the listener to virtual sound source. \mathbf{g} is the loudspeaker gain vector. The \mathbf{p} can be expressed as a linear combination form

$$\mathbf{p} = g_1 \mathbf{l}_1 + g_2 \mathbf{l}_2 + g_3 \mathbf{l}_3. \quad (1)$$

And express it in matrix form,

$$\mathbf{p}^T = \mathbf{g} \mathbf{L}_{123} \quad \text{where} \quad \begin{cases} \mathbf{g} = [g_1 \ g_2 \ g_3] \\ \mathbf{L}_{123} = [\mathbf{l}_1 \ \mathbf{l}_2 \ \mathbf{l}_3]^T, \end{cases} \quad (2)$$

thus vector \mathbf{g} can be solved by inverse matrix operation.

$$\mathbf{g} = \mathbf{p}^T \mathbf{L}_{123}^{-1} = [p_1 \ p_2 \ p_3] \begin{bmatrix} l_{1x} & l_{1y} & l_{1z} \\ l_{2x} & l_{2y} & l_{2z} \\ l_{3x} & l_{3y} & l_{3z} \end{bmatrix}^{-1} \quad (3)$$

For satisfying $g_1^2 + g_2^2 + g_3^2 = C$, gain factors have to be normalized using the equation (4). [7]

$$\mathbf{g}^{scaled} = \frac{C}{\sqrt{\sum_{n=1}^3 g_n^2}} \mathbf{g} \quad (4)$$

3-D VBAP rendering in our system, triangles formed by three loudspeakers are connected one by one without intersecting. Total 12 loudspeakers ($U_1, U_4, U_7, U_{10}, U_{13}, U_{16}, L_1, L_4, L_7, L_{10}, L_{13}$ and L_{16}) were used in VBAP rendering as shown in Figure 2. The triangles are made with two adjacent loudspeakers and another nearest loudspeaker in the opposite side layer. Figure 2 shows that a virtual sound source S is located on a triangle made by U_4, L_4 and L_7 loudspeaker.

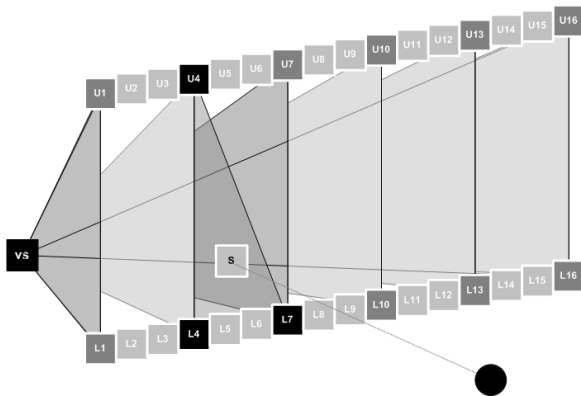


Figure 2: Implementation of 3-D VBAP rendering: an example of virtual source localization by U_4, L_4, L_7 .

Basically, VBAP is a kind of amplitude panning method by coherent signals applied to loudspeakers. Thus loudspeaker positions have to be equidistant from the listening point. In this system, because loudspeaker arrays are straight lines on the frontal plane, the distances from listener to each loudspeaker vary by position. Time delay compensation was a necessary procedure to reproduce accurate virtual sources.

WFS Vertical Panning

First of all, two dimensional wave field is synthesized by virtual loudspeaker array which is located between two real array layers. Each column of upper and lower loudspeaker pair generates virtual loudspeakers, and elevation vectors of each loudspeaker are calculated from the layout of virtual source, upper and lower

loudspeaker pairs. Then each virtual loudspeaker signal is extended to real upper and lower loudspeaker pairs by amplitude panning using elevation vectors. Only vertical amplitude panning was used because horizontal acoustic images are localized by 2-D WFS.

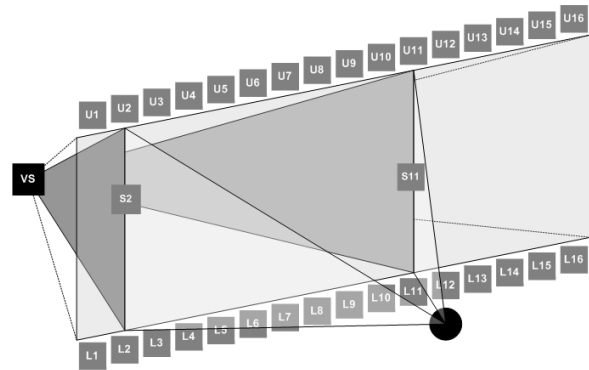


Figure 3: An example of 2-D WFS expanded to 3-D by VBAP.

The main idea of 3-D expansion is as shown in Figure 3, namely rendered secondary source S on horizontal plane (2-D) is panned to upper layer U and lower layer of L . In practice, because the distance from virtual source to each loudspeaker differs, the exact elevation of secondary source may be different such as S_2 and S_{11} in Figure 3. Moreover, in case of VBAP was used for vertical panning, the location of listening point should be the basis for calculating vectors from listener to virtual sound source.

However, sound images were already localized on horizontal plane; only the vertical component from locations of sound source and listener was considered for 3-D expansion model. Thus, virtual sound source was located by vertical components of amplitude panning method based on the arrangement: the height of listener is 1.1 m; distance from loudspeaker arrays to listener is 2 m. Figure 5 represents the final concept WFS vertical panning method for 3-D expansion.

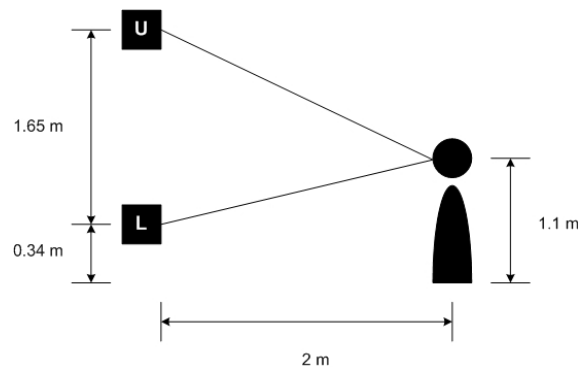


Figure 4: Side view of loudspeaker arrays and listener.

LOCALIZATION TESTS

In this paper, double layered loudspeaker arrays were used to reproduce virtual sources on the vertical plane in front of a listener. The system is built up with 32 channel loudspeakers. Each array has 16 channel loudspeakers with 20cm spacing. Two layers have 1.65m interval in height. Lower layer stands at intervals of 0.34 m away from the floor as described in Figure 4. For the convenience of subjects' test procedure, the grid reference was made by colored strings at intervals of about 30 cm on the vertical plane of loudspeaker arrays.

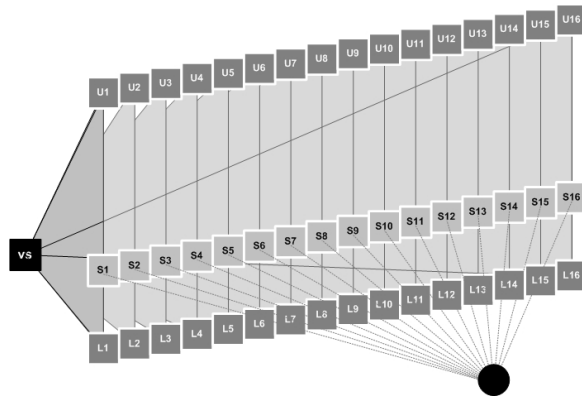


Figure 5: Proposed 3-D expansion model of WFS system.

Test Environment

The localization tests were carried out in a listening room. The room dimension is 4.7 m × 5.3 m × 2.4 m. Reverberation time (RT₆₀) is 0.32 at 1 kHz. The subjects were sitting on a chair 2 m away from the middle of the loudspeaker array (Figure 4). The subjects were free to move their heads during the test.

Subjects Test Panel

Five male subjects ranging from 24 to 33 were participated for the experiment including two authors. All the subjects had previous localization test experiences and had no hearing problems.

Stimuli

Two sound sources were used for localization experiments. First is pink noise bursts with 3 seconds duration. Second is mixed sounds of bicycle ring and ratchets. Each source was rendered by WFS vertical panning and 3-D VBAP methods.

Reference Experiment

The reference experiment was done prior to the localization tests. This test was designed to investigate how accurate the subjects recognize the positions of loudspeakers where the virtual sources are located exactly at the loudspeaker positions. Subjects referred this test to find the position of virtual sources located on the frontal area between two loudspeaker arrays, although reference test had real sources only on the loudspeaker array positions. Pink noise same as localization test stimuli was used. Random set of 20 loudspeaker channels were used in each test. A new random set of loudspeaker channels were generated per each subject. In reference test set, each loudspeaker channel was avoided overlapping. Subjects could repeat test signals until the decision made. During listening tests, subjects used test program implemented by MATLAB. They entered the numbers of the perceived loudspeaker until test ends.

Localization Test Procedure

15 positions of virtual sound source were used during localization tests. All the test procedure used random test set. Each test set had random positions of virtual sound source. A new random set of virtual source positions were generated per each subject as like in reference experiment. Positions of virtual sound source have five different elevations and three different azimuths like 5 × 3 matrix form. All virtual sound sources were rendered as located 1m behind the plane made by loudspeaker arrays. Horizontal positions are -1.2 m, 0 m, and +1.2 m from center. Vertical positions are -0.8 m, -0.4 m, 0 m, +0.4 m and +0.8 m from center. The test results of right side virtual source positions (+0.4 m and +0.8 m) were converted to left side by

symmetry to median plane. In each test, subjects listened to the test stimulus, then point the location of sound source using GUI program implemented by MATLAB. Subjects could repeat the same test stimulus until the decision made in the same manner in reference experiment. In GUI program, loudspeaker arrangements were displayed on screen, and also the same grid with on the loudspeaker arrays. Subjects moved a cursor on the screen by mouse, and clicked when their decisions were made. After first localization test, the same procedure was started with other sound source (bicycle bell sounds).

RESULTS

The test results of reference experiment are shown in Figure 6. 32 gray squares with white circle represent 32 loudspeakers used in this system. *x*-axis is identical to the index of loudspeaker numbers. *z*-axis has the same proportion to the *x*-axis. Namely, the relative distances between squares have same scale with actual loudspeaker arrangements, and the distance between two layers is about 8 times the length of horizontal interval. The asterisks denote average points of each perceived loudspeaker. Horizontal errors are represented by 45° inclined form for avoid overlapping. Vertical errors are represented by dash dot bars. All bars denote 95% confidence intervals. Listening point is projected on the center of the plot for the reference.

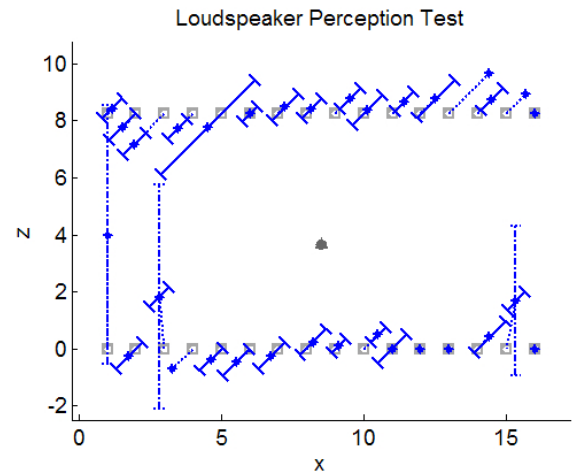


Figure 6: Result of loudspeaker perception tests. Asterisks denote average points of perceived loudspeakers. Diagonal bars denote 95% confidence intervals by horizontal errors. Dash dot lines denote confidence intervals by vertical errors. Listening point is projected on the center of the plot.

There are notable vertical errors in lower side loudspeakers such as L1, L3 and L15. This shows that some subjects selected upper loudspeakers because they perceived sound originated from upper side position, even though the lower loudspeakers were driven actually. When comparing L1 and L3, it is possible to say that the farther toward sides, the more error occurs. From the point of horizontal errors, it seems to be more error occurs from center to sides. The errors have tendency to grow farther toward sides. However in the sections between loudspeaker number 1 to about 3, or about 14 to 16, the errors decrease. Subjects had comparatively correct answers at the middle and at both ends. This is thought to be related with cone of confusion [6].

Results of localization tests are represented in Figure 7, and Figure 8. In test 1, pink noise signal was used for virtual sound sources, and bell sound was used for test 2 as mentioned earlier. In most results of WFS vertical panning method (WFS VP), resolution of horizontal direction was quite accurate. However in test of 3-D VBAP, the side results were slightly further from the intended sound source. It seemed to be that it is related to

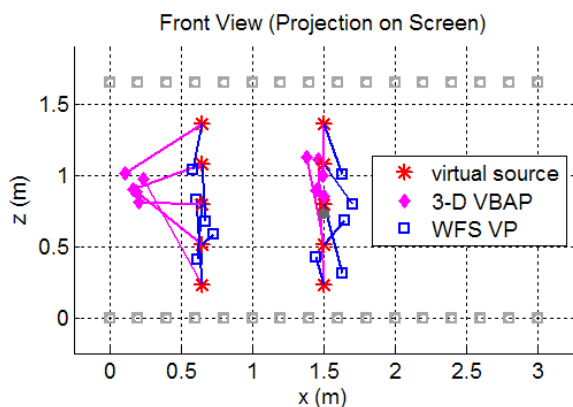


Figure 7: Total results of mean values in test 1.

the feature of VBAP. Secondary image made by three loudspeakers are located only on the triangular plane made by three loudspeakers. Thus the side virtual sources reproduced by our system might be projected toward the screen in normal direction not along the direction of the listening point. It is recognizable by specific results represented in Figure 10 and Figure 11. In both two methods, the results were gathered toward middle of vertical direction. Though results of WFS VP could be classified by elevation, VBAP results were somewhat poor.

In test 2, there was notable tendency that most results were biased close to upper loudspeaker layer. It is more remarkable in VBAP results. Subjects confused to find exact elevation because some subjects felt sound was separated by upper and lower layer, or altered by moving their sights. It is regarded that this is related to the spectrum of signal (Figure 9) used in test 2. Tonal component of the signal might effect as spectral cue for judging elevation. Thus the result of test 2 was not successful than test 1, although it is intended for using familiar sound signals.

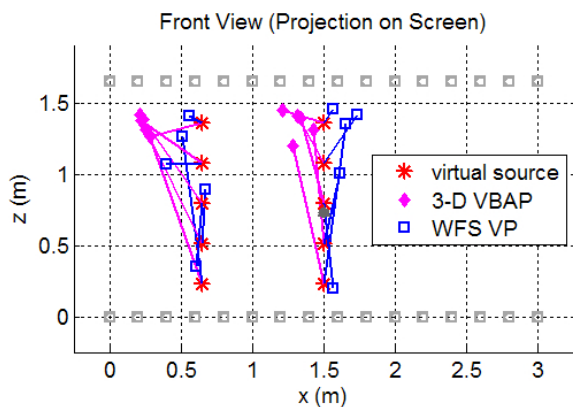


Figure 8: Total results of mean values in test 2.

In Figure 10 and 11, specific test results are represented. In each plot, pair of same elevation and different azimuth test results is shown. Diamonds denote average point of perceived 3-D VBAP localization tests. Squares denote average point of perceived WFS VP localization tests. Ellipses are made by 95% confidence intervals of both horizontal and vertical errors. For the comparison of 3-D VBAP test results and virtual sources in frontal view, the 'original virtual source' was plotted at the

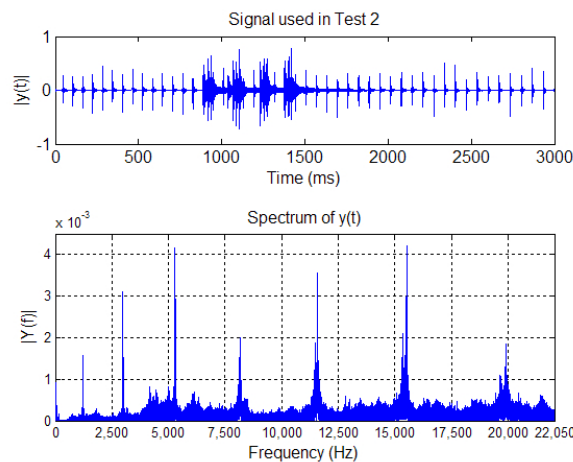


Figure 9: Test 2 signal in time domain and spectrum.

end of extension line from listening point via 'projected virtual source'. Projected virtual source means the cross point from listening point to virtual source through the virtual screen made by two loudspeaker arrays. It is the same as vectors which subjects are staring into. Original virtual source is projection of normal direction to screen which is the same as vectors in the view of loudspeakers. Horizontal localization quality of tested results was acceptable in both methods. 3-D VBAP had very narrow horizontal errors in test 1 although not fine vertical resolution relatively. In case of wide band noise (test 1), middle elevated virtual source had good localization quality especially in 3-D VBAP. However localization quality was better when virtual source was located close to upper array layer than middle or lower case at test 2. Or it could be said that most test results had a tendency to be originated from upper elevation in test 2.

CONCLUSIONS

A WFS based loudspeaker array system was implemented with two layers. WFS vertical panning method was proposed to expand sound images on horizontal plane to three dimensions. Based on conventional WFS methods, WFS VP generate virtual loudspeaker array in the middle of two real loudspeaker layers and reproduce them by vertical amplitude panning. 3-D VBAP was also studied and implemented suitably to this system's loudspeaker arrangement. Subject tests were conducted to evaluate localization quality of this system. By comparing 3-D VBAP and WFS VP, the horizontal localization resolution was observed to be fine in both systems. WFS VP system was somewhat better than 3-D VBAP system. However in the point of vertical localization quality, large errors occurred in both systems, although WFS VP had relatively small variances. Results showed that it is feasible to expand 2-D sound images to 3-D having both azimuth and elevation using double layered WFS array system, although it had errors to perceive clear virtual images in 3-D field. To compare both algorithms more accurately, calibrating 3-D VBAP rendering algorithm in this system remains in future work. Also subjective listening test procedure considered by psychoacoustical methods will be researched to obtain more robust experiments.

ACKNOWLEDGEMENT

This work was supported by the IT R&D program(2008-F-011, Development of Next Generation DTV Core Technology) of KEIT, KCC and MKE, Korea.

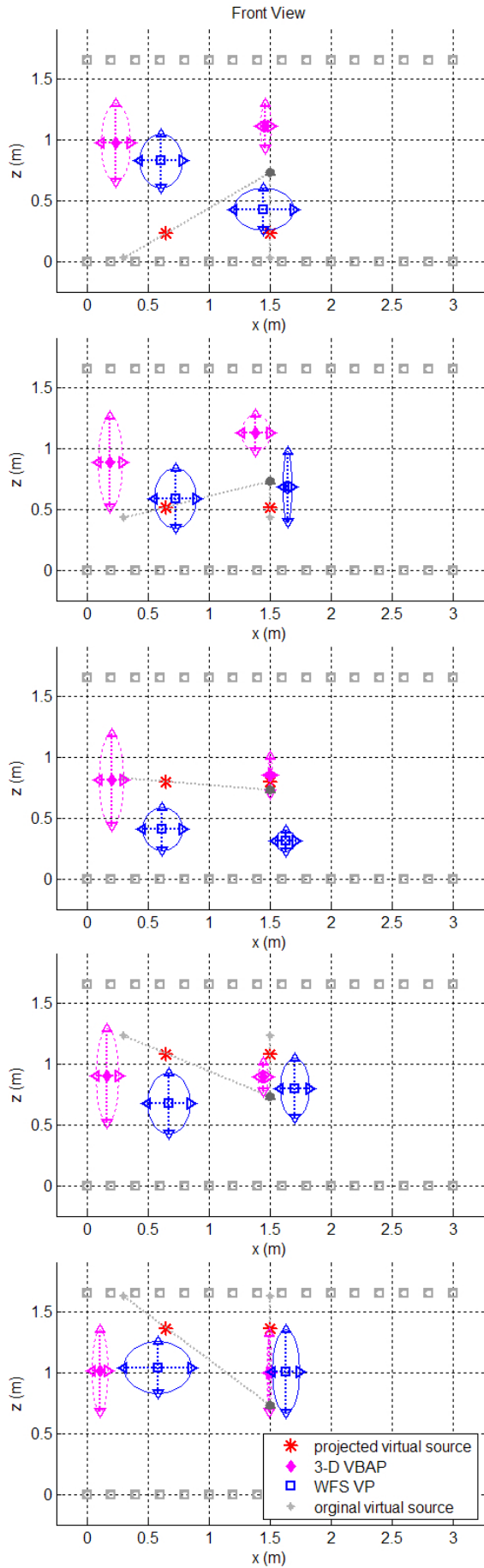


Figure 10: A set of results in test 1.

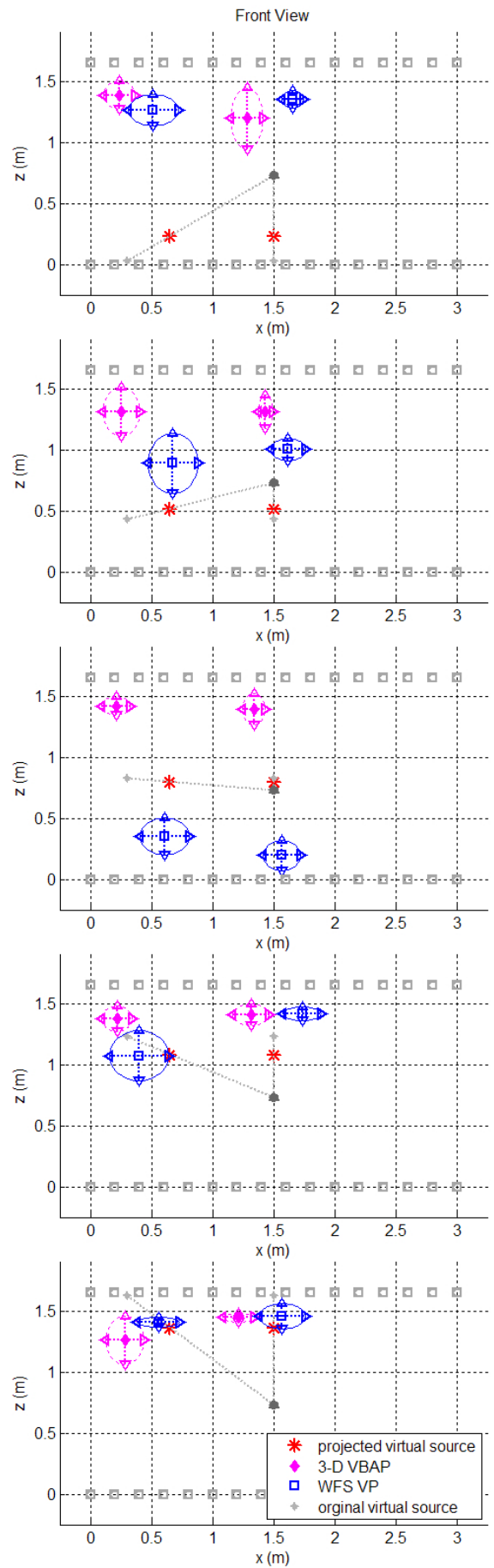


Figure 11: A set of results in test 2.

REFERENCES

- [1] Judith Liebetrau et al. “Localization in Spatial Audio - from Wave Field Synthesis to 22.2”. 123rd AES Convention. New York, NY, 2007.
- [2] A. J. Berkhout. “A Holographic Approach to Acoustic Control”. *J. Audio Eng. Soc.* 36.12 (1988), pp. 977–995.
- [3] A.J. Berkhout, D. de Vries, and P. Vogel. “Acoustic Control by Wave Field Synthesis”. *J. Audio Eng. Soc.* 36 (1993), pp. 977–995.
- [4] K. Hamasaki, K. Hiyama, and R. Okumura. “The 22.2 multichannel sound system and its application”. 118th AES Convention. Barcelona, Spain, May 2005.
- [5] Jose J. Lopez, Maximo Cobos, and Basilio Pueo. “Rear and Side Reproduction of Elevated Sources in Wave-Field Synthesis”. 17th European Signal Processing Conference. Glasgow, Scotland, 2009.
- [6] Brian C. J. Moore. *An Introduction to the Psychology of Hearing*. 5th ed. Academic Press, 2004.
- [7] Ville Pulkki. “Virtual Sound Source Positioning Using Vector Base Amplitude Panning”. *J. Audio Eng. Soc.* 45.6 (1997), pp. 456–466.
- [8] Joseph Sanson, Etienne Corteel, and Olivier Warusfel. “Objective and subjective analysis of localization accuracy in Wave Field Synthesis”. 124th AES Convention. Amsterdam, The Netherlands, May 2008.