Supporting Sounds: Design and Evaluation of an Audio-Haptic Interface

Emma Murphy[1,*], Camille Moussette[2], Charles Verron[3,**],
and Catherine Guastavino[1]

[1] Centre for Interdisciplinary Research in Music Media & Technology and School of Information Studies, McGill University, Canada
emma.murphy@dcu.ie
[2] Umeå Institute of Design, Sweden
[3] Laboratoire de Mécanique et d'Acoustique de Marseille, France

Abstract. The design and evaluation of a multimodal interface is presented in order to investigate how spatial audio and haptic feedback can be used to convey the navigational structure of a virtual environment. The non-visual 3D virtual environment is composed of a number of parallel planes with either horizontal or vertical orientations. The interface was evaluated using a target-finding task to explore how auditory feedback can be used in isolation or combined with haptic feedback for navigation. Twenty-three users were asked to locate targets using auditory feedback in the virtual structure across both horizontal and vertical orientations of the planes, with and without haptic feedback. Findings from the evaluation experiment reveal that users performed the task faster in the bi-modal conditions (with combined auditory and haptic feedback) with a horizontal orientation of the virtual planes.

Keywords: Auditory feedback, Haptic feedback, Target-finding, User Evaluation.

1 Introduction

Despite ongoing research on the evaluation of systems that employ audio-haptic feedback, there is still much to be explored in the context of human-computer interaction. The present study investigates the relative contribution of audio and audio-haptic cues on performance and perceived usability, using a non-visual target finding task. Previous research has investigated the addition of haptics to visual interfaces for target-finding [1, 2]. Kim and Kwon [3] implemented a haptic and audio grid in order to enhance recognition for ambiguous visual depth cues. The authors implemented a haptic vertical grid and pitch variation to convey a target location to users and subsequent evaluations revealed that the multimodal cues increased precision, particularly in the vertical axis [3]. Magnusson and Grohn [4] evaluated a set of audio-haptic

---

[*] Currently at Dublin City University, Ireland.
[**] Currently at REVES-INRIA, Sophia Antipolis, France.

C. Magnusson, D. Szymczak, and S. Brewster (Eds.): HAID 2012, LNCS 7468, pp. 11–20, 2012.

feedback cues using a Phantom Omni haptic device with 3D auditory cues for visually impaired to locate non-visual objects in a virtual environment using a memory game task. Findings from this study show that this technique of 3D sound mapping aided users' understating of the spatial layout of the virtual space. Furthermore the addition of 3D sound cues to a virtual interface with visual and haptic cues was shown to improve collaboration between users [4] by enhancing awareness of ongoing work between participants. In [5] an audio-haptic interface was evaluated to investigate if users could integrate information from audio and haptic sensory channels in order to achieve a target selection task. The results of this study revealed that users had a preference for haptic feedback, even when audio could be considered better suited to the task. Ménélas et al. [7] investigated the addition of audio, haptic and combined audio-haptic to enhance target finding tasks in visual 3D environments with multiple and obscured targets. The findings from this study illustrated that haptic feedback and combined audio haptic feedback were more effective when compared to the audio-only condition to enhance the visual cues. The use of a full loudspeaker array is not always feasible for the design of user interfaces and binaural synthesis of 3D sound [8] over headphones is often more appropriate. Gonot [9] et al. has investigated the use of binaural synthesis in comparison to stereophonic implementations for navigating in constrained virtual environments. Findings from this study reveal that binaural synthesis can yield better results in terms of usability, cognitive load and subjective evaluation. Wall et al. [10] explored the potential of haptic feedback in the form of virtual magnetic cues to enhance a visual target finding task in comparison to the addition of stereo visual graphics. The magnetic effect applied to draw the user towards the target did not enhance participant's temporal performance although it did improve accuracy, while the stereo graphic condition improved both user accuracy and timings. Hwang et al [11] have illustrated that haptic force feedback in the form of gravity wells can enhance user performance for both time and error rates in multi target-finding tasks.

2 Methods

The virtual environment was composed of five equally spaced parallel planes. Practical applications for this virtual structure include the design of a non-visual menu or browser interface. By including five planes, the structure represented an interesting stack-like configuration, and it matched reasonably well with the physical constraints of the haptic device. Furthermore a restricted visual space could be increased on a small screen device by using the virtual planes to represent alternative views. Two configurations have been considered, with either horizontal or vertical planes. A virtual target was located at a random location on one of the planes as shown in Figure 1.

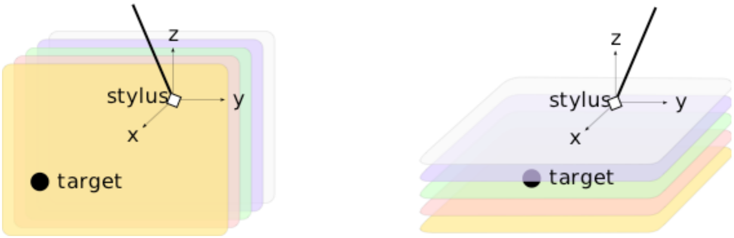

Fig. 1. Structure of 3D planes in both vertical and horizontal orientations with target located randomly on one of the planes

In order to find the target the user had to navigate through the planes and identify which one contained the target. A virtual bowed sound was played when the target and the stylus of the haptic device were located on the same plane (horizontal or vertical, according to the configuration). By identifying the position of the virtual sound source, the listener could navigate on the plane and locate the target. The aim of the experiment was to compare the differences in usability and navigation using the two different orientations of the virtual structure: horizontal and vertical, and the two different types of feedback: audio and audio-haptic. Specifically, we investigated whether it would be possible to navigate the interface and find the target without the support of the haptic planes. Due to the fact that the interface was not designed with redundant information between modalities we hypothesized that it would be more difficult for users to navigate the environment without haptic feedback.

2.1 Haptic Feedback

Haptic effects were designed and controlled using H3D[1], an open source haptic graphic API based on X3D[2], and a Phantom Omni haptic device from Sensable Technologies Inc. This desktop unit offers 6 degrees of freedom with a stylus-type grip and provides a workspace area of 160 W x 120 H x 70 D mm.

The five virtual planes were positioned at 10 mm intervals centered on the device's workspace origin. The planes were large enough to completely fill the workspace area on the other axes. This compact arrangement allows us to rotate the stack while keeping the spacing and haptic configuration intact. The interval distance of 10 mm between the planes seemed appropriate and adequate after initial pilot testing. It was the largest value possible to work with both orientations (the Omni depth axis offers only 70 mm of travel). Beyond our rendered haptic features, the users were naturally limited to the mechanical limits of the device in all three axes. The haptic feedback was active only on the planes and the target. Between the planes no forces were applied. The haptic rendering used a magnetic effect from Sensable's OpenHaptics library (MagneticSurface[3]) to create semi-rigid surfaces to hold users to the five planes while

---

[1] `http://www.h3d.org`

[2] `http://www.web3d.org/x3d/specifications`

[3] `http://www.h3dapi.org/modules/apidoc/html/classH3D_1_1MagneticSurface.html`

they were navigating the virtual environment. The magnetic effect consisted of forces (F=-kx) generated in order to keep the point of device's stylus on the defined surfaces. When the pointer was pulled outside a specific delta distance from the planes, it was freed from the magnetic attraction. In our experiment, the haptic planes were built with a stiffness index of 0.4 and 2 mm for the snap distance setting. The randomly positioned target consisted of a 4 mm radius sphere with a high surface friction value and no magnetic force applied. Contact between the pointer and the target's surface automatically registered the successful completion of the task and triggered the corresponding auditory cue. Note that the Phantom Omni can produce a maximum exertable force of 3.3 Newtons. Only the positional X, Y and Z data of the arm were used for the study. The device was situated on a desk and we left the grip position of the stylus open to the participants.

2.2 Audio Feedback

The auditory interface was created using a set of sound samples, triggered according to the actions of the user. We based the choice of audio samples on a musical metaphor with a string instrument. More specifically, the haptic movement through the planes was complemented by a plucked sound, and movements on a given plane were sonified by a smooth bowing sound. Cello samples (retrieved from www.freesound.org/) were chosen over other string instrument samples as the lower register was considered less obtrusive to the listener. Each plane was assigned a plucked pitch corresponding to the first 5 notes of a major scale mapped to the ascending plane number (A3, B3, C#3, D3, E3). Five plucked sounds (corresponding to the five notes) were used to inform the user of a plane crossing. A continuous bowed cello sound was played while the user was on the correct plane (i.e., the plane with the target) until they located the target. The pitch of the bowed sound was the same as the plucked sound for that plane level. Finally, when the user successfully located the target, a distinctive bowed chord auditory cue was played[4].

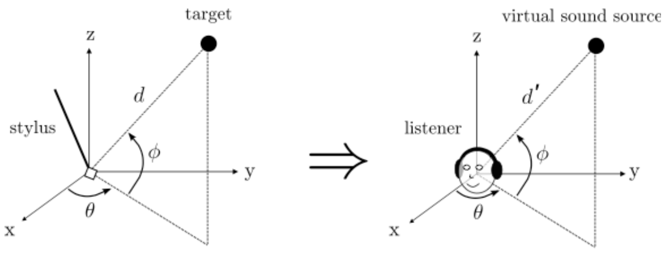

Fig. 2. The "ears in hand" metaphor: a virtual sound source is spatialized to simulate the target position relative to the haptic device stylus. The direction and distance (with a scaling factor) are reproduced with binaural synthesis.

---

[4] A video demo (with monophonic sound rendering) is available at `http://www.computing.dcu.ie/~emurphy/video/Audio-Haptic.mov`

The bowed sound was spatialized using binaural synthesis to guide the user on the plane and find the target. We used the “ears in hand” metaphor similar to that reported in [4]. The target was simulated as a virtual sound source, while the stylus (corresponding to listener’s hand) represented the center of listener’s head. More specifically, the azimuth θ, elevation φ and the distance d of the target were calculated in a coordinate system (x,y,z) whose origin was attached to the point of the haptic device’s stylus (see Figures 1 and 2). Then the bowed cello sound was spatialized with coordinates (θ, φ, d’) in a listener-centric system (Figure 2). The relatively small size of the Phantom Omni workspace (160 W x 120 H x 70 D mm) was compensated by a scaling factor of 60 for simulating the virtual source distance relative to the listener, i.e., d’=60d. The target-stylus distance being comprised between approximately 0 and 0.2 m, the corresponding source-listener distance varied between 0 and 12 m. Note that the orientation of the axes (x,y,z) always matched the orientation of the five planes in the virtual environment, and did not depend on the orientation of the stylus itself.

The Spatialisateur (v3.4.1.1) developed by IRCAM for Max/MSP was used for binaural sound rendering over headphones [12]. Directional cues θ and φ were simulated with Head Related Transfer Function (HRTF) filtering based on KEMAR measurements. The distance cue d’ was simulated by attenuating the direct sound with a gain: g = 1/d’ if d’ >1 m otherwise g=1. The cello sound source level was calibrated to approximately 50 dB-SPL for the reference distance of 1m. For a source-listener distance d’ inferior to 1 m (i.e., a stylus-target distance d inferior to 1.6 cm) the virtual source was moved on the upper hemisphere, reaching φ=+90 degrees when d’=0 m. Artificial reverberation was used to improve the distance perception and enhance the externalisation of the virtual source [13]. We choose a 2 second reverberation time and a global reverberation gain of -24dB. Since the source-listener distance varied from 0 to 12 m in the experiment, this settings lead to direct/reverberant signal ratios ranging from 24 to 10 dB. Doppler and air absorption effects were not simulated in this study. As the sound played only when the target and the stylus were located on the same plane, in the horizontal configuration the elevation of the virtual source was always 0 degree and the azimuth varied continuously between -180 to +180 degrees; in the vertical configuration, the azimuth of the virtual source was either -90 or +90 degrees, and the elevation varied continuously between -90 and +90 degrees.

2.3 Experimental Design

The experimental design was based on the following independent variables;

- Feedback: audio only, audio-haptic
- Orientation of the virtual planes: vertical, horizontal

This resulted in four experimental conditions: Audio-Haptic Vertical, Audio-Haptic Horizontal, Audio-Only Vertical, Audio-Only Horizontal.

2.3.1 Participants and Procedure

Twenty-three participants volunteered for the study, 20 males and 3 females between the ages of 24 and 55 (mean age: 32, SD: 8.2). Participants were either involved in an eNTERFACE ‘08 workshop or part of the wider research community at LIMSI, CNRS, France, involved in audio, computer science or engineering related research. While most users had experience testing and working with non-speech sound only 6 users reported prior experience testing or working with a haptic device.

Participants were provided with both a short training introduction and a practical trial session before beginning the main experiment. The initial training introduction involved presenting a visual description of the virtual structure (as represented in Figure 1). While the entire experiment was non-visual it was considered useful to show the users a visual representation of the virtual structure so that they could visualise the task mentally. However, users were not given any information about the 3D audio mappings or the haptic feedback.

In a practical training session, users were presented with 8 trials, 2 of each condition. Users were asked to navigate the planes, find the plane with the target and locate that target. For the main experiment participants were presented with 44 trials (11 per condition), except for the first 3 participants who completed 40 trials (10 per condition). The order of presentation of both orientation and feedback were radomised across trials. The target position within and across planes was also randomised. In order to move onto the next trial participants were asked to press a button on the Phantom Omni stylus when they were ready to move on. Timings were recorded from this point until the user located the target. There was an automatic detection of the end of task target location. User trajectories across and within the virtual planes were also recorded. Participants were also asked to complete a post-task questionnaire concerning perceived effectiveness and ease of use of the audio-haptic cues and the participant's cognitive strategies for finding the target.

3 Results

Users successfully located the target across all trials for the audio-haptic condition. In the audio-only condition, 4 users failed to locate the target for a combined total of 10 trials (6 horizontal, 4 vertical). Their timings were recorded for the unsuccessful trials (Average: 167s) but not included in the data for completion times. Furthermore, 62 trials out of 1000 trials were excluded from the data analysis due to technical difficulties.

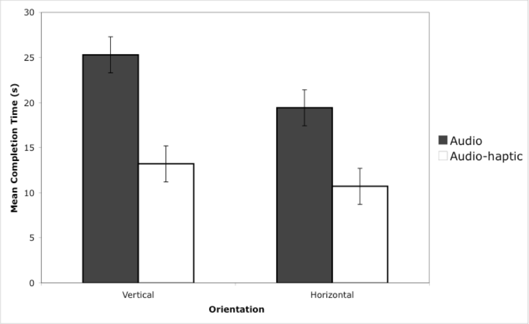

Fig. 3. Mean completion times for vertical and horizontal orientations across both audio only and audio-haptic conditions. Error bars display Standard Error.

Figure 3 illustrates the mean completion times for each condition. A repeated measures factorial ANOVA for completion times revealed significant effects of orientation and feedback. The time taken for users to complete the task in the horizontal condition was significantly less than that of vertical condition (F(1, 936) = 6.1, p=0.02). In addition, completion times for the audio-only condition were significantly longer than those of the audio-haptic condition (F(1, 936) = 49.3, p<0.001). Furthermore, post-hoc tests revealed a significant difference between the vertical and horizontal orientations for both audio-only (t(243) = -2.94, p two-tailed =0.004) and audio-haptic conditions (t(228) = -3.07, p two-tailed =0.002). There were no interaction effects between feedback and orientation.

3.1 Trajectory Analysis

We recorded the angular change of the arm of the Phantom Omni for each trial and each participant to gather information on the user's navigation path in addition to their completion times. A repeated measures factorial ANOVA for angular change revealed a significant main effect of orientation (F(1,936) = 5.6, p<0.05) and main effect of feedback (F(1,936)=18.0, p<0.0001)). There were no interaction effects between feedback and orientation. The mean angular data confirms the time data analysis in that the user trajectory paths were longer in the vertical orientation for both feedback conditions.

We also recorded the number of times that the users crossed planes when searching for the plane with the target for each trial. A repeated measures factorial ANOVA for mean plane crossings revealed a significant main effect of feedback (F(1,936) = 39.0, p<0.0001) but no main effect of orientation. No interaction effects were observed between feedback and orientation conditions. It is interesting that orientation did not have a significant effect on the mean plane crossings before users identified the target. The mean number of plane crossings are almost the same for both vertical (5.6) and horizontal orientations (5.7) in the audio-haptic feedback condition. Considering that there were 5 planes in the design, these ratings illustrate that the magnetic haptic effect was effective to control the participant's movements in both conditions. The greater mean number of crossings in the audio only conditions (14.7) demonstrated that this it was more difficult to remain on a given plane without haptic feedback, especially in the vertical condition.

3.2 Post-task Feedback, Observations, User Comments

In post-task questionnaires users were asked to rate the audio, haptic and combination of audio haptic cues on a scale of 1-5 (5 as the highest value) in terms of effectiveness and ease of use. A one-way ANOVA revealed a significant effect for ease of use ratings on the three types of feedback (F(2, 67= 8.7, p<.001). Post-hoc tests revealed a significant difference between ratings for effectiveness for haptic and audio-haptic cue conditions (t(44) = 2,3, p two-tailed =0.03), with audio-haptic cues perceived as significantly more effective than haptic cues. A one-way ANOVA revealed no significant effect for ease of use ratings across the three types of feedback.

It should be noted that participants never experienced the haptic feedback in isolation. As a result, some users interpreted rating the haptic cue as evaluating the haptic feedback in terms of whether the task would have been possible with haptic feedback alone. Overall the subjective feedback for effectiveness and ease of use of cues were relatively high (Mean Rating: 3.8, SD: 0.4) and in verbal feedback as part of the post-task questionnaires, users reported a pleasurable experience interacting with the virtual environment and found the task enjoyable.

As part of post-task questionnaires users were asked to describe their strategies for finding the target using the multimodal interface. Most users described a process of first navigating the virtual space to determine the orientation of the planes, then using the auditory cues to first determine the correct plane and concentrate on the location cue to find the target. From observation the most efficient users were those who immediately grasped the structure of the virtual environment and understood the 3D audio cues. In the post-task questionnaires 13 out of 23 users included descriptions of the spatial audio cues as part of their strategy for locating the target. From questionnaire analysis the remainder of the users did not refer to the fact that the cues were spatialized and instead commented on the changes in intensity for the location cues. In terms of the 3D audio rendering, no front-back or up-down reversals were mentioned by participants in the post-test questionnaire, nor did we observe any inversions in the analysis of the trajectories. Such reversals are typical artifacts of static binaural synthesis that can be reduced by head-tracking [13]. Head movements were not tracked in our experiment, but the "ears in hand" metaphor allowed subjects to explore dynamically the auditory space by means of the haptic stylus. It is possible that this exploration helped reducing the front-back and up-down confusions by providing dynamic localization cues to the listener.

4 Discussion

Our findings provide support for our research hypothesis, that haptic feedback improves performance of a target-finding task. The highly significant difference in completion times between audio vs. audio haptic feedback can be accounted for by the design of the interface, as the haptic cues were designed to support the auditory interface without any redundant information. Although users found it more difficult to remain on the virtual planes, it should be noted that they were able to complete the task using the auditory cues within a reasonable length of time without haptic feedback. But users may have been applying the mental model of the virtual structure previously constructed during the training session or previous trials using audio-haptic condition. This would need to be further investigated in a separate audio-only evaluation study.

The significant difference of completion times in the horizontal for both audio and audio-haptic conditions is a reflection of the mappings for the binaural rendering of 3D sound as this was the main cue to allow users locate the target. An explanation for the longer completion times in the vertical orientation is that users were relying on elevation cues (with non-individualized HRTF) to navigate the surface of the planes in this condition. This finding is in agreement with previous studies that have illustrated that participants find elevation cues difficult to perceive [14]. From this finding it is possible to recommend that in the design of auditory interfaces using 3D sound to convey a virtual 3D space, it is better to map 3D binaural rendering along a horizontal plane. Vertical movements may better be conveyed through other non-speech auditory parameters such as pitch or intensity. Some users alternated their hand position to compensate for the changing orientation of the planes. It should be acknowledged that there are physiological constraints in terms of the gestural control of the haptic device and this could have an effect on user performance in the vertical orientation.

5 Conclusions and Future Work

We foresee that the design idea of virtual planes developed during this experiment could be favorably adapted to ungrounded interfaces like portable devices where small screen sizes are often a limiting factor in building complex interfaces. The haptic cues are generated from the device itself as it moves and changes orientation. The use of an expanded navigation space using the proposed 3D plane structure with audio and haptic feedback provides interesting capabilities for developing meaningful interfaces that are contextually relevant and engaging. Simple but very valuable haptic feedback could support browsing of various layers of data (maps, menu structures, subsets of a larger collection, etc) or enhance selection task from a few discrete items with haptic detents through 3D space, like a scroll wheel on a computer mouse. Haptic layers could be enhanced by using spatial audio to render options not accessible in a limited visual display. For example the 3D spatial audio design presented in this paper could be extended to include audio pan-and-zoom techniques proposed in [15]. The target finding interface could be exploited to test a combination of audio, haptic and visual cues to implement a collaborative game for blind and sighted users. Interaction design using audio and haptic feedback has the potential to create learning objects to include children otherwise marginalised as a result of disabilities such as visual impairment or learning disabilities.

Acknowledgments. This study was supported by an NSERC-SRO team grant on ENACTIVE Interfaces (P.I.: M. Wanderley), We would like to thank Orange Labs, Dr. Antoine Gonot and Dr. Ilja Frissen.

References

1. Oakley, I., McGee, M.R., Brewster, S., Gray, P.: Putting the feel in look and feel. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2000), pp. 415–422 (2000)
2. Wall, S.A., Paynter, K., Shillito, A.M., Wright, M., Scali, S.: The effect of haptic feedback and stereo graphics in a 3d target acquisition. In: Proceedings of Eurohaptics, Edinburgh, UK, July 8-10 (2002)

3. Kim, S.-C., Kwon, D.-S.: Haptic and Sound Grid for Enhanced Positioning in a 3-D Virtual Environment. In: Oakley, I., Brewster, S. (eds.) HAID 2007. LNCS, vol. 4813, pp. 98–109. Springer, Heidelberg (2007)
4. Magnusson, C., Danielsson, H., Rassmus-Gröhn, K.: Non Visual Haptic Audio Tools for Virtual Environments. In: McGookin, D., Brewster, S. (eds.) HAID 2006. LNCS, vol. 4129, pp. 111–120. Springer, Heidelberg (2006)
5. Moll, J., Huang, Y., Sallnas, E.: Audio makes a difference in haptic collaborative virtual environments. Interact. Comput. 22(6), 544–555 (2010)
6. Picinali, L., Menelas, B., Katz, B., Bourdot, P.: Evaluation of a Haptic/Audio System for 3-D Targeting Tasks. In: Proceedings of the 128th Convention of the Audio Engineering Society, London, UK (2010)
7. Ménélas, B., Picinali, L., Katz, B., Bourdot, P.: Audio haptic feedbacks in a task of targets acquisition. In: IEEE Symposium on 3D User Interfaces (3DUI 2010), Waltham, USA, pp. 51–54 (2010)
8. Blauert, J.: Spatial hearing: the psychophysics of human sound localization. MIT Press, Cambridge (1997)
9. Gonot, A., Chateau, N., Emerit, M.: Usability of 3-D Sound for Navigation in a Constrained Virtual Environment. In: Proceedings of the Audio Engineering Society (AES) Convention (2006)
10. Wall, S.A., Paynter, K., Shillito, A.M., Wright, S., Scali, M.: The effect of haptic feedback and stereo graphics in a 3d target acquisition. In: Proceedings of Eurohaptics, Edinburgh, UK, pp. 23–29 (2002)
11. Hwang, F., Keates, S., Langdon, P., Clarkson, J.: Multiple haptic targets for motion-impaired computer users. In: Proceedings of CHI 2003, Florida, USA, pp. 41–48 (2003)
12. Jot, J.-M.: Real-time Spatial processing of sounds for music, multimedia and interactive human-computer interfaces. ACM Multimedia Systems Journal 7(1), 55–69 (1999)
13. Begault, D.R., Wenzel, E.M., Anderson, M.R.: Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. Journal of the Audio Engineering Society 49(10) (2001)
14. Lokki, T., Grohn, M.: Navigation with auditory cues in a virtual environment. IEEE Multimedia 12(2) (2005)
15. Bouchara, T., Katz, B., Jacquemin, C., Guastavino, C.: Audio-visual renderings for multimedia navigation. In: Proceedings of the 16th International Conference on Auditory Display, Washington, DC (2010)