Localization Performance of Real and Virtual Sound Sources

M.Sc.E.E. Jan Abildgaard Pedersen
AM3D A/S
Riihimäkivej 6
DK-9200 Aalborg
Denmark

M.Sc.E.E. Torben Jørgensen
Terma A/S
Hovmarken 4
DK-8520 Lystrup
Denmark

E-mail: jap@am3d.com / toj@terma.dk

ABSTRACT

This paper describes how a 3D-Audio system for use in fighter aircrafts was evaluated in an experiment, by comparing localization performance between real and virtual sound sources. Virtual sound sources from 58 selected directions were evaluated, while 16 of these directions were also evaluated using real sound sources, i.e. loudspeakers. 13 pilots from the Royal Danish Air Force and 13 civil persons were used in the test. The localization performance was split into a constant and a stochastic difference between the perceived direction and the desired direction (stimulus). The constant difference is a localization offset and the stochastic difference is a measure for the localization uncertainty. Stimuli length of both 250 ms and 2 s enabled investigation of the importance of head movements, i.e. using head tracking. Real and virtual sound sources could be located with an uncertainty of 10o and 14o degrees for azimuth while the uncertainty for elevation was 12o and 24o (real and virtual sound sources). No significant localization offset was found for azimuth, while an average offset for elevation of 3o – 6o degrees was found using long stimuli. A significant difference between the localization offset obtained in different directions was found – especially for elevation, where the offset was found to have a strong correlation to the stimuli elevation.

1   INTRODUCTION

3D-Audio is used in fighter aircrafts to enhance the situational awareness. This includes 3D-Audio indication of the directions of detected approaching missiles as well as channel separation of several simultaneous sound signals, e.g. speech from ground control, wingman and alarm signals. The 3D-Audio system included headphone playback system, head tracker and a digital signal processor (DSP). This paper describes the design and the results of a psychoacoustic test, which was used to evaluate such a 3D-Audio system. In particular the localization performance was evaluated, i.e. the capability of positioning sound sources in 3D space at predetermined positions, which enables listeners to localize sound sources at desired positions with some uncertainty. The experiment focused only on the localization performance in relation to direction - distance was not considered. Separate work focusing on distance perception has been presented earlier by other authors [1], [2], [3].

Localization performance has been a subject for research for many years and a long list of authors has presented work including localization performance of real sound sources, phantom sound sources and

Report Documentation Page		*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **01 APR 2005**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Localization Performance of Real and Virtual Sound Sources**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **AM3D A/S Riihimäkivej 6 DK-9200 Aalborg Denmark**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001856, New Directions for Improving Audio Effectiveness (Nouvelles orientations pour l'amelioration des techniques audio)., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT **UU**	18. NUMBER OF PAGES **30**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

virtual sound sources [4], [5], [6], [7], [8], [9], [10], [11]. Many of these papers focus on the localization performance of virtual sound sources, i.e. binaural sound reproduction systems, where test subjects wear headphones and the signals fed to each ear has been recorded using a dummy head or has been processed by head-related transfer functions (HRTF) [5]. Either the recording or the processing will enable listeners to localize sound sources in full 3D as if the listener was present at the recording position or at the synthesized listening position. Head tracking enables the signals fed to the two ears to change accordingly to head movements, so that the perception of the virtual sound source remains while moving the head. The perception of the virtual sound source not only remains because of the head tracking system, it is improved dramatically [5], [10], [11]. Especially front-back ambiguity is almost solved in 3D-Audio systems, which includes a head tracker [11].

When evaluating the localization performance of 3D-Audio systems different strategies has been proposed. One way is to determine the difference between the direction of stimuli (desired direction) and the perceived direction. This will show how large the errors can be when trying to position a virtual sound source in 3D space, i.e. an absolute measure of the angle between the desired and observed position. In a given application this could be compared to the actual needs or requirements for localization performance. However an obvious reference would be real life localization performance, i.e. localization performance of real sound sources [9]. Most 3D-audio systems are implementing an approximation to a real sound field, which means that if the localization performance of the 3D-Audio system is approximately the same as for real sound sources, then the localization performance of the 3D-Audio system is optimal.

The experiment described in this paper included a listening test, where the same test subjects evaluated the localization performance of both a 3D-audio system and a setup of real sound sources. Care was taken to ensure that all conditions were equal for the two evaluations. The only difference was the fact that test subjects did wear headphones when they evaluated the 3D-Audio system, which was not the case when they evaluated the real sound sources.

2   STRATEGY FOR THE EXPERIMENT

2.1   Purpose of the experiment

The purpose of this experiment was to provide information about the localization performance of a 3D-Audio system in relation to usage in "Mission Critical" systems. Specifically the performance measures for systems, where alarm signals are given a direction corresponding to objects that need attention, were of interest. It was a desire to determine the localization performance in situations that resemble the real life use of an auditory display for "Mission Critical" systems. The localization performance for virtual sources was compared to the localization performance for real sound sources. In this experiment headphones were used for virtual sound sources and loudspeakers were used for real sound sources.

2.2   Psychoacoustic method

First the subjects heard one sound signal from one direction either via headphones (virtual sources) or via one loudspeaker (real sources) and were then asked to point to the direction, where they perceived the sound source, i.e. the single stimuli method. Simply pointing to the direction where they perceived the sound source to come from was an intuitive method, which should decrease the need for very long and intensive training sessions.

A laser pointer was mounted on the pointing device in such a way, that a light dot gave visual feedback to the subject to ensure, that the answered direction was indeed the direction, where they perceived the sound

to originate from. Different and independent random sequences of directions and sound source type (real or virtual) were used for each subject.

The main experiment was preceded by a structured training session, which introduced the test subjects to the concept of direction of sound sources as well as running test sessions using both virtual and real sound sources. The aim of the training sessions were to minimize the increased variance introduced by using subjects, who were naive in the task, methods and procedures used in the experiment.

Visibility of the real sound sources, i.e. loudspeakers, would have introduced a bias into the experiment, because subjects would only point to directions, where they could see a loudspeaker – not in directions in between. This bias was avoided by installation of an acoustically transparent curtain (side, top and bottom) surrounding the subjects in such a way that no loudspeakers were visible.

2.3   Localization Performance and definition of direction

The localization performance was split into a constant and a stochastic difference between the perceived direction and the desired direction, i.e. stimulus. The constant difference is a localization offset and the stochastic difference is a measure for the localization uncertainty. A given direction relative to the head of the listener was characterised by the two Euler angles: azimuth and elevation.

Azimuth defines a rotation about the z-axis in a coordinate system, where origo is the centre of the listeners head, z-axis pointing upward and y-axis pointing straight forward, i.e. x-axis pointing to the right hand side. Azimuth is 0 for a direction straight ahead, i.e. in the direction of the y-axis. Azimuth is positive when turning in the direction to the left (counter clockwise), e.g. azimuth is $+90^o$ for a direction directly to the left of the listener and 270° for a direction directly to the right of the listener.

Elevation is a rotation around the x-axis, where a positive angle means upward and a negative angle means downward relative to the horizontal plane for the listeners head, e.g. $+90^o$ is right above the listeners head and -90° is directly below the listener. Elevation is 0° in the horizontal plane.

2.4   Virtual and real sound sources

The perceived direction of a sound will be different from the desired/real direction of the presented sound. This is true both when using virtual and real sound sources. This means, that when judging the localization performance of a 3D-Audio system (virtual sources), one could compare the results to results obtained using real sound sources placed in the desired direction relative to the subject.

A large number of directions were used for virtual sound sources because the 3D-Audio system was the main focus of the experiment. Azimuth and elevation were considered to be two independent variables in the experiment, which meant that separate selection of values for azimuth and elevation were performed. At first 107 directions were considered for use in the experiment for virtual sources, however due to left/right symmetry of the basic set of directions this number was lowered to 58 asymmetric directions, which basically covered the same directions as the basic set. This decision was made to lower the time consumption of the experiment and because verification of left/right symmetry was not part of the purpose of this experiment.

For practical reasons, a significant lower number of real sound sources were used: 16 loudspeakers. These 16 directions for the real sources were a subset of the 58 directions used for virtual sources. A test session was either a loudspeaker session or a virtual source session, so that test subjects did not have to put on headphones or remove them during a test session.

2.5   Controlled parameters in the experiment

The controlled parameters of the experiment were:

- Azimuth angle

- Elevation angle

- Test subject

- Stimuli

- Sound source type: virtual or real

3   DESIGN OF THE EXPERIMENT

3.1   Stimuli

3.1.1   Spectral and temporal characteristics

White noise was selected as the basic stimuli. Both short and long bursts were used: 250 ms and 2 s. Short bursts of white noise were used to prevent subjects from being able to turn their head as part of localizing the sound sources. This stimulus was used to evaluate the static localization performance. Earlier work has proved that the head movement has a dramatic impact on the localization performance, i.e. the difference between static and dynamic sound localization performance [5], [10], [11], [13]. Dynamic localization performance, where then evaluated using the long noise bursts, where subjects had time to use head movements as part of the localization task. During the training sessions only long bursts were used. If a test session lasted longer than 15 min. a break was automatically inserted by the program, which controlled the test. After a break of approximately 5 min. the test session was continued.

Training sessions:

- Long bursts of white noise: 2.00 s.

Experiment sessions:

- Short bursts of white noise: 0.25 s.
- Long bursts of white noise: 2.00 s.
- Maximum 15 min. without breaks

The level of the stimuli was adjusted to 84 dB SPL(LIN) = 75 dB SPL(A) at the listening position while playing the noise signal through a loudspeaker. A dummy head, "Valdemar" [5], [6], [7], were placed on the listening position and used to ensure the same reproduction level using loudspeakers and headphones.

3.1.2 Directions of virtual and real sound sources

Azimuth and elevation had to be independent since they were regarded as two controlled parameters of the experiment. Basically it was decided to use 15 values for azimuth and 9 values for elevation. This corresponds to the 107 directions, because azimuth are undefined for the two elevation values +90° and -90° .The values for azimuth angle were 0°, 24°, 48°, 72°, 96°, 120°, 144°, 168°, 192°, 216°, 240°, 264°, 288°, 312° and 336°. The values for elevation angle were -90°, -66°, -44°, -22°, 0°, 22°, 44°, 66° and 90°. These 107 directions were then reduced to 58 simply by removing either the left or the right direction in a couple, which were left/right symmetric. The used 58 direction for virtual sound sources is shown in figure 3.1.
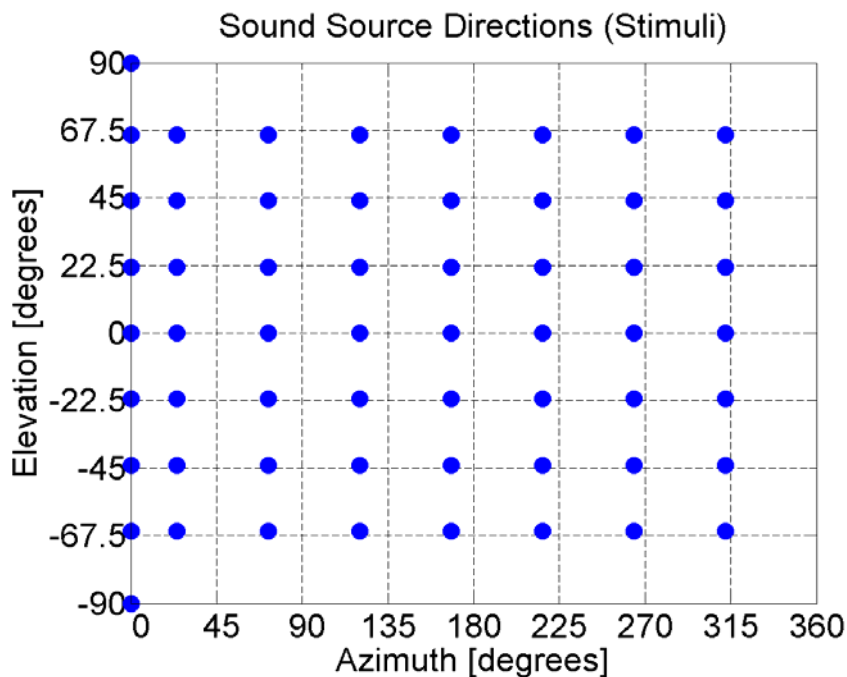

Figure 3.1: Directions for the 58 virtual sound sources without left/right symmetry.

This means that the number of values used for azimuth reduced to 8:

- Azimuth: 0°, 24°, 72°, 120°, 168°, 216°, 264°, 312°
- Elevation: -90°, -66°, -44°, -22°, 0°, 22°, 44°, 66°, 90°

16 of the 58 directions were then selected for real sound sources, i.e. loudspeakers were mounted in these 16 directions. Figure 3.2 shows these 16 directions in a similar way as in figure 3.1. Finally table 3.1 gives the complete list of the 58 directions including "Direction ID", elevation angles, azimuth angles and an indication of "Real Source", which means that such a direction were used both as a virtual and a real sound source.
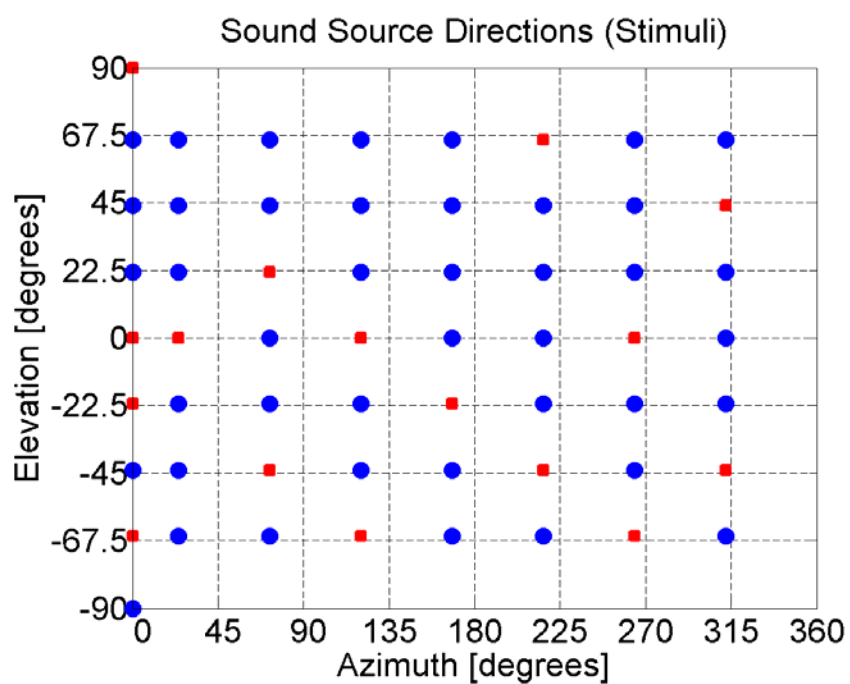
Figure 3.2: Directions for the 16 real sound sources (loudspeakers) shown as red squares, which are a subset of the 58 directions used for virtual sources (blue disks).

Table 3.1: Direction of all 58 virtual and 16 real sound sources.

Direction ID	Elevation	Azimuth	Real Source
1	-90º	0º	-
2	-66º	0º	X
3	-66º	24º	-
4	-66º	72º	-
5	-66º	120º	X
6	-66º	168º	-
7	-66º	216º	-
8	-66º	264º	X
9	-66º	312º	-
10	-44º	0º	-
11	-44º	24º	-
12	-44º	72º	X
13	-44º	120º	-
14	-44º	168º	-
15	-44º	216º	X
16	-44º	264º	-
17	-44º	312º	X
18	-22º	0º	X
19	-22º	24º	-
20	-22º	72º	-
21	-22º	120º	-
22	-22º	168º	X
23	-22º	216º	-
24	-22º	264º	-
25	-22º	312º	-
26	0º	0º	X
27	0º	24º	X
28	0º	72º	-
29	0º	120º	X
30	0º	168º	-
31	0º	216º	-
32	0º	264º	X
33	0º	312º	-
34	22º	0º	-
35	22º	24º	-
36	22º	72º	X
37	22º	120º	-
38	22º	168º	-
39	22º	216º	-
40	22º	264º	-
41	22º	312º	-
42	44º	0º	-
43	44º	24º	-
44	44º	72º	-
45	44º	120º	-
46	44º	168º	-
47	44º	216º	-
48	44º	264º	-
49	44º	312º	X
50	66º	0º	-
51	66º	24º	-
52	66º	72º	-
53	66º	120º	-
54	66º	168º	-
55	66º	216º	X
56	66º	264º	-
57	66º	312º	-
58	90º	0º	X

3.1.3 Processing signals to create virtual sound sources

Signals for right ear and for left ear were convolved in real time with the Head-Related Transfer Functions (HRTF) which corresponded to the direction of the desired sound source position relative to the orientation and position of the listeners head. This was enabled by fitting the headphone with a Head Tracker, which continuously did sent the positions and orientations to the program, which was performing processing of the signals. The used HRTF database did hold measurements from 11950 different directions, i.e. 2° angular resolution [5]. The complete real time 3D-Audio system including HRTF database, head tracker, convolution, binaural reverberation etc. was a product developed by AM3D A/S [12]. The processing also included equalization for the transfer function of the used headphone [5]. A calibration of the Head Tracker was performed before the start of each test session.

3.2 Physical setup

3.2.1 Reproduction system for real sound sources

The physical setup consisted of 16 real sound sources, i.e. loudspeakers, positioned in a distance of 2.13 m (7 feet) from a listening position, see figure 3.3. A curtain cylinder surrounded the listening position to avoid visibility of the loudspeakers. The light in the room was switched off during all tests and light was shined on the curtain from inside the cylinder, which ensured no visibility of objects outside the cylinder. The curtain cylinder was fitted both with a circular top and a circular bottom. The height of the curtain cylinder was 3.70 m and the diameter was 2.90 m.

The test subjects were standing on an adjustable platform, which elevated the ears of all subjects to a height of 2.15 m above floor level. Figure 3.3 shows the used elevation angles, while figure 3.4 shows a top view, where the used azimuth angles are shown. The physical setup was established in a large Television Studio at the facilities of AM3D A/S. The adjustable platform measured 30 cm by 30 cm, which constrained the possibility for test subjects to move away from the reference listening position. However the subjects were allowed to turn their whole body around in order to localize the sound and for precise pointing toward the position, where they perceived the sound to originate from.

The sound pressure responses of all 16 loudspeakers were measured at the listening position, and individual equalization were applied to each loudspeaker to ensure very similar response from all 16 loudspeakers. The resulting amplitude responses of all the loudspeakers were flat in the range from 140 Hz to 15 kHz.
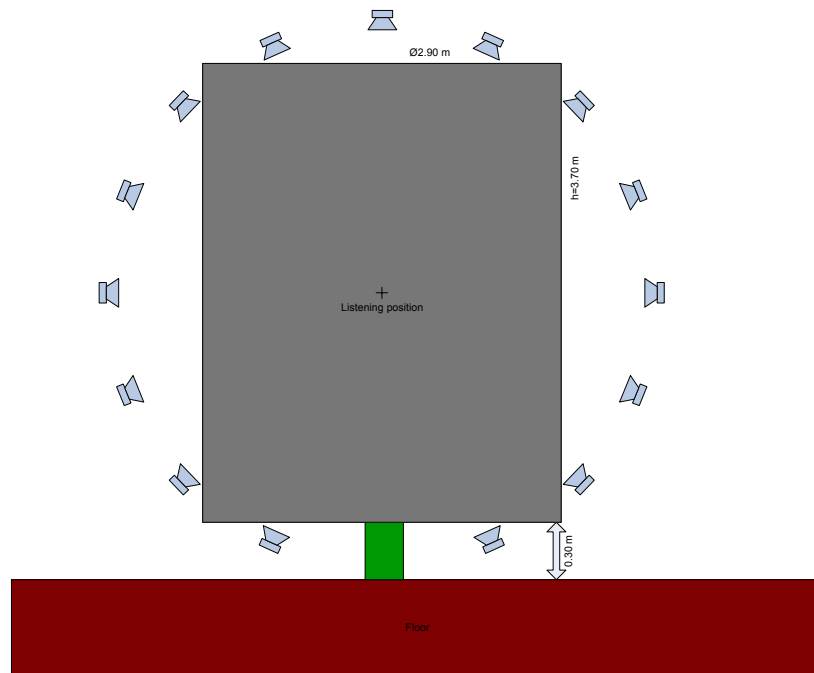
Figure 3.3: Side view of the physical setup, where 16 loudspeakers were positioned around a listening position in a distance of 2.13 m. A curtain cylinder (grey) surrounded the listening position. Diameter of the cylinder was 2.90 m and the height was 3.70 m. Both the top and bottom of the cylinder was fitted with curtain disks to close up the cylinder. Test subjects were standing on a platform (green) to ensure an ear height of 2.15 m.

Figure 3.5 shows the complete setup including the curtain cylinder and the loudspeakers mounted around it. This picture was taken from the left hand side, i.e. from the direction, where azimuth was +90 degrees. It follows from this, that the loudspeakers seen in the left side of the picture, were the loudspeakers directly in front of the listener, i.e. in the direction, where azimuth was 0 degrees.
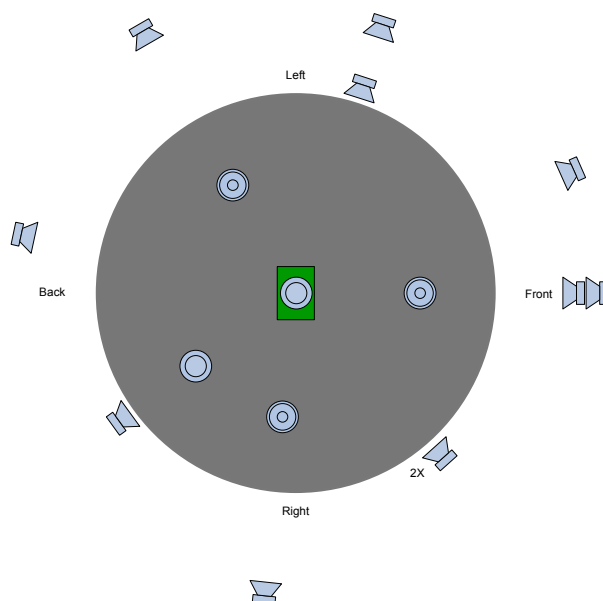
Figure 3.4: Top view of the physical setup, where 16 loudspeakers were positioned around a listening position in a distance of 2.13 m. A curtain disk (grey) forms the two ends of the closed cylinder. The diameter of the disks was 2.90 m. Three loudspeakers were pointing almost upwards (elevation = -66°), which is indicated by the upward facing loudspeakers. Two Loudspeakers were pointing downwards or almost downwards, which is indicated by the downward facing loudspeakers. All loudspeakers were positioned in a distance of 2.13 m from the listening position, but show up at different projected distances in this 2D top view diagram due to different values of elevation angle for the 16 different loudspeakers.

Figure 3.5: Picture of the curtain cylinder with loudspeakers mounted around it.

3.2.2 Reproduction system for virtual sound sources

The reproduction setup for virtual sources consists of signal processing inside the 3D-Audio system and the test subject wearing headphones, which were equipped with a head tracker. Based on the position and orientation of the test subjects head combined with the desired direction of the virtual sound source digital filters were applied to the short or long white noise burst as described in section 3.1.3. The selected headphone was a "Beyerdynamic – DT 990 PRO", which has a specified frequency range of 5 Hz – 35 kHz. The head tracker updated the position and orientation of the headphones with a rate of 60 Hz. The maximum latency was 35 ms. Each headphone session started with a calibration of the head tracker by instructing the subject to look straight forward towards a fixed point on the curtain, which was in the same height as the eyes of the subject and in the direction defined as zero azimuth.

3.2.3 Pointing device

The test subjects had to show their perceived direction by pointing. For this a toy gun equipped with a tracking device was used. The test subjects were instructed to point by aiming at a target using both hands and straight arms. This ensured that the toy gun was held in a position directly in front of the test subjects, i.e. not of the either right or left hand side. The test subject were also instructed to turn their whole body as part of the localization task and when pointing they should be facing the position, where they perceived the sound to originate from. This should prevent subject from pointing and answering in an imprecise way, e.g. by shooting over their shoulder to point to a position behind them. Subjects were instructed to notice where they were pointing by observing where the laser dot shined on the curtain. The update rate of the tracker mounted on the toy gun was also 60 Hz, and this tracker was calibrated at the start of each session by instructing the test subject to point and aim at the same fix point on the curtain as described in section 3.2.2. This was performed at the same time as the calibration of the head tracker.

3.2.4 Automated test system

Presentation of stimuli in the predetermined directions, adjusting the signal processing according to the position and orientation of the head tracker, performing all the necessary signal processing and recording the answered directions by reading the data from the gun tracker was all performed by an integrated system, which was build around a PC including test system software and signal processing software, head tracker control and power amplifiers. Figure 3.6 shows the structure of this system. The system included a user interface which was used by the test operator to control the test.
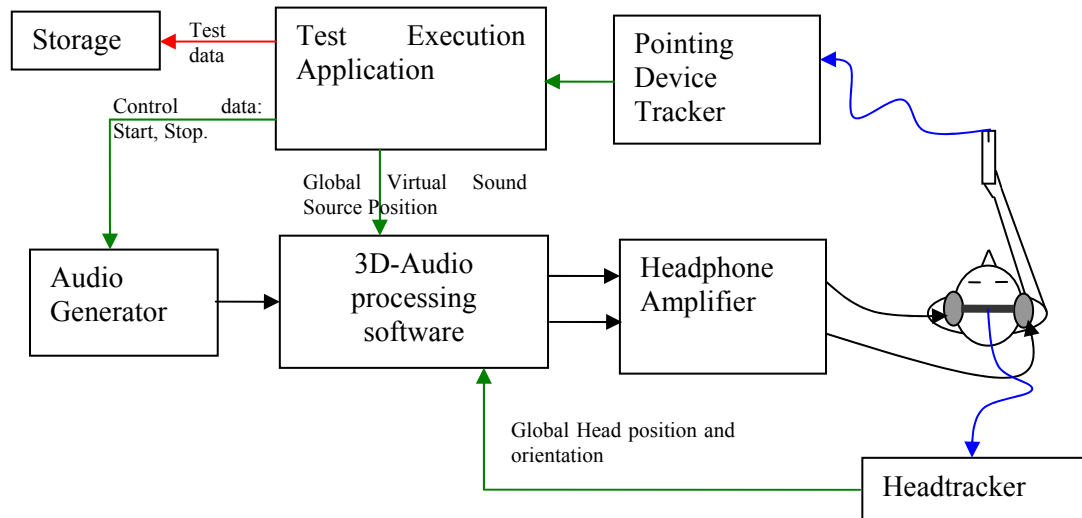
Figure 3.6: Structure of the automated test system

3.3   Procedure for the experiment

3.3.1   Selection of test subjects

The 3D-Audio system under test is targeted towards applications in fighter aircrafts. This led to the desire to test if fighter pilots have special requirements for a 3D-Audio system. In practise this was ensured by recruiting half of test subjects from the Royal Danish Air Force: 13 pilots from the Air Force Bases in Aalborg and Karup, Denmark. The other half of test subjects was civil persons, primarily students from Aalborg University, Denmark. This led to a total number of test subjects of 26 persons.

3.3.2   Screening and instruction of test subjects

All test subjects had a test of hearing threshold, i.e. audiogram. All test subjects were ensured to fall within a range of hearing loss between +20 dB to -10 dB, based on the standard hearing threshold (ISO 389). Then the test subjects had their picture taken, determination of leading eye, collection of personal facts (name, age, sex, experience in listening tests, ear height above the floor and total height) and they had to read a 2 page instruction, which described their task in the experiment. A short discussion between the test operator and the test subject should minimize the risk of misunderstandings. The height of the platform in the bottom of the curtain cylinder was then adjusted according to the measured height of the ears of the test subjects to ensure that the height of the ears when standing on the platform was 2.15 m in all cases. The subjects were not permitted to view the loudspeakers in the setup prior or during the experiment.

3.3.3   Test sessions

The whole experiment was split into 6 test session for each test subject:

- Headphone familiarization (training: 16 directions, 3 repetitions, long stimuli)

- Virtual sound sources, 2 s stimuli

- Virtual sound sources, 250 ms stimuli

- Loudspeaker familiarization (training, 16 directions, 3 repetitions, long stimuli)

- Real sound sources, 2 s stimuli

- Real sound sources, 250 ms stimuli

The order of these 6 sessions was made random for each person by random selection of 1 out of 4 possible sequences:

- Sequence A: Virtual sound sources first   2 s stimuli first
- Sequence B: Virtual sound sources first   250 ms stimuli first
- Sequence C: Real sound sources first   2 s stimuli first
- Sequence D: Real sound sources first   250 ms stimuli first

Following this terminology then sequence A is similar to the sequence given in the start of section 3.3.3.

Each sound source direction was used 3 times for each subject using both headphones and loudspeakers.

The total number of stimuli and answers were:

Headphone familiarization: 26 persons x 16 directions x 3 repetitions = 1248 stimuli
Virtual sources: 26 persons x 58 directions x 3 repetitions x 2 stimuli length = 9048 stimuli
Loudspeaker familiarization: 26 persons x 16 directions x 3 repetitions = 1248 stimuli
Real sources: 26 persons x 16 directions x 3 repetitions x 2 stimuli length = 2496 stimuli

Total: 14040 stimuli, which was equal to 39 hours of effective test (approx. 10 sec/stimuli)

This was equal to 1.5 hour of effective testing time for 1 test subject, but screening, instruction and pauses led to a total testing time for 1 test subject of approx. 2.5 hours. The automatic control software inserted a pause in any test session after each 15 min. of testing time. The 2.5 hour for 1 test subject corresponded to a total testing time of 65 hours.

3.3.4   One test session

The procedure for any one session was:

- The test subject was positioned in the reference position: standing on the platform in the centre of the curtain cylinder facing the fix point on the curtain, which indicated zero azimuth and zero elevation.

- The subjects calibrated both the head tracker and the gun tracker by aiming with two-handed grip and straight arms towards the fix point on the curtain and by looking straight ahead towards the same fix point. The calibration was effectuated by pressing the trigger on the toy gun. In the case of real sound sources then only the gun tracker was calibrated.

- Before each stimulus, the test subject reported "Ready" to the system by pressing the trigger after returning to the reference position facing the fix point.

- Each stimulus signal was presented

- Answered direction was given by the subject through aiming with two-handed grip and straight arms toward the perceived direction and pressing the trigger.

4   RESULTS

4.1   Basic data analysis

The localization error, calculated as the difference between the answered direction and the desired direction, was regarded as a random variable. The distribution was assumed to be a Normal distribution. Unbiased estimators for mean (average), μ, and variance, σ^2, were calculated from the raw directional data obtained, and standard deviation, σ, of the error was simply calculated as the square root of the estimated variance [14]. A 95% confidence interval centred at the estimated mean (average) having a width equal to 3.92 times the standard deviation was then determined as a model describing the basic data [14].

The basic data analysis was performed using data from all 26 test subjects but limited to 1 specific direction, 1 length of stimuli and 1 specific reproduction method. Data from virtual sound sources, 2 s stimuli in the direction ID=28 are plotted in figure 4.1. Direction ID=28 is located in the horizontal plane, i.e. elevation=0°. The azimuth for the direction is 72° (to the left relative to front direction). For such a specific subset of data 78 answers were obtained (26 persons x 3 repetitions).
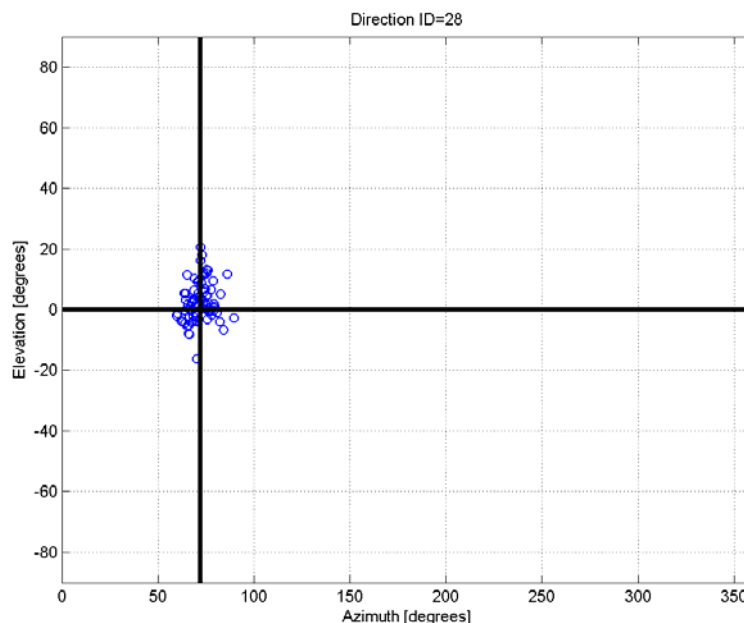

Figure 4.1: 78 answers for the direction indicated by the black horizontal and vertical lines. Each answer is indicated by a blue circle. Data were obtained using virtual sources and 2 s stimuli.

The azimuth and elevation values for each answer were then made relative to the stimuli direction by subtracting the stimuli azimuth and elevation. Figure 4.2 shows these relative answers, i.e. the localization errors. The red rectangle indicates 95% confidence intervals for azimuth and elevation, while the average azimuth error, μ_a, and average elevation error, μ_e, are indicated by black horizontal and vertical lines. The statistical parameters for these data are given in table 4.1.
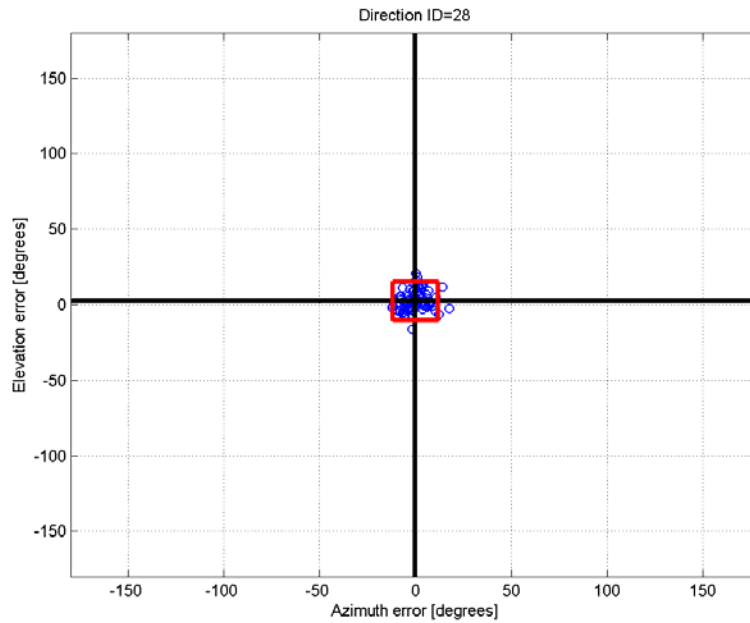

Figure 4.2: Localization error calculated from the 78 answers obtained using virtual sources and 2 s stimuli in direction 28 (azimuth=72°,elevation=0°). The red rectangle indicates 95% confidence interval for azimuth and elevation errors. Each localization error is indicated by a blue circle. Black horizontal and vertical lines indicate the global average azimuth and elevation errors.

Table 4.1: Statistical characteristics for the localization error obtained in the direction ID=28 using virtual sources and 2 s stimuli.

	Azimuth	Elevation
Average error (μ)	-0.1^o	2.6^o
Standard deviation (σ)	6.0^o	6.5^o
95% confidence interval	[-11.9^o ; 11.6 ^{o}]	[-10.1 ^{o} ; 15.4 ^{o}]

Equations 4.1 and 4.2 give the stochastic model for the localization error both for azimuth and elevation.

$$error_{azimuth} \in N(\mu_a, \sigma_a^2) \tag{4.1}$$

$$error_{elevation} \in N(\mu_e, \sigma_e^2) \tag{4.2}$$

4.2 Direction offset and uncertainty across all directions

The analysis process described in section 4.1 was repeated for all 58 directions used for virtual sources. Figure 4.3 shows the average azimuth error, μ_a, for all 58 directions, which shows that the average error, μ_a, must be modelled as a stochastic variable itself. This stochastic variable is also assumed to be a Normal distribution. The black line indicates the over all average for azimuth error, i.e. a global average azimuth error, μ_a, and the 95% confidence interval for the average azimuth error is indicated by red horizontal lines.
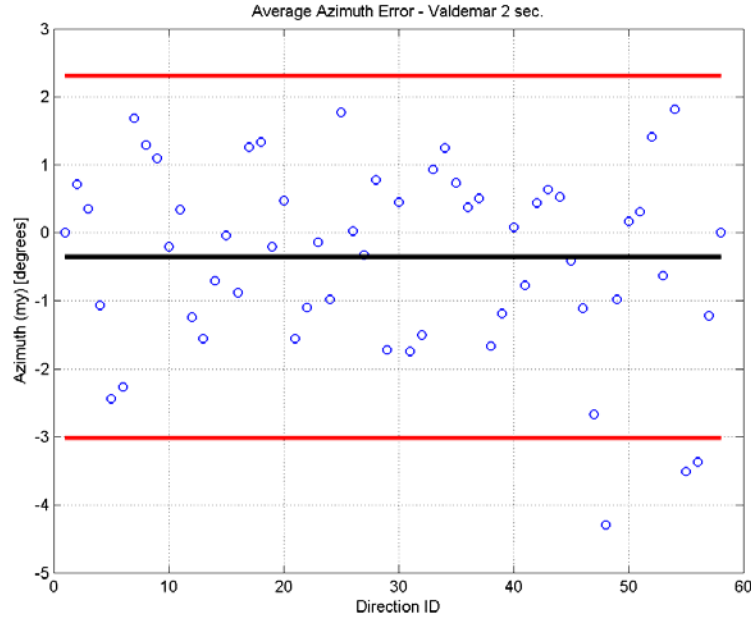

Figure 4.3: Average azimuth error calculated for the 58 directions used for virtual sound sources indicated by blue circles. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the global average azimuth error, and the red horizontal lines indicate the 95% confidence interval for the average azimuth error.

Equation 4.3 gives the stochastic model for the distribution of average azimuth error, μ_a, at different directions, as shown in figure 4.3. The standard deviation of the average error is termed: $\sigma_{average,a}$. The global average azimuth error, $\mu_{a,global}$, was calculated as given in equation 4.4 based on the average azimuth error, $\mu_a(i)$, in all 58 directions.

$$\mu_a \in N(\mu_{a,global}, \sigma^2_{average,a}) \tag{4.3}$$

$$\mu_{a,global} = \frac{1}{58} \bullet \sum_{i=1}^{58} \mu_a(i) \tag{4.4}$$

The localization error can now be modelled as the sum of two stochastic variables, both Normal distributed, which in turn yields one Normal distributed stochastic variable with an increased variance. Equation 4.5 then gives the total stochastic model for the localization error, where $\sigma_{a,total}$ is the total

standard deviation of the azimuth error, which can be calculated using equation 4.6 [14]. This was done by assuming that the stochastic element within isolated directions was independent from the stochastic element between directions [14].

$$error_{azimuth} = N(\mu_{a,global}, \sigma^2_{a,total}) \tag{4.5}$$

$$\sigma^2_{a,total} = \sigma^2_{average,a} + \sigma^2_a \tag{4.6}$$

Figure 4.4 shows the average elevation error for the 58 directions used for virtual sources, 2 s stimuli. Here it should be noted, that the average error seems to correlate to the direction ID. In fact the average elevation error was found to be a function of the stimuli elevation and the global average elevation error, $\mu_{e,global}$. Figure 4.5 shows the relationship between the average elevation error, μ_e, the stimuli elevation, S_e, and the global average elevation error, $\mu_{e,global}$, which is also described in Equation 4.7.

$$\mu_e = \alpha \bullet S_e + \mu_{e,global} = -0.25 \bullet S_e + 5.7^o \tag{4.7}$$

α is an "Elevation error factor", which describes how large a fraction of an increase in stimuli elevation is reflected in an increased elevation error, i.e. larger stimuli elevation yields larger elevation error. Figures 4.6 and 4.7 show the standard deviations for azimuth errors and elevation errors as a function of stimuli direction ID. Figure 4.8 shows how the standard deviations for elevation errors increased when stimuli moved away from the horizontal plane (elevation=0°).
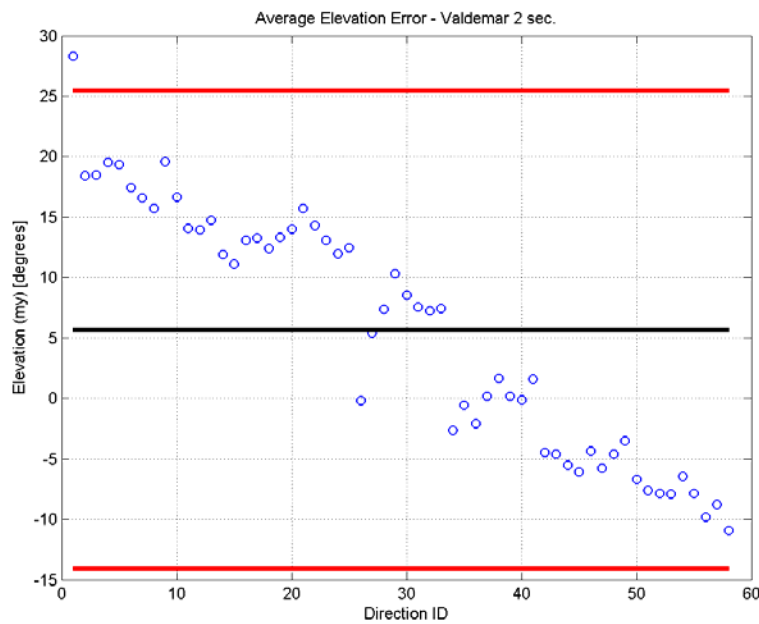

Figure 4.4: Average elevation error calculated for the 58 directions used for virtual sound sources indicated by blue circles. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the global average elevation error, and the red horizontal lines indicate the 95% confidence interval for the average elevation error.
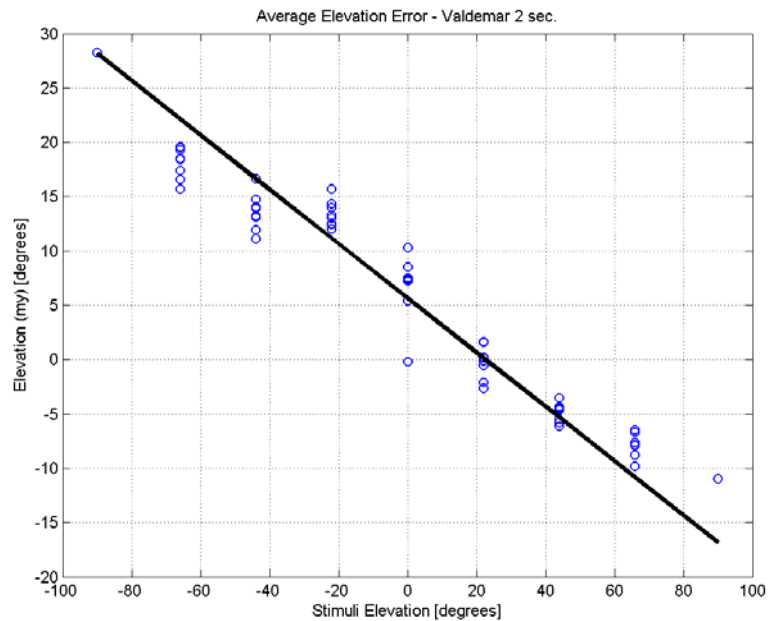
Figure 4.5: Average elevation error calculated for 58 directions plotted as a function of stimuli elevation. The basic data were obtained using virtual sources and 2 s stimuli. The black straight line indicates a simple relationship between average elevation error and stimuli elevation.
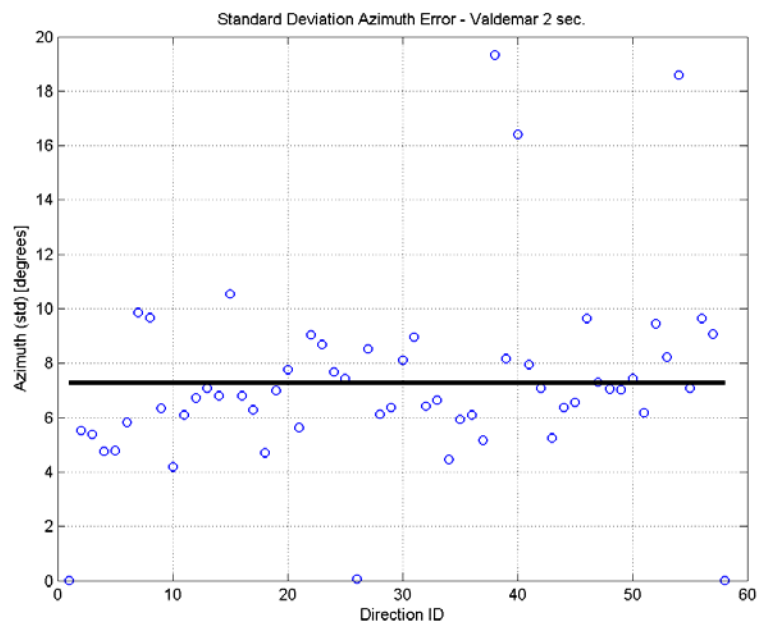

Figure 4.6: Standard deviations for azimuth error calculated for the 58 directions used for virtual sound sources indicated by blue circles. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the average standard deviation across directions.
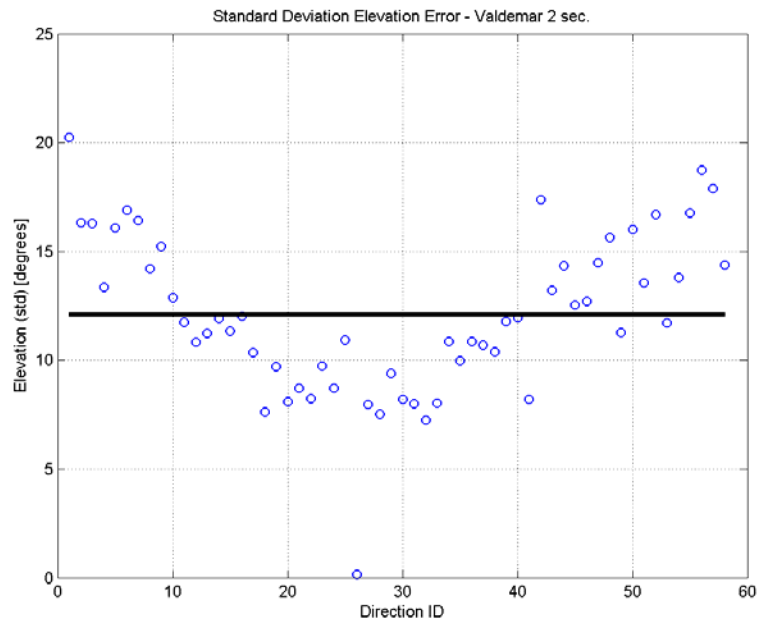
Figure 4.7: Standard deviations for elevation error calculated for the 58 directions used for virtual sound sources. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the average standard deviation across directions.
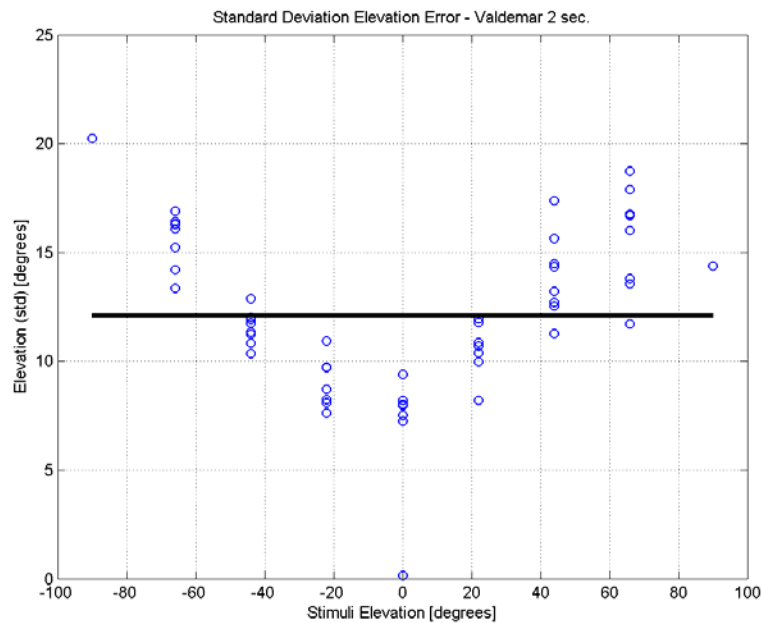

Figure 4.8: Standard deviations for elevation error calculated for 58 directions plotted as a function of stimuli elevation. The basic data were obtained using virtual sources and 2 s stimuli. The black horizontal line indicates the average standard deviation across directions.

Figure 4.9 shows all 4524 answers obtained using virtual sound sources and 2 s stimuli. The black horizontal and vertical lines show the global average for azimuth error, $\mu_{a,global}$, and elevation error, $\mu_{e,global}$. The red rectangle indicates 95% confidence intervals for azimuth and elevation, which are based on the statistical parameters given in table 4.2.
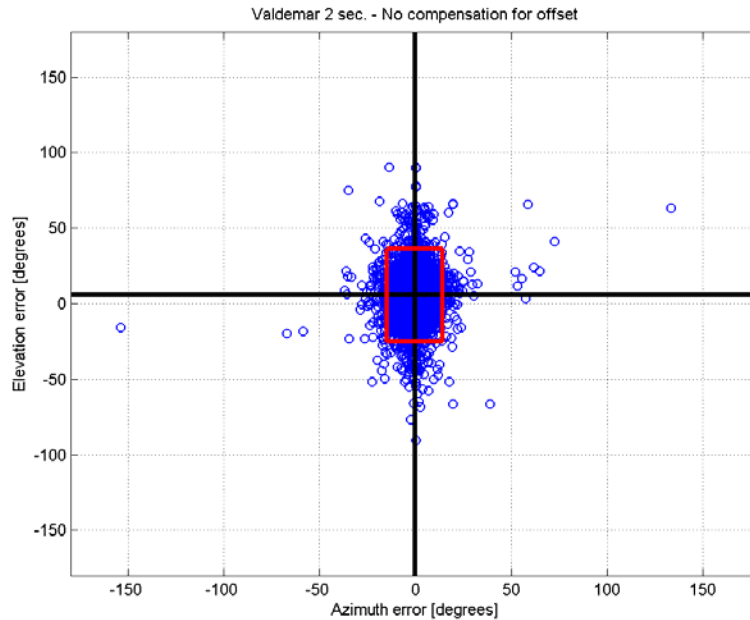

Figure 4.9: Localization error calculated from 4524 answers obtained using virtual sources and 2 s stimuli in all 58 directions. The red rectangle indicates 95% confidence interval for azimuth and elevation errors. Each localization error is indicated by a blue circle. Black horizontal and vertical lines indicate the global average azimuth and elevation errors.

Table 4.2: Statistical characteristics for the localization error obtained using virtual sources and 2 s stimuli.

	Azimuth	Elevation
Average error (μ_{global})	-0.4^o	5.7^o
Standard deviation (σ_{total})	7.4^o	15.7^o
95% confidence interval	[-14.8^o ; 14.1^o]	[-25.2^o ; 36.5^o]

The localization offset was then removed by calculating the localization error as the difference between the answered directions and the average of the answered directions for a given stimuli direction. This is equivalent to compensating for the localization offset in the different directions, see figures 4.3, 4.4 and 4.5. It follows from this that the average localization error becomes zero in all stimuli directions, and the global average localization error vanishes to zero. The standard deviation of the average localization error also becomes zero because the localization error was constant (zero) in all stimuli directions. Though the

total standard deviation for the localization error, σ_{total}, now becomes equal to the standard deviation for the localization error in the individual directions, σ, because no increase in variance is found due to variation of the average localization error across stimuli directions. Figure 4.10 shows all 4524 answers obtained using virtual sound sources and 2 s stimuli after compensation for the localization offset found in the 58 individual stimuli directions. The black horizontal and vertical lines show the global average for azimuth error, $\mu_{a,global}$, and elevation error, $\mu_{e,global}$, which are both trivially equal to 0° after compensation. The red rectangle indicates 95% confidence intervals for azimuth and elevation, which are based on the compensated statistical parameters given in table 4.3.
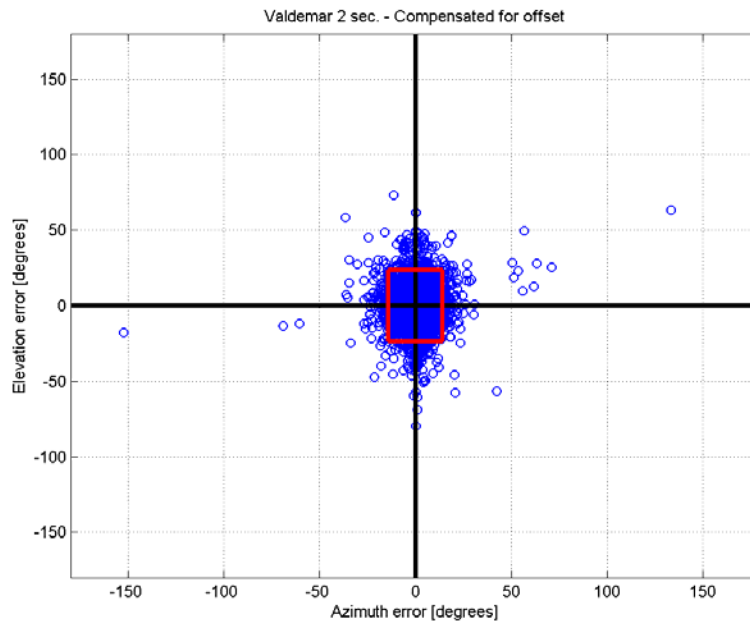

Figure 4.10: Compensated localization error calculated from 4524 answers obtained using virtual sources and 2 s stimuli in all 58 directions. The red rectangle indicates 95% confidence interval for azimuth and elevation errors. Each localization error is indicated by a blue circle. Black horizontal and vertical lines indicate the global average azimuth and elevation errors

Table 4.3: Statistical characteristics for the localization error obtained using virtual sources and 2 s stimuli compensated for the localization offset found in each of the 58 stimuli directions.

	Azimuth	Elevation
Average error (μ_{global})	0.0^o	0.0^o
Standard deviation (σ)	7.3^o	12.1^o
95% confidence interval	$[-14.2^o ; 14.2^o]$	$[-23.7^o ; 23.7^o]$

4.3   Accuracy of physical setup and directional data recording

A special session was run in order to determine the accuracy of the physical setup and the system for recording the directional data, i.e. the positions of the real sound sources (loudspeakers) and the tracker system, which determines the position and orientation of the toy gun. A test subject was placed on the platform inside the curtain cylinder and was instructed to point to each of the 16 loudspeakers one at a time. The test subject was assisted in this task by a person outside the curtain, who ensured that the laser dot shined in the centre of the loudspeaker (+/- 0.5 cm). The light in the room was turned on during this session to enable visibility of the loudspeaker from inside the cylinder. The obtained directional data was then analyzed in a similar way as described in section 4.2. Table 4.4 gives the results of this analysis.

$\sigma_{average}$ is a measure for the accuracy of the positions of the loudspeakers relative the theoretical correct positions as defined in table 3.1. σ on the other hand is a measure for the accuracy of the system for recording the directional data, i.e. gun tracker. The uncompensated confidence interval gives the total localization uncertainty including both the uncertainty of loudspeaker positions and the gun tracker. The compensated confidence interval is the uncertainty only for the gun tracker. This compensation for localization offset was described in section 4.2.

Two methods were tried out for determining the answered direction:

- Direction given by the orientation of the toy gun
- Direction given by a straight line passing through the centre of the head of the test subject and the calculated position of the laser dot on the curtain.

Similar results were obtained by the two methods, which led to usage of the orientation of the toy gun for all directions in this paper.

Table 4.4: Statistical parameters for accuracy of physical setup and directional data recording.

	Azimuth	Elevation
Average localization error (μ_{global})	-0.1$^\mathrm{o}$	-0.9$^\mathrm{o}$
Standard deviation for the average ($\sigma_{average}$)	1.6$^\mathrm{o}$	2.2$^\mathrm{o}$
Standard deviation in each stimuli direction (σ)	0.9$^\mathrm{o}$	0.6$^\mathrm{o}$
Total standard deviation (σ_{total})	1.8$^\mathrm{o}$	2.3$^\mathrm{o}$
95% confidence interval	[-3.7$^\mathrm{o}$; 3.5$^\mathrm{o}$]	[-5.4$^\mathrm{o}$; 3.6$^\mathrm{o}$]
95% confidence interval (compensated)	[-1.8$^\mathrm{o}$; 1.8$^\mathrm{o}$]	[-1.2$^\mathrm{o}$; 1.2$^\mathrm{o}$]

4.4 Summery of results for sound source type and stimuli length

Table 4.5, 4.6, 4.7 and 4.8 gives a summery of the results obtained using the 4 combinations of long/short stimuli and virtual/real sound sources.

Table 4.5: Summery of analysis results obtained using virtual sources and 2 s stimuli.

	Azimuth	Elevation
Average localization error (μ_{global})	-0.4º	5.7º
Elevation error factor (α)	-	-0.25
Standard deviation for the average ($\sigma_{average}$)	1.4º	10.1º
Standard deviation in each stimuli direction (σ)	7.3º	12.1º
Total standard deviation (σ_{total})	7.4º	15.7º
95% confidence interval	[-14.8º ; 14.1º]	[-25.2º ; 36.5º]
95% confidence interval (compensated)	[-14.2º ; 14.2º]	[-23.7º ; 23.7º]

Table 4.6: Summery of analysis results obtained using real sources and 2 s stimuli.

	Azimuth	Elevation
Average localization error (μ_{global})	0.1º	2.9º
Elevation error factor (α)	-	-0.125
Standard deviation for the average ($\sigma_{average}$)	1.3º	4.8º
Standard deviation in each stimuli direction (σ)	4.9º	6.2º
Total standard deviation (σ_{total})	5.1º	7.8º
95% confidence interval	[-9.9º ; 10.0º]	[-12.4º ; 18.2º]
95% confidence interval (compensated)	[-9.6º ; 9.6º]	[-12.1º ; 12.1º]

Table 4.7: Summery of analysis results obtained using virtual sources and 250 ms stimuli.

	Azimuth	Elevation
Average localization error (μ_{global})	-2.8º	0.9º
Elevation error factor (α)	-	-0.50
Standard deviation for the average ($\sigma_{average}$)	13.6º	20.5º
Standard deviation in each stimuli direction (σ)	21.4º	19.2º
Total standard deviation (σ_{total})	25.4º	28.1º
95% confidence interval	[-52.5º ; 46.9º]	[-54.2º ; 55.9º]
95% confidence interval (compensated)	[-42.0º ; 42.0º]	[-37.7º ; 37.7º]

Table 4.8: Summery of analysis results obtained using real sources and 250 ms stimuli.

	Azimuth	Elevation
Average localization error (μ_{global})	0.7º	4.4º
Elevation error factor (α)	-	-0.125
Standard deviation for the average ($\sigma_{average}$)	6.0º	8.7º
Standard deviation in each stimuli direction (σ)	8.8º	11.2º
Total standard deviation (σ_{total})	10.7º	14.2º
95% confidence interval	[-20.2º ; 21.6º]	[-23.5º ; 32.2º]
95% confidence interval (compensated)	[-17.3º ; 17.3º]	[-22.0º ; 22.0º]

4.5   Pilots vs. civil persons

Separate analysis has been performed on data from the 13 pilots and from the 13 civil persons to investigate if differences could be found. Table 4.9 gives the uncompensated confidence intervals and table 4.10 gives the confidence intervals, which are compensated for localization offset.

Table 4.9: 95% confidence intervals – no compensation for localization offset.

	Azimuth	Elevation
virtual sources – 2 s – pilots	[-13.1 $^{\circ}$; 9.5 $^{\circ}$]	[-26.1 $^{\circ}$; 35.2 $^{\circ}$]
virtual sources – 2 s – civil persons	[-15.2 $^{\circ}$; 17.5 $^{\circ}$]	[-24.2 $^{\circ}$; 37.7 $^{\circ}$]
real sources – 2 s – pilots	[-8.0 $^{\circ}$; 6.7 $^{\circ}$]	[-11.6 $^{\circ}$; 17.9 $^{\circ}$]
real sources – 2 s – civil persons	[-10.3 $^{\circ}$; 11.9 $^{\circ}$]	[-12.8 $^{\circ}$; 18.3 $^{\circ}$]
virtual sources – 250 ms – pilots	[-57.2 $^{\circ}$; 48.2 $^{\circ}$]	[-52.4 $^{\circ}$; 57.5 $^{\circ}$]
virtual sources – 250 ms – civil persons	[-47.6 $^{\circ}$; 45.3 $^{\circ}$]	[-55.6 $^{\circ}$; 54.1 $^{\circ}$]
real sources – 250 ms – pilots	[-21.6 $^{\circ}$; 21.6 $^{\circ}$]	[-25.3 $^{\circ}$; 36.0 $^{\circ}$]
real sources – 250 ms – civil persons	[-18.4 $^{\circ}$; 21.1 $^{\circ}$]	[-21.2 $^{\circ}$; 28.0 $^{\circ}$]

Table 4.10: Compensated 95% confidence intervals.

	Azimuth	Elevation
virtual sources – 2 s – pilots	[-11.0 $^{\circ}$; 11.0 $^{\circ}$]	[-22.7 $^{\circ}$; 22.7 $^{\circ}$]
virtual sources – 2 s – civil persons	[-15.9 $^{\circ}$; 15.9 $^{\circ}$]	[-24.2 $^{\circ}$; 24.2 $^{\circ}$]
real sources – 2 s – pilots	[-6.9 $^{\circ}$; 6.9 $^{\circ}$]	[-11.4 $^{\circ}$; 11.4 $^{\circ}$]
real sources – 2 s – civil persons	[-10.6 $^{\circ}$; 10.6 $^{\circ}$]	[-12.0 $^{\circ}$; 12.0 $^{\circ}$]
virtual sources – 250 ms – pilots	[-41.1 $^{\circ}$; 41.1 $^{\circ}$]	[-38.3 $^{\circ}$; 38.3 $^{\circ}$]
virtual sources – 250 ms – civil persons	[-40.5 $^{\circ}$; 40.5 $^{\circ}$]	[-36.0 $^{\circ}$; 36.0 $^{\circ}$]
real sources – 250 ms – pilots	[-16.4 $^{\circ}$; 16.4 $^{\circ}$]	[-24.0 $^{\circ}$; 24.0 $^{\circ}$]
real sources – 250 ms – civil persons	[-16.9 $^{\circ}$; 16.9 $^{\circ}$]	[-18.8 $^{\circ}$; 18.8 $^{\circ}$]

4.6   Front/back reversals

Front-back ambiguity has been found to be a significant problem for 3D-Audio systems, however systems including a head tracker have been shown to almost solve this problem [11]. The front-back ambiguity or front/back reversal occurs when a sound event is perceived in a position in front of the person when the

sound source in fact is positioned behind the person. The characteristics of the misperceived position relative to the true position is a direct font-to-back or back-to-front mirroring in a vertical plane that passes through the two ears of the person (front and back hemisphere). A special analysis to uncover this problem has been performed, where the localization error was calculated to be the smaller of two possibilities:

- The difference between the perceived direction and the stimulus direction
- The difference between the mirror image of the perceived direction and the stimulus direction

Table 4.11 gives the results for this analysis, where front/back% is a measure of how many times the mirrored position was used compared to the total number of positions, i.e. how often front/back reversals were identified.

Table 4.11: Localization offset compensated 95% confidence intervals for azimuth, where front/back reversals mirrored back to positions closer to the stimuli directions.

	Azimuth - normal	Azimuth - mirrored	Front/back%
virtual sources – 2 s	$[-14.2^{\,o}\,;14.2^{\,o}]$	$[-12.7^{\,o}\,;12.7^{\,o}]$	5.1%
real sources – 2 s	$[-9.6^{\,o}\,;9.6^{\,o}]$	$[-8.2^{\,o}\,;8.2^{\,o}]$	4.2%
virtual sources – 250 ms	$[-42.0^{\,o}\,;42.0^{\,o}]$	$[-27.2^{\,o}\,;27.2^{\,o}]$	21.3%
real sources – 250 ms	$[-17.3^{\,o}\,;17.3^{\,o}]$	$[-14.3^{\,o}\,;14.3^{\,o}]$	9.1%

5   DISCUSSION

One of the interesting findings in the results was the strong correlation between average elevation error and stimuli elevation, which was shown in figure 4.5. Equation 4.7 formulated a simple relationship between stimuli elevation, S_e, and the average elevation error, μ_e. The global average elevation error, $\mu_{e,global}$, formed an offset for the average elevation error as seen in equation 4.7. One way to exploit this knowledge is to compensate the stimulus direction in such a way, that taking the average elevation error into account will yield a perceived direction, which closely approximates a desired elevation, D_e. Equation 5.1 states the relationship between the elevation error, $error_e$, stimulus elevation, S_e, and the answered elevation, A_e. The average elevation error, μ_e, can then be calculated from equation 5.2, where E[x] is the expected value operator [14].

$$error_e = A_e - S_e \tag{5.1}$$

$$\mu_e = E[A_e - S_e]$$
$$\Updownarrow \tag{5.2}$$
$$\mu_e = E[A_e] - S_e$$

The expected value of the answered elevation, i.e. the average answered elevation, $E[A_e]$, should be equal to the desired elevation, D_e. Combining this with equation 4.7 leads to equation 5.3, which gives the

relationship between the desired elevation, D_e, and the stimuli elevation, S_e. Isolating the stimuli direction, S_e, leads to equation 5.4, which specifies the needed stimuli direction for a given desired elevation in order to compensate for the found relationship between average elevation error and stimuli direction.

$$D_e = \mu_e + S_e = \alpha \bullet S_e + \mu_{e,global} + S_e = (1+\alpha) \bullet S_e + \mu_{e,global} \qquad (5.3)$$

$$S_e = \frac{D_e - \mu_{e,global}}{1+\alpha} \qquad (5.4)$$

Both the global average elevation error, $\mu_{e,global}$, and the elevation error factor, α, can be found in tables 4.5, 4.6, 4.7 and 4.8. Using equation 5.4 for calculating the needed stimuli leads generally to stimuli elevation angles larger than the desired elevation angles, e.g. for a desired elevation angle of 60° require a stimuli elevation angle of 72.4^o using virtual sources and 2 s stimuli (table 4.5).

However stimuli elevation angles can't go above $+90^o$ or below -90°. This fact imposes a limit onto the range of possible desired elevation angles. Setting the stimuli elevation, S_e, to $+90^o$ and -90° in equation 5.3 leads to a lower and upper limit for desired elevation angles to -61.8^o and $+73.2^o$. The impact of this statement is that positions near the "poles of the sphere", i.e. -90° elevation and $+90^o$, represents a special challenge, which require further research.

Why should stimuli elevation angle be larger than the desired elevation angle? One possible answer could be that humans are more specialized in localizing in the horizontal plane due to everyday life, where the majority of sound events take place in the same level above the ground as our ears. This combined with the fact that the human ears are located in a way, which favours the discrimination from left to right rather than down to up, may explain why humans has a default localization near the horizontal plane?

The results given in table 4.5, 4.6, 4.7 and 4.8 supports a statement saying that: "The more a human is in doubt about where the sound originates from – the more restricted range exists for the answered elevation angles". E.g. using virtual sources and short stimuli (250 ms) yields a possible range for the desired elevation angles to $[-44.1^o ; 45.9^o]$. The experience is that sound localization becomes more difficult for shorter stimuli and for virtual sources compared to real sources. Table 5.1 gives the range of possible desired elevation angles in the 4 cases originating from tables 4.5, 4.6, 4.7 and 4.8.

Table 5.1: Ranges for possible desired elevation angles.

	Possible desired elevation angles
virtual sources – 2 s	$[-61.8^o ; 73.2^o]$
real sources – 2 s	$[-75.9^o ; 81.7^o]$
virtual sources – 250 ms	$[-44.1^o ; 45.9^o]$
real sources – 250 ms	$[-74.4^o ; 83.2^o]$

Using azimuth and elevation to specify a direction in 3D may represent a problem: when calculating the average elevation error near the "poles of the sphere", then the answered directions are limited/clipped to

the range +/- 90°. E.g. for stimuli elevation angle = +90° , where test subjects will point to directions distributed in an area around the "pole of the sphere". Any position away from the "pole of the sphere" will yield a negative elevation error – the elevation error can't be positive in this situation by the definition of elevation angle. It follows from this, that even if the answered directions are evenly distributed around the "pole of the sphere", then the average elevation error will not be zero, i.e. it will be negative. This led to a special analysis, where the average localization error wasn't spilt into azimuth and elevation error – it was performed as an average of complex numbers, which represented the localization error in polar form. However this didn't change the result dramatically, but the standard deviation for elevation error did decrease.

Figure 4.9 and 4.10 confirms in a graphical way the fit between the obtained localization errors and the stochastic model including the calculated parameters. Figure 4.10 shows how compensation for the localization offset do decrease the overall localization uncertainty in a significant way. It follows from this, that it is very important to separate the variance due to different localization offsets in different direction from the variance due to localization uncertainty in each direction.

Table 4.4 shows that both the accuracy of the physical setup and the accuracy of the directional data recording system were sufficient compared to the results obtained in the experiment.

Standard deviation for azimuth using real sources and 2 s stimuli was found to be 4.9^o (table 4.6), which is comparable to results obtained by other authors: 1.5^o (in the stimuli range +/-30°) [4], 8° (average absolute localization error) [9] and 1° to 3° [8]. The standard deviation for elevation errors using real sources and 2 s stimuli was found to be 6.2^o, which again could be compared to the average absolute localization error found by [9]: 8°. Differences in choice of stimuli, stimuli length, physical setup, range and distribution of sound source positions may be possible explanations for these differences. It should be noted, that this experiment covered the full sphere of directions, i.e. combinations of azimuth and elevation covering the full sphere.

Virtual sources and stimuli length of 2 s yielded a standard deviation of 7.3^o for azimuth and 12.1^o for elevation. The average absolute localization error for virtual sources was found by [9] to be 11° and 10° by [13]. Using short stimuli (250 ms) this was found to increase to 21.4^o for azimuth and 19.2^o for elevation. This compares to the data presented by [13]: 18°, which was a test restricted to the horizontal plane.

The standard deviations obtained for pilots generally were found to be smaller compared to civil persons. A significantly lower standard deviation was found for azimuth errors for pilots compared to civil persons: 5.0^o vs. 8.0^o. The ability to focus on a task at hand and to use the input from the auditory system may be an explanation for this result?

The front/back reversal rate was found to be 4.2% (2 s stimuli) and 9.1% (250 ms stimuli) for real sources, which is lower than the results obtained by [9]: 12%. However this shows that even when using real sources the auditory system has a significant rate of front/back reversals. The rate for virtual sources was found to be 5.1% (2 s stimuli) and 21.3% (250 ms). In [9] the front/back reversal rate was found to be 20% for virtual sources and in [13] the rate was found to be 4% (2 s stimuli) and 10% (250 ms). It is clear, that short stimuli, i.e. no head movement, greatly increases the problem with front/back reversals both for real and virtual sources, however the increase was substantially larger for virtual sources.

The main results of the paper show that the uncertainty for azimuth for virtual sources does compare to the real sources. However the uncertainty for elevation is significantly higher for the virtual sources. Actually the uncertainty for elevation for virtual sources (2 s stimuli) compares well with real sources for short stimuli (250 ms). The uncertainty both for azimuth and elevation increased for shorter stimuli, where head movements can't be used in the localization task. This was particularly clear for virtual sources.

6   CONCLUSION

The localization uncertainty was much higher for short stimuli (0.25 sec.) compared to long stimuli (2 sec.). The long stimuli enabled head movements to be used in the localization task.

Pilots had a lower localization uncertainty compared to civil persons. Head movements, i.e. long stimuli, greatly reduced the problem of front/back reversals, especially for virtual sound sources.

No significant localization offset was found for azimuth, while an offset for elevation of 3 – 6 degrees was found using long stimuli. A significant difference between the localization offsets (average error) obtained in different directions was found – especially for elevation, where the offset was found to have a strong correlation to the stimuli elevation. It follows from this, that a significant part of the uncertainty for elevation can be removed by compensating for these individual offsets in the different directions (Compensated for localization offset).

A simple formula was devised for calculating the needed stimuli elevation from a desired elevation angle, which generally introduces larger stimuli elevation angles compared to the desired elevation angle.

Due to the needed compensation of elevation angle a restricted range of desired elevation angles was determined, which showed a problem of how to position sound sources near the "poles of the sphere", i.e. near elevation = $+90^o$ and -90°.

The results of the analysis of the obtained directional data compared well to results presented by other authors under the observation that significant differences existed between the different experiments: stimuli, stimuli length, reproduction setup, range and distribution for azimuth and elevation angles and sound source type. This experiment covered the full sphere of directions: both azimuth and elevation.

The localization uncertainty for virtual sound sources was found to be larger than for real sound sources, especially for elevation. However the localization performance of virtual sound sources was comparable to the performance of real sound sources:

95% confidence interval (compensated for offset):

Real Sound Sources: Azimuth=[-9.6 ; 9.6]   Elevation=[-12.1 ; 12.1]

Virtual Sources: Azimuth=[-14.2 ; 14.2]   Elevation=[-23.7 ; 23.7]

7   ACKNOWLEDGMENT

8   REFERENCES

[1]   Douglas S. Brungart: "A Speech-Based Auditory Distance Display", presented at the AES 109th convention, Los Angeles, 2000, September 22–25. Preprint number 4816.

[2]   S. H. Nielsen: "Auditory Distance Perception in Different Rooms", *Journal of the Audio Engineering Society*, Volume 41, 1993, pp. 755–770.

[3]   A. W. Bronkhorst and T. Houtgast: "Auditory distance perception in rooms", Nature, Volume 397, 1999, pp. 517-520.

[4]   Sylvain Choisel and Karin Zimmer: "A Pointing Technique with Visual Feedback for Sound Source Localization Experiments", presented at the AES 115th convention, New York, 2003, October 10–13. Preprint number 5904.

[5]   Pauli Minnaar: "Simulating an Acoustical Environment with Binaural Technology – investigations of binaural recording and synthesis", Ph.D. thesis obtained at Aalborg University, Denmark, September 2001.

[6]   Pauli Minnaar, Søren Krarup Olesen, Flemming Christensen and Henrik Møller: "Localization with Binaural Recordings from Artificial and Human Heads", *Journal of the Audio Engineering Society*, Volume 49, No. 5, 2001, pp. 323–336.

[7]   H. Møller, D. Hammershøi, C. B. Jensen and M. F. Sørensen: "Evaluation of artificial heads in listening tests", *Journal of the Audio Engineering Society*, Volume 47, No. 3, 1999, pp. 83–100.

[8]   J. Blauert: "Spatial hearing: The psychophysics of human sound localization", MIT Press., Cambridge, USA, Revised edition, 1997.

[9]   A. W. Bronkhorst: "Localization of Real and Virtual Sound Sources", *Journal of the Acoustical Society of America*, Volume 98, No. 5, 1995, pp. 2542–2553.

[10]   D. R. Begault, E. M. Wenzel, A. S. Lee and M. R. Anderson: "Direct comparison of the impact of head tracking, reverberation and individualized head-related transfer functions on the spatial perception of a virtual speech source", presented at the AES 108th Convention of the Audio Engineering Society, Paris, France, 2000, February 19–22. Preprint number 5134.

[11]   F. L. Wightman and D. J. Kistler: "Resolution of front-back ambiguity in spatial hearing by listener and source movement", *Journal of the Acoustical Society of America*, Volume 105, No. 5, 1999, pp. 2841–2853.

[12]   Company web site for AM3D A/S, including explanation of technology: www.am3d.com

[13]   Douglas S. Brungart *et al*: "The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources", Proceedings of ICAD 04- Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004.

[14]   Sheldon M. Ross: "Introduction to Probability and Statistics for Engineers and Scientists". ISBN 0-471-60815-7.