In: Nordahl, R. (Ed.): HAID 2010, LNCS 6306, Berlin : Springer-Verlag, 46-54

Audioworld: a Spatial Audio Tool for Acoustic and Cognitive Learning

André Melzer[1], Martin Christof Kindsmüller[2] and Michael Herczeg[2]

[1] Université du Luxembourg, Campus Walferdange, L-7201 Walferdange, Luxembourg
andre.melzer@uni.lu
[2] University of Luebeck, Institute for Multimedia and Interactive Systems,
Ratzeburger Allee 160, D-23538 Luebeck, Germany
{mck, herczeg}@imis.uni-luebeck.de

Abstract. The present paper introduces *Audioworld*, a novel game-like application for goal-oriented computer-supported learning (CSL). In *Audioworld*, participants localize sound emitting objects depending on their spatial position. *Audioworld* serves as a flexible low cost test bed for a broad range of human cognitive functions. This comprises the systematic training of spatial navigation and localization skills, but also of verbal skills and phonetic knowledge known to be essential in grammar literacy, for example. The general applicability of *Audioworld* was confirmed in a pilot study: users rated the overall application concept novel, entertaining, and rewarding.

Keywords: Audio-based localization, computer-supported learning, human cognitive functions, spatial navigation.

1 Introduction

Multimodal human-computer interfaces make use of the synergetic effects of human vision and auditory perception. In visual tasks, for example, sound enables the user to monitor the state of background processes while remaining focused on the visual task [1, 2]. Sound thus reduces the demands on visual attention and increases the efficiency in common multiple task situations during computer use (e.g., [3]). In addition to the complementary function of hearing, audio enhanced interfaces grant access to digital information when visual input is either not available or impossible (e.g., text-to-speech utilities for users with special needs [4]).

The present study continues and extends the latter line of research, yet focuses on intact vision. We introduce Audioworld, a software application for standard consumer hardware, which aims at providing a tool for multifaceted audio-based learning in a game-like environment. In Audioworld, sound is the main sensory mode for navigation and orientation. By using their spatial hearing abilities, players navigate a desktop virtual environment to localize sound emitting objects (e.g., musical instruments or verbal information).

2 Binaural Hearing, Modeling, and Rendering Spatial Audio

Processing spatially presented auditory cues requires binaural hearing, which mirrors stereoscopic vision [5, 6]. Interaural Time and Level Differences (ITD, ILD) are the essential cues in human binaural source localization. These cues arise from the separation of the ears on the head and their resonant and reflective anatomical characteristics. ITDs and ILDs are affected by stimulus features like frequency and intensity, offsets in the interaural baseline, and imposition of amplitude modulation.
Spatial audio in virtual 3D environments (VE) is either based on the perceptual systems or the physical systems approach [7, 8]. Physical auditory VE systems require direct sound and early reflections to be calculated on the basis of environment geometry and acoustic properties of the individual listener. Perceptual auditory VE systems employ signal-processing algorithms without specific individualized physical models. Volume panning methods model a sound field with different intensities according to the direction of the virtual source [7, 8]. Volume panning is implemented in most toolkits targeted at the consumer entertainment market (e.g., Microsoft's DirectSound®).

Though volume panning has certain limitations (e.g., see [7]), the *Audioworld* application presented in this paper is based on this method. This was mainly done because we were focusing on evoking specific auditory perceptions, not on physical accuracy. In addition, it was our goal to use only off-the-shelf hardware.

3 The *Audioworld* Concept

The main task in *Audioworld* requires localization of meaningful sound emitting objects distributed across a spatial 2D array. Sound localization is the result of complex cognitive processing. It requires binaural hearing in order to determine the direction and distance to a target object. Additionally, distracting noise or competing sounds need to be suppressed in order to maintain the focus of attention on the auditory target dimension specified by the task. Finally, actions have to be initiated towards the target object location (i.e., moving one's own body or the body of an avatar). At the same time, movements towards lures (e.g., irrelevant sound emitting objects) have to be avoided.

3.1 Technological Aspects

We designed *Audioworld* using the 3D Game Studio (3DGS; Conitec, http://www.conitec.com) authoring suite for 2D and 3D real-time applications based on Microsoft's DirectX® graphics rendering standard. *Audioworld* uses a first person camera perspective. In addition to its binaural hearing requirements, the first person perspective specifically engenders psychological identification with the main acting character, a major aspect of cognitive and emotional learning (i.e., perspective taking [9]).

3.2 Main Function

The center of the *Audioworld* GUI's main menu shows a level preview, and two level select buttons. The button row located at the bottom of the screen comprises start, options, name of player, and quit (Fig. 1). Options refer to level and sound options. Level settings include an adjustable counter for a time limit, the item positioning function, and the sound item list that includes an "add" button. Additional wave files for sound items may be loaded into the system. *Audioworld* automatically loops wave files for continuous playback.

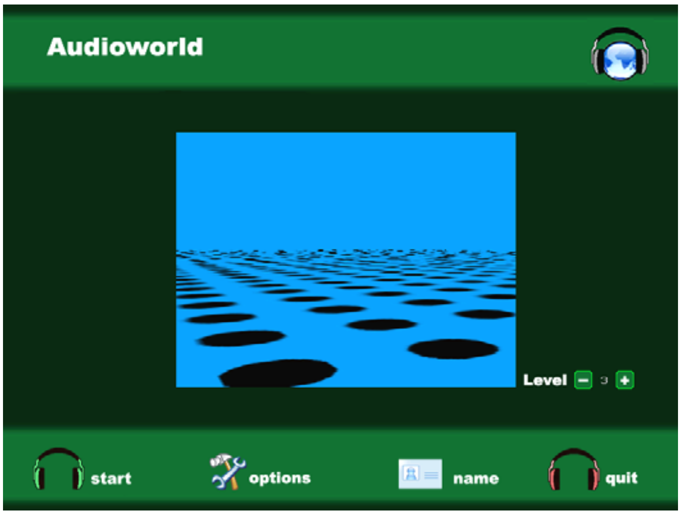

Fig. 1. Main menu in *Audioworld* with a preview of the Syllables game level.

Sound item positioning may be done by randomly distributing items across the two-dimensional level map ("random"), or by using automatic positioning according to a fixed spatial array ("default"). Manually placing ("drag-and-drop") selected sound items on the 2D map preview supports systematic individual learning ("rearrange", c.f. Fig. 2). The volume panning in *Audioworld* models a sound field with different intensities according to the direction of each sound item, which renders a realistic spatial audio impression.

Sound options were conceived to enable gradual and individual learning by adjusting task difficulty. Hence, sound options comprise a three-step adjustment of the sonic radius for sound items, and a collection of distracting sounds (e.g., street noise, pink and white noise, schoolyard noise), which are independent of the player's spatial position. Like sound items, new distracting sounds may also be loaded into the system.

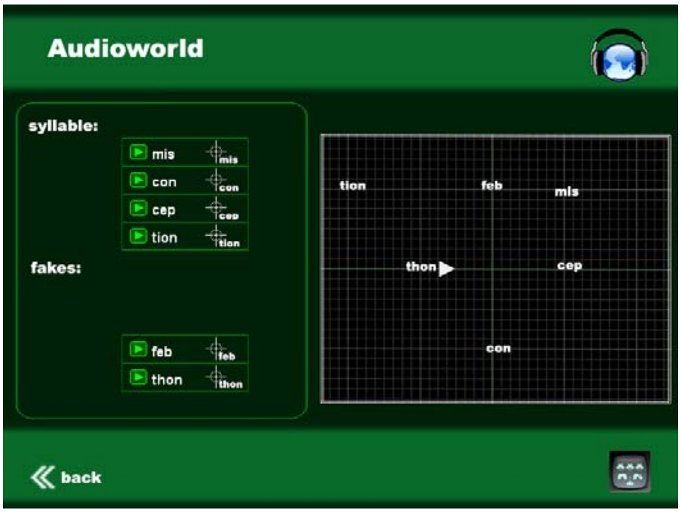

Fig. 2. Manual “drag-and-drop” positioning in Audioworld.

We included distracting sounds because sound localization requires suppression of competing sounds. Hence, additional processes are needed to focus attention. The detrimental effect of additional noise or competing sound on various cognitive tasks is well documented (e.g. [10]). Thus, and compared to a control condition, additional distracting sounds make localization of target sounds more difficult. This is typically reflected by behavioral dependent measures. Hence, distracting sounds both prolong the time needed for task completion and increase the number of errors compared to a control condition without distracting sounds. We tested this hypothesis and the overall applicability of *Audioworld* in a pilot study.

4 Pilot Study

A total of fourteen participants took part in the pilot study. By using the arrow keys on the computer keyboard, they navigated game levels and localized sound emitting target objects as quickly as possible. Time needed for completion and errors were logged automatically. At the end of the experiment participants answered a questionnaire. Participants rated the playability for each game level. Also, they estimated the game’s effect on their personal level of motivation. Finally, participants rated how difficult it was to localize sound objects on the basis of auditory perceptions.

4.1 Design, Stimuli, and Apparatus

The pilot study used a 2x2 mixed factorial design. Distracting sound was the between-subjects factor. Half of the participants were presented with a task-irrelevant distracting sound (schoolyard noise) throughout levels, while the other half was not distracted by irrelevant sound. Game level (nonspeech-based: Music World, speech-based: Syllables) was manipulated within subjects, that is, each participant played both game levels.

We designed two different game levels and one training level (see below). Four different sound emitting objects were used per level. Objects (i.e., instrument sounds or syllables, see below) were distributed across each level map such that a square-like configuration would have resulted if objects had been interconnected. Hence, item positioning was fixed and nearly identical across levels.

4.1.1 Dark World (Training Level)

This level was exclusively used for training purposes. Players navigated an open arena, which was surrounded by an enclosure. Both enclosure and floor consisted of abstract and dark textures. The four target sound objects were invisibly placed on the map (i.e., copier, water tap, telephone, and washing machine). Hence, object localization in Dark World required participants to rely on their hearing sense. Each target location emitted the typical sound of the assigned object (e.g., water dripping from the tap, telephone ringing). Characteristic object sounds were therefore used as auditory icons (earcons) [1, 11, 8].

4.1.2 Music World (Game Level 1)

This game level was also based on the map of Dark World. Yet, it provided even less visual information. This was achieved by adding plain white textures to the level. Therefore, the only visual information came from a ground evenly textured with black dots (c.f. Fig. 1). Every movement of the player caused an immediate and coherent displacement of the dots. The resulting optical flow induced an impression of "walking".

In Music World, sound objects were again invisible and presented as auditory icons. We used prerecorded sounds of musical instruments (e.g., drums, guitar, and vibes). The target sound objects were presented as isolated musical instruments that were part of a looped musical pattern. Participants listened to this instrumental pop tune played by all instruments together before starting Music World.

4.1.3 Syllables (Game Level 2)

The second game level was a variation of Music World. It also provided only minimum visual information, but appeared in light blue color (Fig. 1). Instead of using instruments as constituents of a coherent song, the German word "Missverständnis" (i.e., misconception) was split into its syllables "miss–ver–ständ–nis" that formed four different target objects. The syllables were spoken by a male voice and then looped. Before collecting the syllables, participants listened to the intact word several times.

4.1.4 Apparatus

Standard 14" notebooks (HP Compaq Evo N620C) equipped with stereo headphones (Stagg SHP-2200) were used.

4.2 Participants

Participants were randomly assigned one of two experimental conditions. In the distracting sound condition, participants' (2 female and 5 male) mean age was 28. In the non-distracting condition, participants' (7 males) mean age was 34. All participants were naïve to the subject of the experiment. They reported normal hearing abilities and spoke German as first language. No participant indicated problems with reading or writing skills (e.g., dyslexia).

4.3 Procedure

In a training session, participants made themselves familiar with Audioworld. This included the handling of game controls for navigation and the game's GUI. Most importantly, they learned to localize sound objects on the basis of their binaural hearing. Participants moved forward (up arrow on standard keyboard), backward (down arrow), turned left (left arrow), or turned right (right arrow) depending on their acoustic perceptions of sound intensity and direction. Localizing (i.e., "running into") one of the invisible target sound objects immediately triggered verbal feedback (e.g., "excellent") spoken by a pre-recorded male voice.

At test, participants started either with Music World or Syllables. In the distracting sound condition, participants selected "schoolyard noise". Participants in the non-distracting condition turned off the distracting sound function.

In Music World, two additional musical instruments (i.e., sitar and double bass) appeared as lures. These "false instruments" fitted the harmonies of the song, but sounded inappropriate for the pop tune. In Syllables, two extra spoken German syllables (i.e., "ge" and "da") were added as lures. "Running into" lures always triggered an immediate short two-tone failure sound and was coded as an error. After participants had finished test levels, they were instructed to fill in the questionnaire.

5 Results

First, level completion data will be reported. Then, results of the questionnaire will be presented.

5.1 Time Needed for Level Completion

The 2x2 mixed factorial analysis of variance performed on the mean time scores needed for game level completion indicated no significant result[1] (*F*s≤1.73). It took participants longer, albeit not significantly, to complete Syllables (*M*=205 sec, *SD*=75.80) compared to Music World (*M*=175 sec, *SD*=77.63). Participants in the

[1] For all statistical analyses, the alpha value was set at .05.

distracting sound condition (*M*=192 sec, *SD*=87.71) were not slower than participants in the non-distracting condition (*M*=189 sec, *SD*=67.34). Figure 4 illustrates the mean time scores in the different conditions. The failure to observe significant effects, especially with game level, is probably due to the small overall number of participants.

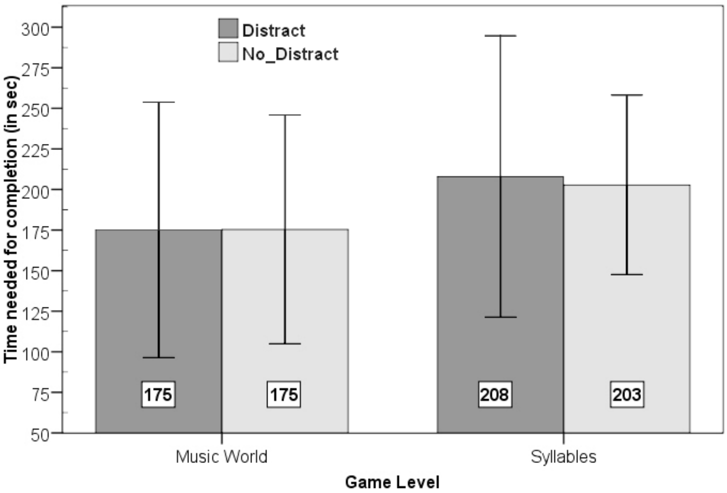

Fig. 3. Means (in sec) for task completion. Error bars represent 95% confidence intervals.

5.2 Errors

For Music World, there were only 6 errors in the non-distracting condition and only one error in the distracting sound condition. For Syllables, 4 errors were recorded in each distracting condition. Due to the overall low number of errors we cannot rule out the possibility that participants might have applied different strategies that caused a distortion in time scores. Hence, careful avoidance of lures (i.e., false instruments or syllables) may have resulted in slow game play. Again, further testing with a greater number of participants is needed.

5.3 Questionnaire

Participants' responses were coded for quantitative analyses. A four-point Likert scale was used. Lower numerical values indicated stronger agreement (e.g., 1=*strongly agree* to 4=*strongly disagree*). Scores were calculated by taking the mean of the scores given for each condition. Preliminary analyses indicated no substantial differences between distracting sound conditions. Therefore, data were collapsed across conditions.

5.3.1 Navigation

Participants reported that it was *“very easy”* to *“easy”* to use the arrow keys for navigation (*M*=1.64, *SD*=0.74). This is important, because it suggests that transferring already existing navigation skills to game controls in *Audioworld* was easy for participants. The findings from the behavioral data thus seem to reflect the intended task of locating sound objects via the auditory sense, rather than the participants’ struggle with game controls.

5.3.2 Gaming Enjoyment, Motivation, and Overall Concept

Participants apparently enjoyed playing *Audioworld* (*M*=1.50, *SD*=0.52). In addition, they felt motivated to continue playing the game with additional levels (*M*=1.71, *SD*=0.73). Finally, the overall concept was considered novel (*M*=1.57, *SD*=0.65).

6 Discussion

Results clearly indicated a good acceptance of *Audioworld*. The concept received good to excellent ratings in terms of handling, degree of difficulty, and navigation. There was only a numerical advantage for the non-speech sounds in Music World compared to the verbal information in Syllables: it took participants slightly longer to “collect” the target syllables than the musical instruments. Based on the literature, we had expected an advantage for non-speech sounds because these sounds are good at giving rapid feedback on actions, and quickly presenting continuous data and highly structured information. In contrast, speech in general is preferred for giving instructions and absolute values [1]. In addition, the verbal task in Syllables was even more difficult, due to the higher mental workload. That is, participants had to keep in mind the correct sequence of the syllables, and they had to recall which syllable would come next. As mentioned above, this finding might reflect overall low power of the experiment.

Finally, we have to address the null finding in the analyses performed on the distracting sound variable. Localizing sound objects requires active and effortful suppression of distractive noise or competing sounds (e.g., [10]). Hence, we had speculated that additional time-consuming and error-prone processes would be needed to focus attention on the acoustic target object. Apparently, this is not true for distractive noise per se. We may speculate that the distracting sound in our study was simply unobtrusive. Therefore, if we (a) had increased distracting noise level, or (b) had devised a more difficult task, we might have obtained a negative effect of distracting sound. However, this will be subject to future testing.

7 Future Work and Conclusion

Interestingly, though instructions were intentionally lacking any information on efficient strategies to localize target sound objects, participants uniformly reported having developed an almost identical strategy. By using the arrow keys they adjusted their position so that both ears received simultaneous sound of similar intensity. Then, they increased the perceived volume level by pressing the arrow keys to approach the sound source. Clearly, participants were actively and deliberately using the hallmarks of binaural hearing, namely the ITDs and ILDs [5]. *Audioworld's* game controls led to a high sense of perceived achievement, which fitted the natural form of sound localization.

We believe that the *Audioworld* concept fits the demands of different learning contexts. First, *Audioworld* may be adapted for entertainment purposes. Several audio games have been successfully introduced recently [12]. Most importantly, *Audioworld* may be used in various learning scenarios. In our pilot study, Music World and Syllables seem particularly promising in this regard. Music World, for example, refers to the learning of music concepts and structures. By using Audioworld, students may learn to differentiate musical instruments, pitch, or scales on the basis of their hearing sense. Using Syllables, *Audioworld* may support the acquisition of reading and writing skills. In a playful context, students may learn to focus on the phonetic and spatial aspects of verbal material (i.e., the individual sounds of verbal constituents), which has emerged as a key component in grammar literacy and is considered to be deficient in dyslexia, for example [13].

Taken together, the innovative and generic audio game concept of *Audioworld* supports learning because it induces continuous high levels of motivation needed for learning; an important, and often neglected, quality factor not only in experimental research but particularly in CSL.

References

1. Brewster, S.: Nonspeech auditory input. In: Jacko, J.A., Sears, A. (eds.) The Human-Computer Interaction Handbook. pp. 220--239. Lawrence Erlbaum Associates, Inc., Mahwah (2003)
2. Fang, X., Xu, S., Brzezinski, J., Chan, S.C.,: A Study of the Feasibility and Effectiveness of Dual-Modal Information Presentations, International Journal of Human-Computer Interaction, 20, 3--17 (2006)
3. Shneiderman, B., Plaisant, C.: Designing the User Interface (4th ed.). Pearson, Boston (2005)
4. Kennel, A., Perrochon, L., Darvishi, A.: WAB: World Wide Web Access for Blind and Visually Impaired Computer Users. In: Proc. of ACM SIGCAPH, 55, pp. 10--15 ACM Press, New York (1996)
5. Blauert, J.: Spatial Hearing: The Psychophysics of Human Sound Localization. The MIT Press, Cambridge (1997)
6. Braasch, J.: Modelling of Binaural Hearing. In: Blauert, J. (ed.) Communication Acoustics, pp. 75--103, Springer, Berlin (2005)
7. Naef, M., Staadt, O., Gross, M.: Spatialized Audio Rendering for Immersive Virtual Environments. In Proc. of VRST'02, pp. 65--72. ACM Press, New York (2002)
8. Novo, P.: Auditory Virtual Environments. In: Blauert, J. (ed.) Communication Acoustics, pp. 277--297. Springer, Berlin (2005)
9. Flavell, J.H., Botkin, P.T., Fry, C.L., Wright, J.W., Jarvis, P.E.: The development of role-taking and communication skills in children. Wiley, New York (1968)
10. Baddeley, A.: Working Memory. Oxford University Press, Oxford (1986)
11. Mynatt, E.D.: Designing with Auditory Icons: How Well Do We Identify Auditory Cues?. In: Proc. of CHI'94, pp. 269--270. ACM Press, New York (1996)

12. Friberg, J., Gärdenfors, D.: Audio Games: New Perspectives on Game Audio. In Proc. of ACE'04, June 3-5, 2004, Singapore, pp. 148--154. ACM Press, New York (2004)
13. Snowling, M.J.: Phonemic Deficits in Developmental Dyslexia, Psychological Research, 43(2), 219--234 (1981)