Publication III

Static and dynamic sound source localization in a virtual room

M. Gröhn, T. Lokki, and T. Takala

Reprinted, with permission, from *Proceedings of AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, pages 337–344, June 2002, Espoo, Finland.

STATIC AND DYNAMIC SOUND SOURCE LOCALIZATION IN A VIRTUAL ROOM

MATTI GRÖHN[1], TAPIO LOKKI[1], AND TAPIO TAKALA[2]

[1] *Telecommunications Software and Multimedia Laboratory, P.O.Box 5400, FIN-02015 HUT, Finland*
{matti.grohn,tapio.lokki}@hut.fi
[2] *Nokia Ventures Organization, P.O.Box 207, FIN-00045 Nokia Group, Finland*
tapio.takala@nokia.com

An audio localization test comparing accuracy of static and moving sound sources was carried out in a spatially immersive virtual environment, using loudspeaker array with vector based amplitude panning for virtual sound sources. Azimuth and elevation error in localization were measured with different sound signals. As was expected errors in azimuth localization accuracy were smaller than errors in elevation accuracy. There were more localization blur with virtual sound sources than sound sources reproduced directly from a single loudspeaker. Localization blur with moving sound sources were in the same level as with static panorated sound sources. Although the sound sources moved steadily, the measurements indicated that subjects perceived the changes in sound source location stepwise due to applied amplitude panning.

INTRODUCTION

Immersive visualization generally takes place in virtual environments, which provide an integrated system of three dimensional (3D) auditory and visual display. Usually models have parts of specific interest, e.g., chemically active part of the protein. Typically these have been highlighted visually. The most interesting parts of the model/data can also be 'highlighted' by using auditory beacons [1] or some other auditory stimuli. In a dynamic representation it is important that the user is able to follow the location of the moving sound source.

The purpose of our test was to find out, how much the movement of the sound source will affect on localization accuracy in a virtual room. We have previously accomplished two static localization experiments [2, 3]. In this paper we compare the dynamic localization results with the results achieved in [3]. All these experiments were accomplished without visual stimulus.

Auditory localization of static 3D sound sources with headphone reproduction have been tested by other researchers in several experiments [4, 5, 6, 7] previously. Sandvad [8] has measured the localization accuracy in a direct loudspeaker reproduction. The localization accuracy in panned loudspeaker reproduction have been reported for example in two articles [9, 10].

Sound reproduction in a virtual room

We run our experiments in a virtual room [1] of the Helsinki University of Technology [11] (Figure 1). Because virtual rooms are multiuser environments, loudspeakers are typically used for sound reproduction instead of headphones.

[1] http://eve.hut.fi

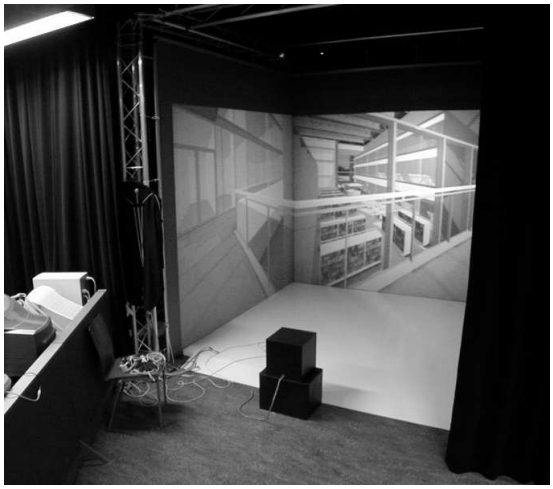

Figure 1: The virtual room of the Helsinki University of Technology. Loudspeakers are behind the screens. No visual stimulus was shown in experiments.

Commonly known and used multichannel loudspeaker reproduction methods are Ambisonics [12], wave field synthesis (WFS) [13] and vector based amplitude panning (VBAP) [14]. In Ambisonics the sound is applied to all loudspeakers all the time. To get an optimal result with Ambisonics the loudspeakers should be in a symmetric layout.

Theoretically the WFS is a superior technique, but unfortunately it is impractical in a virtual room. It is practically impossible to put WFS speaker array in a virtual room without disturbing the visual display.

VBAP is a panning method, where three closest loud-

Loudspeaker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Azimuth	-90	-30	30	90	180	-120	-68	-24	24	68	120	-110	0	110
Elevation	36	36	36	31	30	0	0	0	0	0	0	-34	-39	-34

Table 1: Azimuth and elevation angles of loudspeakers presented from the listening position.

speakers to virtual sound source are used for reproduction. VBAP is less sensitive to listening position than Ambisonics, which is a benefit in a multiuser situation. In addition, it allows more flexible loudspeaker configuration, which is very important because the visual display system limits the possible loudspeaker locations in a virtual room.
In our virtual room we use VBAP for multichannel loudspeaker reproduction. Currently we have 14 Genelec 1029A loudspeakers and their setup is presented in Table 1. The compensation of screens and other part of the implementation of our audio environment are covered in previous article [15].

Direction indication methods

In experiments other researchers have used several different kind of direction indication methods like graphical response screen [16], moving the reference sound [10], pointing schematic drawing of the loudspeaker setup [17], using head mounted laser pointer for pointing [18] or pointing with tracked toy gun [8].
Djelani et al [19] have compared Bochum-sphere technique (also known as GELP), with finger pointing and head pointing. In Bochum-Sphere technique the position of the auditory event is indicated on a sphere representing auditory space. According to their results both pointing methods were superior to the Bochum-Sphere technique. In our experiment we used tracked baton pointing.

1. METHOD

The task of the subjects was to point to the direction of the perceived location of the sound source. The azimuth and elevation values for perceived location were recorded as well as the azimuth and elevation values for sound source location. Subjects did not get any feedback about their pointing accuracy.
Our preliminary static experiment [2] indicated, that in our virtual room non-individualized head related transfer functions (HRTF's) [7] were more inaccurate reproduction method in localization than VBAP.
For this report we accomplished experiment with moving sources. We compared straight and twisted line trajectories. Three different stimuli were used.

1.1. Subjects

For this experiment we had eight non-paid volunteers. Each of them reported to have normal hearing, although this was not verified with audiometric tests. We had seven male subjects and one female subjects. Six of the male subjects were the same as in our previous experiment.

1.2. Stimuli

To utilize both main binaural cues (interaural time differences (ITD) and interaural level differences (ILD)), the sound signal should have enough energy at low frequencies (below 1.5 kHz) and at high frequencies (above 1.5 kHz). There are also other factors in stimulus affecting the localization accuracy like spectral and temporal structure (see for example [20, 21, 22]). It has been found [23] that frequencies near 6 kHz are important for elevation perception.
We used three different stimuli: pink noise, music (excerpt from The Wall by Pink Floyd), and frog croaks. These signals have different kind of spectral content. Pink noise covers the whole audible frequency range, but temporal information is missing. Music is a broadband signal which in addition has a clear temporal structure. Croak sound has most of its energy below 2 kHz. The stimuli were played continuously in a loop.

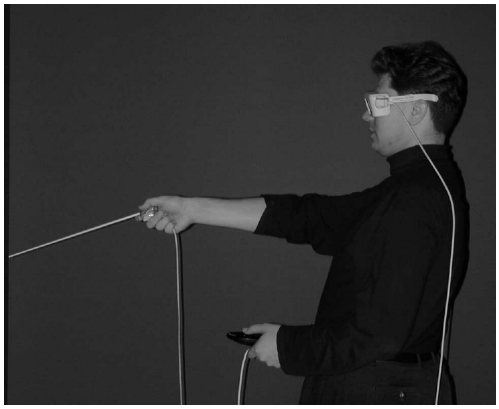

Figure 2: One of the subjects accomplishing the experiment. In his right hand he is holding a tracked baton (used for pointing), and in his left hand he is holding our in-house wandlike device (used for interaction).

1.3. Procedure

During the experiment the user could freely move and turn his head. He pointed the perceived location of the sound source with a baton and indicated that by clicking

a wand button (Figure 2). After the user has clicked the button of the wand, he heard a signal which indicated that sound source has started to move. The task of the subject was to follow the movement of the perceived sound source by pointing it with a baton. The end of the movement was indicated using end signal. There was a short pause before the next trajectory.
There were four straight line trajectories, and four twisted line trajectories. Each stimulus trajectory combination was played twice for each subject. In the first task the straight line trajectories were used, and in the second the twisted line trajectories. Each subject followed in total 48 trajectories.
Before the main experiment subjects have a short three trajectory training task (each stimulus was presented once). The subjects were provided with oral instructions before the training task. After the training task it was checked that the task was clear for the subject.

2. RESULTS OF DYNAMIC LOCALIZATION EXPERIMENT

Azimuth and elevation values of the pointed location were recorded using 60 Hz sampling rate. In addition, the time from start, stimulus, trajectory index, and sound source azimuth and elevation values were recorded.
The results analyzed from stored data are presented by concentrating two main issues. First, the measured localization blur is considered and second the effect of real sources (loudspeaker positions) is discussed. Finally, we compare results of this experiment with results from our previous experiment.

	Azimuth	Elevation
Pink Noise	6.1	16.3
Pink Floyd	8.0	14.3
Frog croaks	7.7	15.4
All signals	6.8	15.3

Table 2: Median values of azimuth and elevation errors for starting points.

2.1. Localization blur

The median values of azimuth and elevations errors [2] in starting points are shown in Table 2. The median azimuth localization error for starting points is 6.8 degrees and in elevation 15.3 degrees.
The median values of azimuth and elevations errors for trajectories[3] are shown in Table 3. As expected the error increase due to a movement of sound source. The median azimuth error in trajectories is 12.5 degrees and median elevation error is 24.1 degrees.

	Azimuth	Elevation
Pink Noise	13.8	26.6
Pink Floyd	12.5	22.8
Frog croaks	11.1	22.7
All signals	12.5	24.1

Table 3: Median values of azimuth and elevation errors for dynamic trajectories for signals.

There is not much difference between the signals in accuracy. In dynamic case the frog croak stimulus produces the least error in azimuth as seen in Table 3. The pink noise produces the largest error for elevation. There is only a small difference in accuracy between the straight and twisted line trajectories as seen in Table 4.

	Azimuth	Elevation
Straight	12.0	22.3
Twisted	13.0	25.5

Table 4: Median values of azimuth and elevation errors for straight and twisted line trajectories.

The example of time dependent behavior of measured trajectories is seen in Figures 3 and 4. There is a thick black line representing the movement of each sound source. Each measured trajectory is plotted using a thin dotted line. On the right in Figure 3 is an example of the twisted line trajectory, the other two trajectories are straight line trajectories.
It is prominent that the measured elevation is in general above the actual sound source location. The average offset is around ten to fifteen degrees above the real location. Also it is quite clear, that it take a little bit time to start to follow the sound source. This delay is variating from approximately one second (see time-azimuth plot on the right in Figure 3) to as long as five seconds (Figure 4).
On the right in Figure 3 it is clearly seen that direction change in azimuth are well perceived after short delay. The changes in elevation are harder to perceive.

2.2. Effect of loudspeaker positions

In our previous static localization experiment the localization blur was smaller in direct loudspeaker reproduction. In other words panned virtual sound sources were not localized as accurate as real sound sources (loudspeakers). Trajectories measured in dynamic experiment suggest, that loudspeaker positions have an effect on trajectories. It seems, that measured trajectories have tendency to bend towards the loudspeaker positions.

[2] azimuth error = abs(perceived azimuth - source azimuth)

[3] Each trajectory was 15 seconds long and measured at.60 Hz sampling rate. All together for one trajectory 900 values of perceived positions was obtained. Error of trajectory is a median of these samples.

Reproduction	Environment	Perception	Pointing
non-optimal loudspeaker positioning	acoustics of a virtual room	localization blur	inaccuracy of pointing
use of VBAP	screens diffuse the signals	MAMA	
	screens low-pass filter the signals		

Table 5: Factors degrading the localization accuracy in a virtual room

The time-azimuth plot on the left in Figure 3 has a time dependent variation in accuracy. This seems to be due to our loudspeaker configuration. In our environment there are loudspeakers in horizontal plane in azimuth positions of -68, -24, 24 and 68 degrees. Near these positions the measured azimuth is more close to source azimuth than in between the loudspeakers, where virtual sound source is panned.
The azimuth-elevation plot on the right in Figure 3 shows, that measured trajectories have bent towards the loudspeaker positions. Especially the top of triangle trajectory has been hard to localize precisely. The time-elevation plot for the same trajectory indicates, that subjects have perceived mainly two different elevations. These two levels of elevation are correlated with loudspeaker positions plus the earlier mentioned ten to fifteen degrees elevation offsets.
The azimuth-elevation plot in Figure 4 shows clearly, that panned end position of the trajectory was very hard to localize. The only measured trajectories having near the correct elevation bent clearly towards to loudspeaker in floor level behind the front screen.

2.3. Comparison of accuracy with our previous static experiment

The median azimuth localization error for starting points is 6.8 degrees and in elevation 15.3 degrees. These are in line with the overall median accuracy in our previous static experiment (median error for azimuth 6.9 and for elevation 15.9).
The median azimuth error in trajectories is 12.5 degrees and median elevation error is 24.1 degrees. These errors are in line with the errors for VBAP reproduction in our previous static experiment (median error for azimuth 11.9 and for elevation 23.4).
Consistency of results in our two experiment suggest, that our test method is reliable.

3. DISCUSSION

According to Blauert [7] the localization blur in azimuth is approximately one degree in optimal conditions. The localization blur for elevation is more signal dependent and it can variate from four degrees (white noise) to seventeen degrees (continuous speech by unfamiliar person). Familiarity with the signal also plays a role in elevation perception.
In our environment there is more localization blur than in optimal conditions. That is natural because there are several factors degrading the localization in a virtual room as listed in Table 5. On the other hand localization blur in direct loudspeaker reproduction in our environment [3] is in line with the accuracy that Sandvad [8] achieved in an anechoic chamber.
Pulkki [10] has found in his listening tests that perceiving elevation of a virtual source is highly individual in VBAP reproduction. Our results support his findings. He also reported that his subjects used only few panning steps between elevation positions from -15 to 30 degrees. The same stepwise perception of elevation is found in our experiment (see Figures 3 and 4).
Grantham [24] has defined a minimum audible movement angle (MAMA), that should be exceeded before it is possible to perceive the direction of the moving sound source. Under the optimal circumstances (slowly moving sound presented directly in front of the subject) the MAMA is between two to five degrees (for the azimuth changes in horizontal plane). MAMA is one of the reasons for the delays in the beginning of trajectories and in turns. Another and according to our experiment at least as strong explanation as MAMA is trajectories tendency to bend (or stay near) towards loudspeaker positions.
Ballas et al. [25] explored the effect of auditory rendering on perceived movement. According to their experiment increasing the number of loudspeakers in VBAP reproduction enhanced the accuracy in perceived movement. The problem in multimodal virtual environment is that due to a visual display configuration, there is a lot of areas, where one can not set loudspeaker. In our environment it might be possible to add still a few loudspeakers.

4. AREAS FOR FUTURE RESEARCH

We tried to utilize ANOVA for analysis, but the dataset does not properly fulfill the assumptions of the ANOVA model. In the near future we should find the statistical tools for the dataset that is not normally distributed.
Pulkki's [10] results suggests that although the perception of elevation is highly individual, each subject is consistent in his perceptions. We should analyze the elevation trajectories of each subject to find out, if we can support this consistency hypothesis.
Our results suggest that localization of moving sound source is more accurate near the loudspeakers positions.

More experiments are needed to measure the effect of loudspeaker positions to localization accuracy.

Our experiments have so far had only one auditory stimulus at a time. In the future, we are interested in to explore the effects of several simultaneous auditory stimuli, for localization accuracy of one specific auditory stimulus. These additional sounds can be static or moving.

Because the long term goal is to use these auditory stimuli in immersive visualization, there is need to explore the localization accuracy in multimodal situation, where there are simultaneous auditory and visual stimuli.

5. CONCLUSIONS

Due to a several factors (see Table 5) there is a more localization blur in our virtual room than in optimal conditions. In addition, the panned sound sources had more localization blur than sound sources reproduced directly from a single loudspeaker.

Localization of a moving sound source was as accurate as localization of a static panned sound source. Especially the changes in azimuth are well perceived. Changes in elevation were perceived stepwise. Measured trajectories have tendency to bend towards the loudspeaker positions.

ACKNOWLEDGEMENTS

We would like to thank Mr. Tommi Ilmonen for his work for our audio software and hardware. We would also like to thank Prof. Lauri Savioja for his support and valuable comments during this experiment.

This work has been partly financed by the Helsinki Graduate School in Computer Science and Engineering (HeCSE, http://www.cs.helsinki.fi/hecse), HPY Research Foundation, and KAUTE Foundation.

REFERENCES

[1] G. Kramer. *Auditory Display: Sonification, audification and auditory interfaces.* Addison-Wesley, Reading, MA., 1994.

[2] M. Gröhn, T. Lokki, and L. Savioja. Using binaural hearing for localization in multimodal virtual environments. In *Proc. 17th Int. Congr. Acoust. (ICA 2001)*, volume IV, Rome, Italy, September 2001.

[3] M. Gröhn. Is audio useful in immersive visualization? In *Proc. SPIE*, volume 4660B, San Jose, California, Jan 2002.

[4] F.L. Wightman and D.J. Kistler. Localization of virtual sound sources synthesized from model HRTFs. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'91)*, New Paltz, NY, 1991.

[5] E.M. Wenzel. Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments*, 1(1):80–107, 1992.

[6] E. Wenzel, M. Arruda, D. Kistler, and S. Foster. Localization using non-individualized head-related transfer functions. 94:111–123, 1993.

[7] J. Blauert. *Spatial Hearing, The psychophysics of human sound localization.* The MIT Press, Cambridge, MA, 1997.

[8] J. Sandvad. Dynamic aspects of auditory virtual environments. In *the 100th Audio Engineering Society (AES) Convention*, Copenhagen, Denmark, May 11-14 1996. preprint no. 4226.

[9] V. Pulkki. Localization of amplitude-panned virtual sources I: Stereophonic panning. *Journal of the Audio Engineering Society*, 49(9):739–752, Sept. 2001.

[10] V. Pulkki. Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning. *Journal of the Audio Engineering Society*, 49(9):753–767, Sept. 2001.

[11] J. Jalkanen. *Building a spatially immersive display - HUTCAVE. Licenciate Thesis*. Helsinki University of Technology, Espoo, Finland, 2000.

[12] D.G. Malham and A. Myatt. 3-d sound spatialization using ambisonics techniques. *Computer Music Journal*, 19(4):58–70, 1995.

[13] A.J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustic Society of America*, 93, May 1993.

[14] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, June 1997.

[15] J. Hiipakka, T. Ilmonen, T. Lokki, M. Gröhn, and L. Savioja. Implementation issues of 3D audio in a virtual room. In *Proc. SPIE*, volume 4297B, San Jose, California, Jan 2001.

[16] E. Wenzel. Effect of increasing system latency on localization of virtual sounds. In *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, pages 42–50, Rovaniemi, Finland, April 10-12 1999.

[17] P. Minnaar, S.K. Olesen, F. Christensen, and H. Møller. Localization with binaural recordings form artificial and human heads. *Journal of the Audio Engineering Society*, 49(5):323–336, May 2001.

[18] R.L. Martin, K.I. McAnally, and M.A. Senova. Free-field equivalent localization of virtual audio. *Journal of the Audio Engineering Society*, 49(1/2):14–22, Jan./Feb. 2001.

[19] T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert. An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization. *ACUSTICA acta acustica*, 86:1046–1053, 2000.

[20] F.L. Wightman and D.J. Kistler. Factors affecting the relative salience of sound localization cues. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 1–23. Lawrence Erlbaum Associates Inc., 1997.

[21] R.O. Duda. Elevation dependence of the interaural transfer function. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 49–75. Lawrence Erlbaum Associates Inc., 1997.

[22] J.C. Middlebrooks. Spectral shape cues for sound localization. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 77–97. Lawrence Erlbaum Associates Inc., 1997.

[23] V. Pulkki. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 62, 2001.

[24] D.W. Grantham. Auditory motion perception: Snapshots revisited. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 295–313. Lawrence Erlbaum Associates Inc., 1997.

[25] J.A. Ballas, D. Brock, J. Stroup, and H. Fouad. The effect of auditory rendering on perceived movement: Loudspeaker density and HRTF. In *Proc. ICAD 2001*, pages 235–238, Espoo, Finland, Jul 2001.

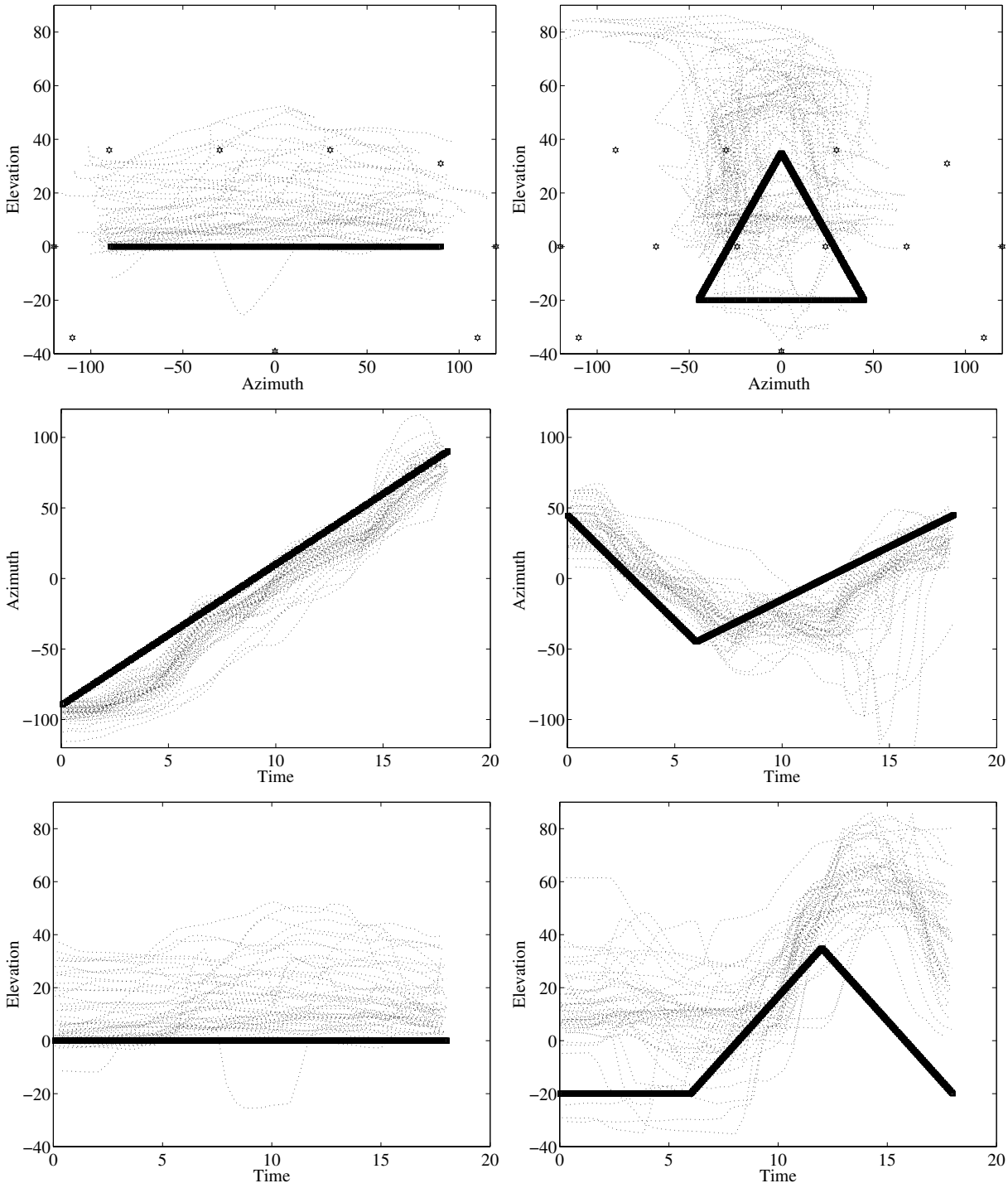

Figure 3: On the left azimuth-elevation, time-azimuth and time-elevation plots for trajectory 1 and on the right the same plots for trajectory 2. Thick black line is the sound source trajectory, and thin dotted lines are measured trajectories. In azimuth-elevation plots the locations of loudspeakers are indicated with star-sign. There are 48 measured trajectories in each figure.

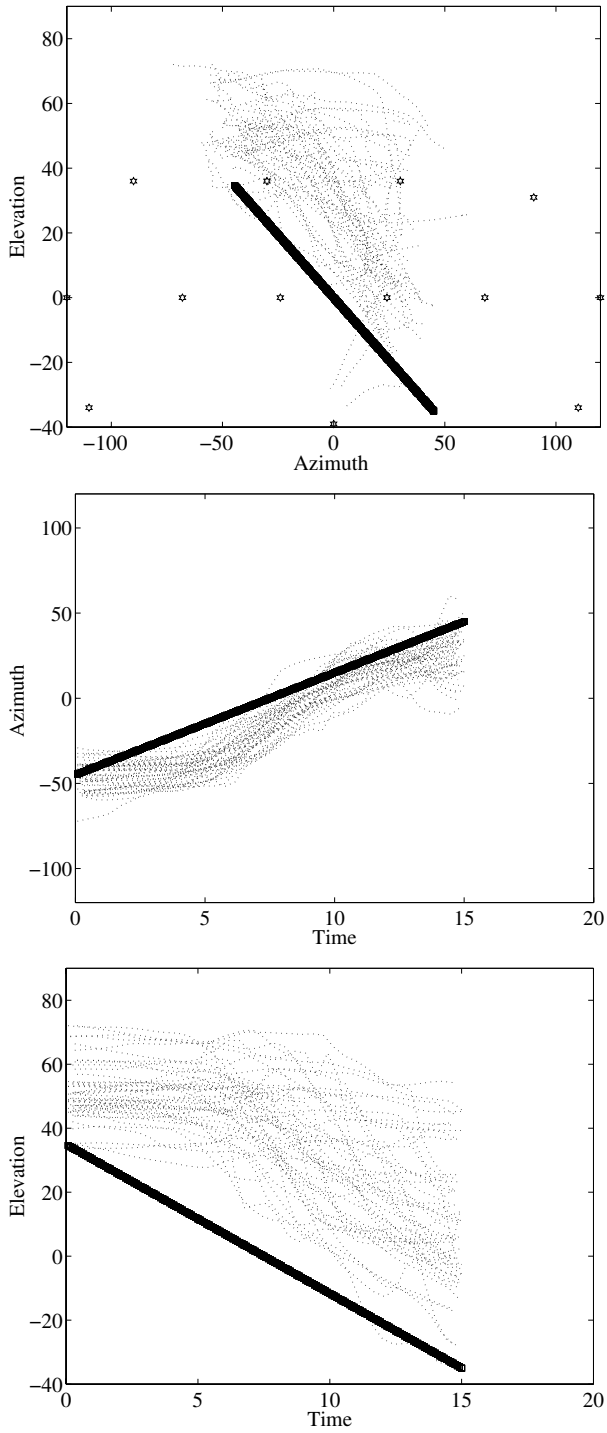

Figure 4: Azimuth-elevation, time-azimuth and time-elevation plots for trajectory 3. Thick black line is the sound source trajectory, and thin dotted lines are measured trajectories. In azimuth-elevation plot the locations of loudspeakers are indicated with star-sign. There are 48 measured trajectories in each figure.