Evaluation of a Low-Cost 3D Sound System for Immersive Virtual Reality Training Systems

Kai-Uwe Doerr, *Member*, *IEEE*, Holger Rademacher, Silke Huesgen, and Wolfgang Kubbat

Abstract—Since Head Mounted Displays (HMD), datagloves, tracking systems, and powerful computer graphics resources are nowadays in an affordable price range, the usage of PC-based "Virtual Training Systems" becomes very attractive. However, due to the limited field of view of HMD devices, additional modalities have to be provided to benefit from 3D environments. A 3D sound simulation can improve the capabilities of VR systems dramatically. Unfortunately, realistic 3D sound simulations are expensive and demand a tremendous amount of computational power to calculate reverberation, occlusion, and obstruction effects. To use 3D sound in a PC-based training system as a way to direct and guide trainees to observe specific events in 3D space, a cheaper alternative has to be provided, so that a broader range of applications can take advantage of this modality. To address this issue, we focus in this paper on the evaluation of a low-cost 3D sound simulation that is capable of providing traceable 3D sound events. We describe our experimental system setup using conventional stereo headsets in combination with a tracked HMD device and present our results with regard to precision, speed, and used signal types for localizing simulated sound events in a virtual training environment.

Index Terms—3D sound simulation, virtual environments, virtual training.

1 Introduction

Some of today's simulations require the use of expensive, heavy, and large equipment. Examples include driving, shipping, and flight simulators with customized visual and motion systems that can only be maintained at specialized facilities. As an alternative, Virtual Reality (VR)-based training systems can be employed to simulate various 3D environments and provide a flexible and cost-effective platform for educational purposes. The F15 flight simulator [1], the Astronaut training to repair the Hubble Space Telescope [2], [3], and the submarine maneuvering trainer [4] are just a few examples of successful use of VR simulations.

In our project, we partially simulate an Airbus A340 cockpit in VR for pilot training [5]. All interaction devices such as side stick, pedals, thrust-levers, knobs, buttons, and dials are modeled as three-dimensional geometric objects. All other parts and surfaces are formed by images (textures). Critical devices such as side sticks, pedals, and thrust-levers are also physically available. All others are replaced by plastic panels to generate a forced feedback for the pilots [6]. A simplified outside visual is available to generate immersive flight simulations. Typically, certain essential training exercises for pilots such as cockpit familiarization and orientation tasks take place during expensive simulator hours. These tasks, however, can be outsourced to Virtual Training Systems (VTS) systems that are designed to provide the required 3D environments for these tasks. The problem with VTS is that due to the limited Field of View (FOV) of HMD systems, trainees can miss out on information that is provided through the training system if the event occurs outside the actual FOV. In this paper, we describe our approach to overcome this issue. The basic idea is to provide "attention getters" through the audio modality (channel) to draw the attention of the user into a specific location/direction.

The ability of people to localize sounds—whether in real or virtual space—has been extensively studied [7], [8], [9], [10]. It is a common practice in psychoacoustic testing to maintain close control over all relevant factors—the stimulus, the source characteristics, and the environment (often an anechoic chamber or a "room" with a single reflection). Two major findings are relevant to our discussion. One has to do with the recognition that room reverberation is very important for the externalization of virtual sounds [11], [12]. However, accurate room modelling is complicated and computationally expensive [13]. The other finding underscores the fact that to achieve accurate vertical localization with virtual sounds, it is necessary to measure individualized head related transfer functions (HRTF), which is also difficult and expensive [14]. Non-individualized HRTFs that are embedded in some commercial sound cards or used in computer games are often viewed as being inadequate. To generate and individualize HRTF's, the acoustic pressure close to the ear is measured for different sound source positions around the listener for an entire acoustic field. With these measurements, a person's HRTF can be computed, which is used to simulate (by filtering) the sound signal that a person can hear. The complex theory behind this methodology is described in Burgess [15].

- *K.-U. Doerr is with the California Institute for Telecommunications and Information Technology, 3100 Calit2 Building, Irvine, CA 92697-2800. E-mail: kdorr@uci.edu.*
- *H. Rademacher is with the Darmstadt University of Technology, Institute of Ergonomics, Petersenstraße 30, Room L1/01 508, D-64287 Darmstadt, Germany. E-mail: rademacher@iad.tu-darmstadt.de.*
- *S. Huesgen and W. Kubbat are with the Darmstadt University of Technology, Institute of Flight Systems and Automatic Control, Petersenstr. 30, D-64287 Darmstadt, Germany. E-mail: {huesgen, kubbat}@fsr.tu-darmstadt.de.*

Manuscript received 5 Mar. 2006; revised 8 July 2006; accepted 2 Oct. 2006; published online 10 Jan. 2007.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-0018-0306.

1077-2626/07/$25.00 © 2007 IEEE Published by the IEEE Computer Society

Meta Platforms, Inc.
Exhibit 1056
Page 001

TABLE 1
EAX Parameters Used

Attribute	Parameter
Environment	EAX ENVIRONMENT ROOM
Environment Size	3x3x3m
Diffusion	default value
Room and Room HF	high freq. filter (reflected sound) - default value
Decay Time	0.1 (for small rooms)
Decay HF Ratio	2.0 (higher frequencies for better localization)
Reflection	default value
Reflection Delay	0.02 (for closer reflection surfaces)

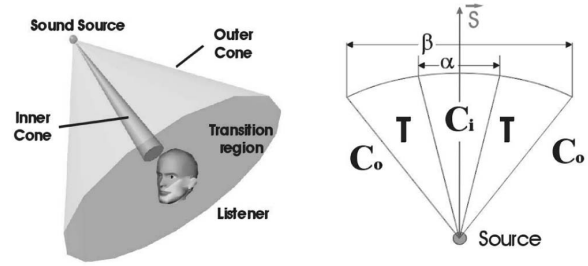


Fig. 1. 3D cone model.

The above described issues have left the impression that a system that provides useful spatial sound for VR applications has to be complex and expensive. In contrast, we hypothesize that it is possible to provide useful sound localization using a judicious combination of simplified models of sources, nonindividualized HRTFs, and rooms. Thus, in this study, we present experimental results that show the level of performance that can be achieved by a low-cost, software-based 3D sound system utilizing consumer-level 3D audio hardware interfaces. Here, we focus on providing a 3D sound simulation that enables users to localize sound events in 3D space (note: we do not focus on implementing a highly realistic 3D sound simulation). In the framework of this project, the usability of a low-cost, real-time 3D sound simulation is evaluated. More specifically, this investigation has two primary objectives: 1) to implement and evaluate a methodology to provide a simplified 3D sound simulation that enables precise and reliable event localization in space, and 2) to select an optimal sound signal type for the above described training scenario.

2 Preliminary Work

For the most part, a 3D sound simulation is used to enhance the gaming experience in computer games. In our approach, we implement and evaluate a 3D sound simulation that provides users the ability to localize and associate sound events with specific positions in 3D space. The purpose here is to use this 3D sound simulation as a way to guide and direct trainees in an environment with dimensions similar to a commercial aircraft cockpit.

2.1 Software Implementation (Sound Engine)

We implemented our sound engine based on the open source library OpenAL [16]. This library provides various models to manipulate sound source files by varying intensity level (volume/gain) and frequency depending on listener and source position in 3D space. OpenAL also takes advantage of the Environmental Audio Extensions (EAX) [17] available on certain sound cards. Although EAX is mostly used in computer games to provide enhanced sound experience, it also provides basic models for calculating sound reflection and reverberation effects in certain environments. These effects can positively affect the perception of sound events and their localization [10]. EAX can simulate these effects and provide a generic HRTF for a variety of environment models for game development. However, first experiments with the full capabilities of OpenAL and EAX negatively affected the ability to estimate source positions. Due to the superimposition of all available effects, test subjects got confused which led to unpredictable localization results. By reducing the number of effects available in both libraries, we developed a setup in which superimpositions were reduced to a minimum. Table 1 lists the effects and parameters we used from the EAX extension. For a complete description of all available parameters, we refer to the EAX user manual [17]. To localize sound sources in a 3D environment, we need to consider the position of the sound source as well as the direction in which the sound source itself emits sound. An object can emit sound in all directions with the same intensity or transmit sound unidirectionally. When a simulated sound source is used with a nondirectional characteristic, only the radiation pattern can be used as a clue to locate the position. Preliminary tests with nondirectional sound sources showed that the ability to estimate a source position in this case is limited to very rough regions. In addition, when multiple sound sources are used, superimposition again leads to problems when specific events related to a 3D position need to be localized. To avoid these problems, we use directional sound sources. OpenAL provides a *3D Cone Model* to simulate the characteristics of directional sources. The basic idea behind this model is to manipulate the volume (gain value) according to the position of the listener as shown in Fig. 1. The cone model consists of an inner and an outer cone. These regions can be defined with α and β as the respective cone angles for the inner and outer cone. If the listener is outside of both cones (C_o region), the sound file will be played with the lowest predefined volume. If the listener moves inside the outer cone (transition region T), the volume increases linearly until the position of the listener is inside the inner cone (C_i) where the maximum volume of the sound can be experienced. By defining the angles, the volume value outside of the cone and the directional vector $\vec{S}$, the model can be specified and adapted. We use this model for our experimental setup in addition to volume variation depending on distances between listener and sound source position. During previous tests, we determined that the inner cone angle α as 1° and the outer angle β as 90° provided the best results.

2.2 Signal Type Selection

The localization of a sound source position strongly depends on the type of signal that is being used. Findings from Schilling [18], Begault [10], [19], and Wenzel [20] concerning frequency spectra and time depending presentation (with respect to a reliable and precise position

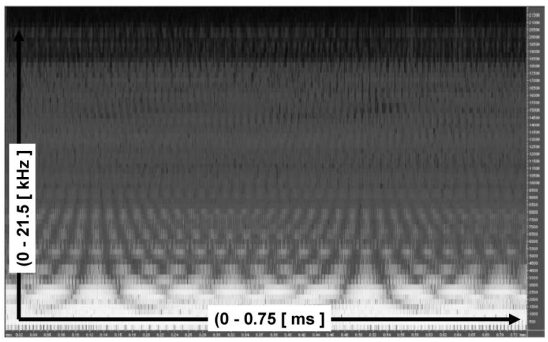


Fig. 2. Frequency spectrum for impulse and prolonged sound file.

localization of 3D sound sources), led to the following general requirements:

- Signals with broad frequency spectra are to be preferred.
- Signals with significant spectral components over 5kHz are to be preferred.
- Extremely short signals (< 0.5 ms) should be avoided.
- If speech signals are used, female voices are to be preferred (because of the higher frequency).

Taking into account these basic requirements, we selected the following signals for our experiments:

- Impulse signal (0.5 seconds).
- Prolonged signal (2 seconds).
- Speech signal (2 seconds).

Fig. 2 shows the spectrogram for the impulse signal and the prolonged signal, and displays the time variation of the spectral distribution of energy. The signal was generated on a keyboard (Model: Technics KN 750), playing a G-Dur accord (G-H-D-G). The bright areas represent regions with higher energy levels, while the dark regions represent low energy contents of the signal. The used signals exhibit recognizable energy contents up to 18kHz and their overall characteristic can be assumed as broad-banded (spectral components from 20Hz to 22kHz). For the impulse signal only a small portion of the sound file was used, while for the continuous signal, the whole file was used. Fig. 3 represents a section from the speech signal that was used. The signal represents the recorded word "attention" spoken by a female voice. The signal was selected because of the high energy content between 2.5 and 13.5 kHz. Due to the energy contents between 500Hz and 6.5kHz the signal can also be assumed as broad-banded.

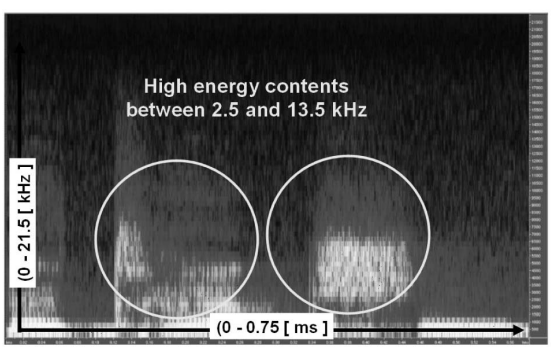


Fig. 3. Frequency spectrum for female speech sound.

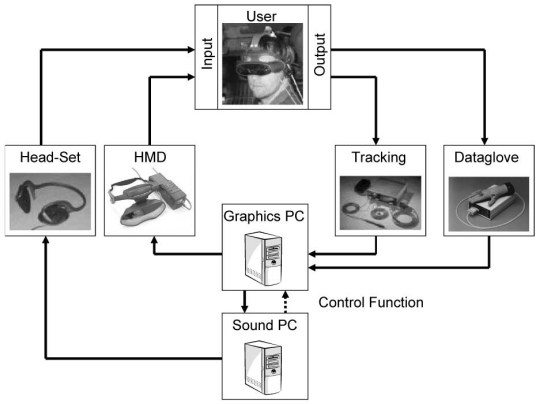


Fig. 4. Hardware setup (simulation components).

3 Experimental Setup

3.1 Hardware Components

The hardware setup is shown in Fig. 4. The user is equipped with a HMD and a Dataglove. The position of both devices is obtained from a magnetic tracking system. The Dataglove serves as a basic interaction device with the Virtual Reality (VR) system by detecting hand gestures. For the perception of audio events, the user is wearing a standard stereo headset. Table 2 describes the hardware components used. The visual and audio simulation is calculated on separate work stations to avoid potential delays due to the computational complexity. The graphic workstation is connected to the tracking system and passes the position information to the 3D sound engine running on a sound server machine. The sound engine determines the user/listener position in our sound simulation and calculates the responding convolution of the sound sources with respect to the movements of the user. The sound sources are stored as sound files using the .wav format. The position, type, and number of the sound sources can be specified through information stored in an ASCII text file. The alteration of the sound signals can be experienced by the users in real time, providing them with immediate feedback to their movements.

TABLE 2
Hardware Components

Device	Specification
HMD Sony LDI-100BE	800x600 resolution (mono)
Tracking Polhemus FASTTRACK	Range 76cm (max. 305cm)
	Res. 0.08cm,0.5°
CyberGlove	18 sensors
Headset	standard stereo

Meta Platforms, Inc.
Exhibit 1056
Page 003

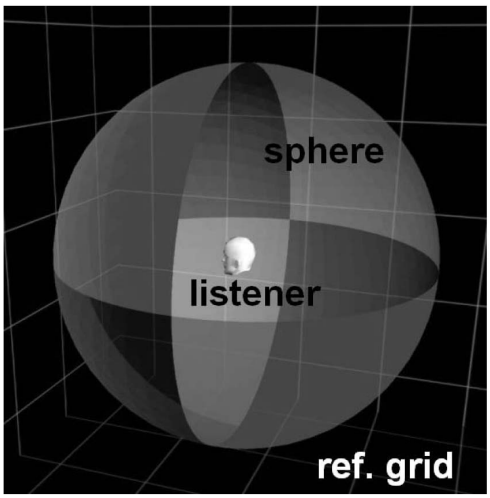


Fig. 5. 3D Test environment (sphere radius 1m).

3.2 Virtual Environment

Fig. 5 shows the visual environment presented to the user. The listener/user is placed in the center of a sphere that is divided in eight sectors. Each sector represents a different aural hemisphere as a position reference for the user. The sphere is surrounded by a cubical reference grid that allows additional depth perception. During the experiment, the simulated sound sources were placed on the surface of the sphere at different positions. For each localization only one source was presented facing toward the center of the sphere ($\vec{S} = (0, 0)$). The sound sources were not visible to the test persons. The task for the test subjects was to estimate a location on the sphere where the sound source was expected with respect to the experienced sound simulation. To interact with the test environment and to determine in which direction a subject was looking, we presented a target cross in the center of the FOV. Since our main focus in this test was to evaluate the precision, the time available for users to estimate the source position was limited to 20 seconds. We repeated the sound event over this time period with a small break (2 seconds) in between the signals. The time constraint was indicated to the test subjects by changing the color of the target cross within 15 seconds (Fig. 6). Once the test subjects were certain about the source position a simple hand gesture (fist) trigged the recording of time and viewing direction.

3.3 Analysis and Evaluation Methodology

To analyze the recorded data the coordinate system described in Fig. 7 was used. In the beginning of each

Fig. 6. Test subject vision and target behavior.

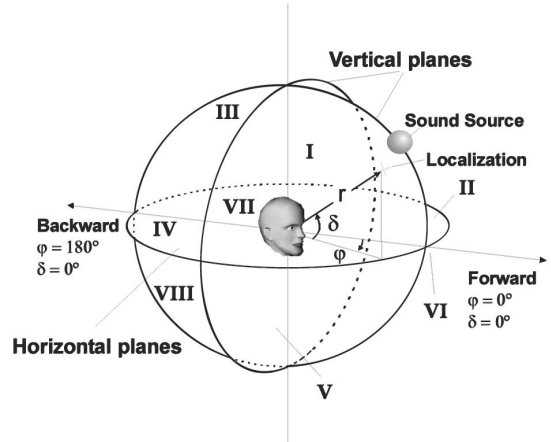


Fig. 7. Test coordinate system.

localization attempt, the test subject was asked to reset to a predefined viewing direction at $\delta = 0°$ and $\varphi = 0°$. To determine localization errors, additional effects like the difference between experienced and nonexperienced test subjects and the frequently reported phenomenon of front-back confusion need to be considered. For a detailed description of these effects, we refer to Wenzel et al. [21]. To calculate the precision of the user's localization, we calculated the deviation $\Delta\varphi$ and $\Delta\delta$ of the recorded viewing direction to the known sound source position as can be seen in (1), (2), (3), and (4).

$$\varphi = arccos\left(\frac{X_s}{\sqrt{X_s^2 + Y_s^2}}\right) \; with \; (0° \leq \varphi < 360°), \quad (1)$$

$$\delta = arcsin\left(\frac{Z_s}{\sqrt{X_s^2 + Y_s^2 + Z_s^2}}\right) \; with \; (0° \leq \delta < 360°), \quad (2)$$

$$\Delta\varphi = |\varphi_{source} - \varphi_{reported}|, \quad (3)$$

$$\Delta\delta = |\delta_{source} - \delta_{reported}|. \quad (4)$$

The normal distribution of the measurement was calculated with the chi-square test. The significance of our statements were proved with statistical analysis using ANOVA tests for repeated measurements in connection with Newman-Keuls-Post-Hoc-Test (N-K-Test) [22].

4 Results

Since it was not feasible to test all possible source positions on the sphere, we focused on selected positions that reflected the working space in a commercial aircraft cockpit for pilots. This reduction was also needed in order to limit the workload for the individual test subjects. It can be assumed that due to symmetry effects the results for the missing source positions are equivalent to their mirrored pendant. To be able to group sound sources in classes in the following diagrams, we used negative values for angles above $180°$ as indicated in Table 3. The following source positions were tested. This test grid provides a total amount

Meta Platforms, Inc.
Exhibit 1056
Page 004

TABLE 3
Test Grid

δ	$\varphi = 0°$	$\varphi = 45°$	$\varphi = -45° \equiv (315°)$
$0°$	X	X	X
$45°$	X	X	X
$90°$	X	X	X
$135°$	X	X	-
$-45° \equiv (315°)$	X	X	X
$-90° \equiv (270°)$	X	-	-
$-135° \equiv (225°)$	X	X	X

of 18 source positions. Each source-position/signal-type combination was tested three times. Therefore, we tested 162 $(18 \times 3 \times 3)$ sound events with each test subject. The sequence of the sound sources presentation was permuted and divided into six groups. Between each group of experiments, the test subjects had a break of two minutes. We tested 28 subjects (2 females and 26 males), which led to 4,536 single measurements for viewing direction and recognition time. The test subjects were informed that every sound source position on the sphere is possible. The results are represented by grouping the positions into the following classes:

- Frontal group $(\varphi \leq \pm 45°)$ group: F,
- Rear group $(\varphi = \pm 135°)$ group: R,
- Top group $(\delta = 45°)$ group: T,
- Middle group $(\delta = 0°)$ group: M, and
- Lower group $(\delta = -45°)$ group: L.

4.1 Sector Test

The first test we conducted was the *Sector Test*. The purpose of this test was to find out if it is possible to distinguish between sound events in the eight sectors. For this test, we used source positions in the center of each sector $(\pm 45°)$. As Fig. 8 and Table 4 show, all signals can be used to associate source position to the eight sectors. The mean percentage for a correct localization is above 90 percent. However, significant differences were found between the frontal and rear group of positions. In the rear group, the percentage of correct localizations was 7 percent lower. Considering the signal type, the impulse signal was significantly better than the other signals (93 percent). In the following sections, we especially focus on the horizontal and vertical localization precision. Since common HMD systems are limited in the FOV provided (horizontal 35-120° and vertical 20-35°), this evaluation will show whether it is possible to guide a test subject to look into a specific direction where a visual stimulus is presented.

TABLE 4
Sector Test Values

	prolonged	impulse	speech	all signals
mean	90%	93%	89%	91%
max	100%	100%	100%	100%
min	79%	71%	71%	78%

4.2 Horizontal Deviation

For the horizontal deviation, we compared first the result for signal types in the top, middle, and lower groups (TML). For all signal types, we found significant differences between the three groups except for the speech signal (see Fig. 9 and Table 5). The precision for the position localization in the middle group $(\delta = 0°)$ was the highest while positions in the lower group $(\delta = -45°)$ for the continuous and impulse signal led to nearly a doubling of deviations. For the speech signal no difference between the upper and lower signal groups could be detected. When we compared the results of the signal types within the groups (TML) we found that the speech signal in the top group was significantly worse than the other two signal types, with a mean deviation of 7.6°. It was also conspicuous that the minimum and maximum deviation for the top (T) and lower (L) group were much higher than for the middle (M) group, which indicates that sound sources presented above or below the users are more difficult to localize.

When the signals were compared between the front and the rear group (FR) a difference could also be detected (see Fig. 10 and Table 6). The sound sources from the front group $(\varphi \leq \pm 45°)$ were significantly better localizable than the sources from the rear group. When we compared the differences between the signal types in the front group, the speech signal was again significantly worse than both other signals. The source signal comparison in the rear group again leads to significant advantages for the impulse signal. Considering the maximum and minimum localization

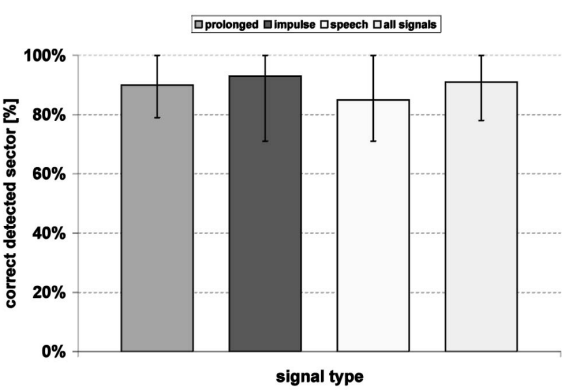


Fig. 8. Sector test (all signals), bars indicate minimum and maximum values.

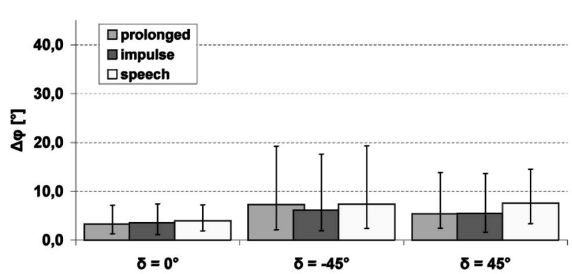


Fig. 9. Horizontal deviation (all signals, TML groups), bars indicate minimum and maximum values.

TABLE 5
Horizontal Deviation Group TML

$\Delta\varphi$ prolonged	$\delta = 0°$	$\delta = -45°$	$\delta = 45°$
mean	3.3°	7.3°	5.4°
min	1.3°	2.1°	2.4°
max	7.1°	19.2°	13.9°
$\Delta\varphi$ impulse	$\delta = 0°$	$\delta = -45°$	$\delta = 45°$
mean	3.6°	6.2°	5.5°
min	1.1°	1.9°	1.6°
max	7.4°	17.6°	13.7°
$\Delta\varphi$ speech	$\delta = 0°$	$\delta = -45°$	$\delta = 45°$
mean	4.0°	7.4°	7.6°
min	1.9°	2.4°	3.4°
max	7.2°	19.3°	14.5°

TABLE 6
Horizontal Deviation Group FR

$\Delta\varphi$ prolonged	$\varphi \leq \pm 45°$	$\varphi = \pm 135°$
mean	3.3°	7.3°
min	1.3°	2.1°
max	7.1°	19.2°
$\Delta\varphi$ impulse	$\varphi \leq \pm 45°$	$\varphi = \pm 135°$
mean	3.6°	6.2°
min	1.1°	1.9°
max	7.4°	17.6°
$\Delta\varphi$ speech	$\varphi \leq \pm 45°$	$\varphi = \pm 135°$
mean	4.0°	7.4°
min	1.9°	2.4°
max	7.2°	19.3°

deviations emphasizes the fact that the localization of sound sources in the rear group was less reliable than in the frontal group.

4.3 Vertical Deviation

We also analyzed our test results with respect to the precision of localizations in the vertical direction. This is even more critical since common HMD systems usually have a stronger limitation regarding the vertical viewing angle. We found again significant differences between the top, middle, and lower source position groups for each type of signal. Fig. 11 shows that the source positions in the middle group $(\delta = 0°)$ led to the best vertical precision results. The diagram also shows that if we compare sound sources in the top $(\delta = 45°)$ and lower $(\delta = -45°)$ group the results are not symmetric. Sound sources presented in the top group exhibit higher vertical deviation results than sound source positions in the lower group. The direct comparison between the sound sources in the respective position groups showed no differences between the signals (see Table 7).

A comparison of the results with respect to the front and rear position group shows a significantly better result for the estimation of positions in the front area (see Fig. 12 and Table 8). The results also indicate that none of the signals exhibits significantly better vertical localization results when compared to each other.

4.4 Localization Time

In addition to evaluating the localization precision in our sound simulation, we measured and compared the time the test subjects needed to estimate a sound source position. Although the localization time was not presented to the test subject as a crucial task, it can be used to identify difficulties in the localization task with respect to signal type and source position. Fig. 13 shows the results when signal types are compared in relation to the sound source positions in the TML group. A comparison of localization times shows that, for a specific signal, the localization time is independent of the source location.

Considering the localization time in respect to the signal types, we found that the prolonged sound type was significantly faster estimated than all other signals $(< 10s)$. The same result was found when the localization time was compared between the frontal and rear source group (see Fig. 14). None of the signal types exhibit differences between the localization time when compared between frontal and rear source positions. Again, the prolonged sound type was localized significantly faster than either impulse or speech signal.

5 Discussion and Conclusion

In this paper, we presented our approach to provide an inexpensive 3D sound simulation for VR training application. We specifically focused on providing a 3D sound environment to enable users to identify and localize sound events in 3D space, which can be used to guide trainees to pay attention to events outside of the FOV of an HMD system. Our experimental evaluation of the system demonstrates that

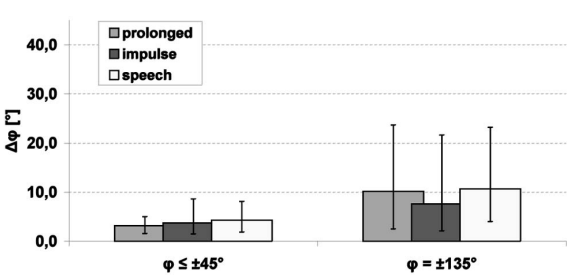


Fig. 10. Horizontal deviation (all signals, FR groups), bars indicate minimum and maximum values.

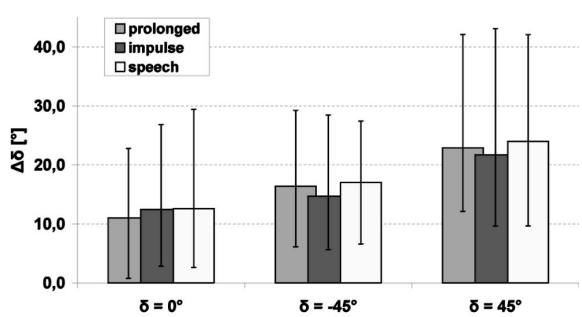


Fig. 11. Vertical deviation (all signals, TML groups), bars indicate minimum and maximum values.

TABLE 7
Vertical Deviation Group TML

$\Delta\delta$ prolonged	$\delta = 0°$	$\delta = -45°$	$\delta = 45°$
mean	11.0°	16.4°	22.9°
min	0.8°	6.1°	12.1°
max	22.8°	29.2°	42.1°
$\Delta\delta$ impulse	$\delta = 0°$	$\delta = -45°$	$\delta = 45°$
mean	12.4°	14.7°	21.7°
min	2.8°	5.7°	9.6°
max	26.8°	28.4°	43.1°
$\Delta\delta$ speech	$\delta = 0°$	$\delta = -45°$	$\delta = 45°$
mean	12.6°	17.0°	24.0°
min	2.6°	6.6°	9.7°
max	29.4°	27.4°	42.1°

TABLE 8
Vertical Deviation Group FR

$\Delta\delta$ prolonged	$\varphi \leq \pm 45°$	$\varphi = \pm 135°$
mean	15.1°	21.8°
min	7.5°	10.8°
max	24.3°	33.3°
$\Delta\delta$ impulse	$\varphi \leq \pm 45°$	$\varphi = \pm 135°$
mean	14.8°	19.4°
min	8.5°	8.0°
max	22.0°	30.0°
$\Delta\delta$ speech	$\varphi \leq \pm 45°$	$\varphi = \pm 135°$
mean	17.2°	20.8°
min	12.7°	11.0°
max	21.5°	29.6°

decent localization accuracy can be achieved by utilizing a software-based 3D sound system in combination with consumer-level 3D sound hardware as provided within commercial off-the-shelf sound cards.

Fig. 15 shows the overall performance of the test subjects for all tested source positions and sound signals with respect to the localization deviation of the simulated sound source position. Fig. 16 demonstrates the frequency of the measured displacements when the results are displayed and combined in $5° \times 5°$ subclasses. The density and frequency of smaller localization deviations in the range of $\Delta\varphi = \pm 10°$ show that the horizontal localization precision is much higher than the vertical precision, where the deviation varies between $\delta = +50°$ and $\delta = -50°$. To interpret the experimental outcome correctly, we also have to take into account that the results depend on sound source position and on the utilized sound source types. The results indicate that all of the signals can be used to identify the eight sectors of the sphere with a mean localization frequency above 90 percent.

In the case of the horizontal deviation, we found that the prolonged and the impulse signals lead to better localization results with a mean deviation of $\Delta\varphi = \pm 5.2°$ and $\Delta\varphi = \pm 5.5°$ in contrast to the speech signal with $\Delta\varphi = \pm 6.4°$. Concerning the vertical deviation, we found that the impulse signal was significantly better localizable with a mean deviation of $\Delta\delta = \pm 16.5°$. The prolonged signal and speech signal in this case lead to deviations of $\Delta\delta = \pm 17.1°$ and $\Delta\delta = \pm 18.2°$. We also compared the localization time needed for each signal and signal position. We found that for a specific signal type, the localization time was independent from the source positions. However, the time needed to localize the prolonged signal was less than 10 seconds, which was significantly shorter than the time needed to localize the other signals.

Our results confirmed previous findings that sound source positions presented in the frontal area $(\varphi <= \pm 45°)$ were more reliably localizable than sources presented in the rear sectors $(\varphi = \pm 135°)$. The results from the top middle and lower source position groups indicate that the middle positions $(\delta = 0°)$ led to significantly higher localization precision than the sound sources presented in the top and the lower position groups. Our experiments also showed that the top and lower position results were not symmetric. The lower source positions exhibited significantly better precision than the top source positions. This effect will be subject to further investigation. In addition, we found that the best signal to provide a localizable 3D sound event was the impulse signal, followed by the prolonged signal. The speech signal was inapplicable due to a significantly lower precision and higher dispersion. The horizontal localization precision is sufficient to fulfill the demands when an HMD is used with a horizontal FOV $>= 15°$. The reliable

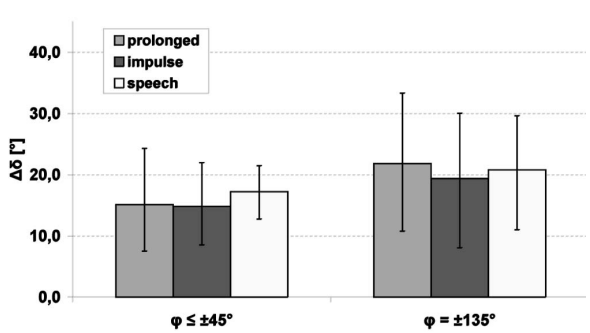


Fig. 12. Vertical deviation (all signals, FR groups), bars indicate minimum and maximum values.

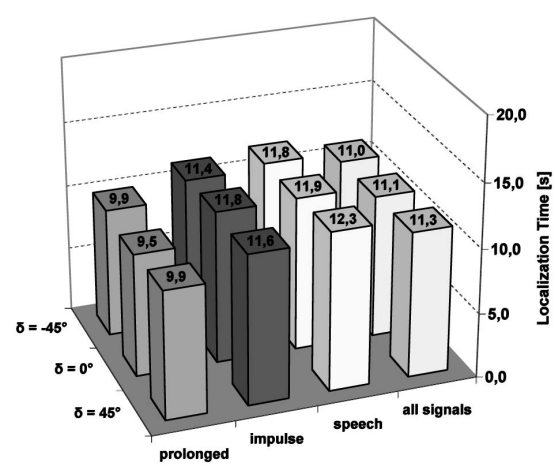


Fig. 13. Localization time (TML groups).

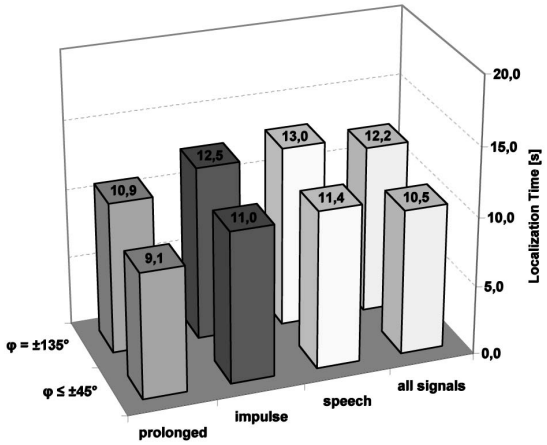


Fig. 14. Localization time (FR groups).

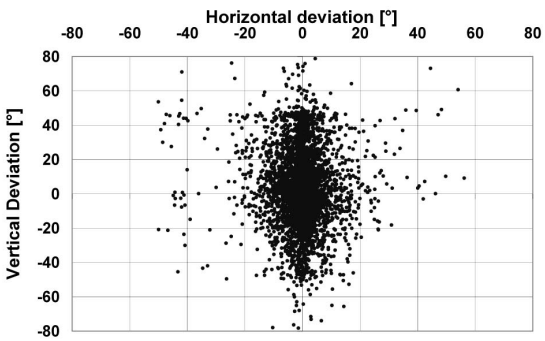


Fig. 15. Localization deviation (all signals and positions).

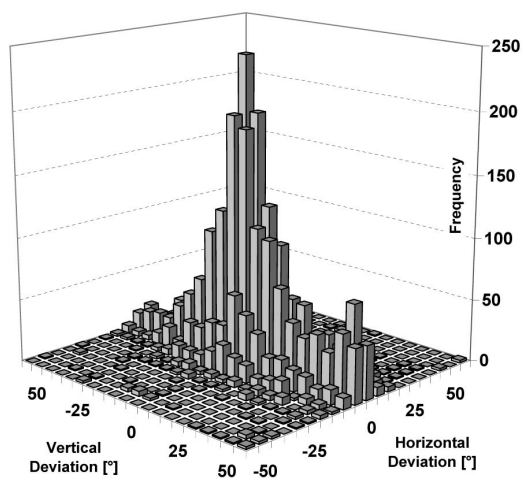


Fig. 16. Localization frequency.

guidance in vertical direction needs to be supported with an additional visual clue for today's HMD's with a vertical FOV between 20-35° depending on the signal being used.

Based on our results, we conclude that the developed 3D sound simulation is capable of providing basic 3D sound events to users who are equipped with a tracked HMD in combination with a conventional stereo head set, and of directing their attention to an object located outside of the HMD-given FOV. The localization accuracies reported indicate what can be achieved by a low-cost, software-based 3D sound system and that retrofitting a VR system with this type of sound simulation can be an affordable and valuable addition to these systems.

Acknowledgments

The authors would like to thank the Fraunhofer Institute for Computer Graphics (IGD) and their software distributor VRcom in Darmstadt, Germany, for supporting them with their Virtual Reality software package "Virtual Design II" [23]. Additional support was provided by members of the Environmental Audio Extensions (EAX) developer team.

References

[1] W.D. McCarty, S. Sheasby, P. Amburn, M.R. Stytz, and C. Switzer, "A Virtual Cockpit for a Distributed Interactive Simulation," *IEEE Computer Graphics and Applications,* vol. 14, no. 1, pp. 49-54, Jan. Feb. 1994.

[2] R.B. Loftin, "Virtual Reality for Aerospace Training," *VR Systems Magazine,* vol. 1, no. 2, pp. 36-38, 1996.

[3] R.B. Loftin and P. Kenney, "Training the Hubble Space Telescope Flight Team," *IEEE Computer Graphics and Applications,* vol. 15, no. 5, pp. 31-37, 1995.

[4] R.T. Hays and D.A. Vincenzi, "Fleet Assessments of a Virtual Reality Training System," *Military Psychology,* vol. 12, no. 3, pp. 161-186, 2000.

[5] K. Doerr, J. Schiefele, and W. Kubbat, "Virtual Simulation for Pilot Training," *Human Factors and Medicine Panel,* NATO's Research and Technology Organisation (RTO), Den Haag, The Netherlands, Apr. 2000.

[6] J. Schiefele, O. Albert, K. Doerr, and W. Kubbat, "Virtual Cockpit Simulation with Force Feedback for Prototyping and Training," Soc. for IMAGE Generation (SIG), Scotsdale, Ariz., Aug. 1998.

[7] J. Blauert and R.A. Butler, "Spatial Hearing: The Psychophysics of Human Sound Localization by Jens Blauert," *The J. Acoustical Soc. of Am.,* vol. 77, no. 1, pp. 334-335, 1985.

[8] S. Carlile, *Virtual Auditory Space: Generation and Application.* R.G. Landes Company, 1996.

[9] J.M. Loomis, C. Hebert, and J.G. Cicinelli, "Active Localization of Virtual Sounds," *The J. Acoustical Soc. of Am.,* vol. 88, no. 4, pp. 1757-1764, 1990.

[10] D.R. Begault, *3-D Sound for Virtual Reality and Multimedia.* Academic Press Professional, Inc., 1994.

[11] N.I. Durlach, A. Rigopulos, X.D. Pang, W.S. Woods, A. Kulkarni, H.S. Colburn, and E.M. Wenzel, "On the Externalization of Auditory Images," *Presence: Tele-Operators and Virtual Environments,* vol. 1, no. 2, pp. 251-257, 1992.

[12] D. Begault, "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems," *J. Audio Eng. Soc.,* vol. 40, no. 11, pp. 895-904, 1992.

[13] B. Blesser, "An Interdisciplinary Synthesis of Reverberation Viewpoints," *J. Audio Eng. Soc.,* vol. 49, no. 10, pp. 867-903, 2001.

[14] H. Moller, "Binaural Technique: Do We Need Individual Recordings?" *J. Audio Eng. Soc.,* vol. 44, no. 6, pp. 451-469, 1996.

[15] D.A. Burgess, "Techniques for Low Cost Spatial Audio," *Proc. Fifth Ann. ACM Symp. User Interface Software and Technology (UIST '92),* pp. 53-59, 1992.

[16] *OpenAL Specification and Reference,* Loki Software, June 2000, http://www.openal.org/.

[17] *Environmental Audio Extention: EAX 2.0,* Creative Technology Limited, 2001, http://developer.creative.com/.

[18] R.D. Shilling and B. Shinn-Cunningham, "Virtual Auditory Displays," *Handbook of Virtual Environments: Design, Implementation, and Applications,* K.M. Stanney, ed., chapter 4, Lawrence Erlbaum Assoc., Inc., 2000.

[19] D.R. Begault, "Challenges to the Successful Implementation of 3-D Sound," technical report, NASA-Ames Research Center, Moffett Field, Calif., 1990.

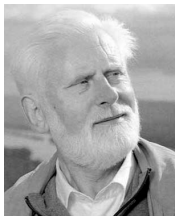
Meta Platforms, Inc.
Exhibit 1056
Page 008

[20] E.M. Wenzel, "Localization in Virtual Acoustic Displays," *Presence: Tele-Operators and Virtual Environments,* vol. 1, no. 1, pp. 80-107, 1992.
[21] E.M. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *The J. Acoustical Soc. of Am.,* vol. 94, no. 1, pp. 111-123, 1993.
[22] B. Winer, *Statistical Principles in Experimental Design,* third ed. McGraw-Hill Humanities/Social Sciences/Languages, 1991.
[23] P. Astheimer, "Virtual Design II: An Advanced VR System for Industrial Applications," Virtual Reality World, Stuttgart, pp. 337-363, 1995.

Kai-Uwe Doerr received the PhD degree from the Darmstadt University of Technology, Germany, in 2004. His expertise includes virtual cockpit simulation, virtual prototyping, computer vision, and 3D database generation. Currently, he is a postdoctoral researcher working jointly with the California Institute for Telecommunications and Information Technology (Calit2), and the Department of Electrical Engineering and Computer Science at the University of California, Irvine. His current work focuses on image-based tracking algorithms, cluster-based large-scale data visualization, and human factors research for interactive 3D visualization technologies. He is a member of the IEEE.

Holger Rademacher received the master's degree in business engineering with mechanical engineering as a technical field from the Darmstadt University of Technology, Germany, in 2005. He conducted the project presented in this paper during his technical study thesis. His expertise includes 3D sound simulation, virtual cockpit simulations, experimental statistics, and human factors research in driver distraction by automotive information and communication systems.

Silke Huesgen studied mechanical engineering at the Darmstadt University of Technology, Germany. Since 2001, she has been working as a PhD student at the Institute of Flight Systems and Automatic Control (FSR) at TU Darmstadt. Her field of activity is comprised of virtual cockpit simulations, avatar animation, and 3D object modeling. Her research mainly focuses on the development of virtual training environments which support and improve system knowledge.



Wolfgang Kubbat received the PhD degree from the Braunschweig University of Technology, Germany, in 1969. His professional background is comprised of three years of research at the Deutsches Zentrum für Luft- und Raumfahrt (DLR) Braunschweig, Germany, 13 years of industrial development work at former MBB Military Aircraft Division Munich, where he finally was head flight guidance & control, and 26 years as head of the Institute for Flight Systems & Control Systems at the Darmstadt University of Technology. He retired in September 2005.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.

Meta Platforms, Inc.
Exhibit 1056
Page 009