

Audio Haptic Feedbacks for an Acquisition Task in a Multi-Target context

Bob Ménélas*

Lorenzo Picinalli†

Brian F. G. Katz‡

Patrick Bourdot§

CNRS-LIMSI, BP. 133 91403 Orsay, France

ABSTRACT

This paper presents the use of audio and haptic feedbacks to reduce the load of the visual channel in interaction tasks within virtual environments. An examination is made regarding the exploitation of audio and/or haptic cues for the acquisition of a desired target in an environment containing multiple and obscured distractors. This study compares different ways of identifying and locating a specified target among others by the mean of either audio, haptic, or both feedbacks rendered simultaneously. The analysis of results and subjective user comments indicate that active haptic and combined audio/haptic conditions offer better results when compared to the audio only condition. Moreover, that the association of haptic and audio feedback presents a real potential for the completion of the task.

Index Terms: H.5.2 [Information interfaces and presentation]: User interfaces—Haptic I/O; H.5.2 [Information interfaces and presentation]: User interfaces—Auditory (non-speech) feedback;

1 INTRODUCTION

We consider here the case of virtual environments within which visual objects can often be hidden by others. While various studies have been carried out to address this situation using visual feedback, the use of haptic or audio have rarely been explored for such a task. One can refer to [5] for a survey of such techniques.

However, to fully take advantage of the sensory capabilities of the human system, a benefit would certainly come from exploiting other sensory channels such as the haptic and audio ones. Indeed, during the last decade many studies investigated the benefit of adding haptic and auditory cues to purely visual systems in target selection tasks. In addition to speed improvements for the completion of the task, Dennerlein et al. [4] underlined that the use of an attractive force-feedback may reduce the muscle skeletal load during computer mouse use. In the same way, in [7] a haptic and audio grid has been used in order to enhance recognition of ambiguous visual depth cues for position selection. In [2], Cockburn et al. outlined how multimodal interaction could provide substantial improvements in a single target task, when compared with visual only feedback. On the other hand, Wall et al. [11] highlighted that in the presence of multiple distractors, adding force feedback (a virtual magnet) to a 3D stereoscopic virtual rendering improved subjects' accuracy, but did not improve the time taken to reach the target. Moreover, Hwang et al. [6] noted that positioning the distractors along the route of the target was detrimental to performances of motion-impaired users. Although some work has investigated navigation in virtual environment by the mean of audio/haptic systems (non-visual), to our knowledge no work has addressed the exploitation of non-visual feedbacks in selection of occluded targets. This

paper tackles this particular issue; it represents the first step in a work which aims at alleviating the visual channel by exploiting an audio/haptic method in the selection of occluded targets. While the visual channel could present some relevant information related to the total virtual scene, alternate feedbacks can be taken advantage of in order to provide additional information and interactions. For instance in [8] the vibration capabilities of a 6 DoF device are exploited in order to provide information about the local area explored. Singular points located in the immediate environment of the users are rendered visually by colored spheres and haptically through a sinusoidal vibration.

In the following experiment we present a protocol hypothesis that, in the working context of the user, the visual channel is either too saturated or is presenting alternate information, such that the selection task cannot benefit from the visual rendering. There are numerous cases where this situation can occur, for example in the case of multidimensional data exploration where selection is based on a multitude of parameters. Following this hypothesis, this study focuses on the usage of purely audio and haptic feedbacks for acquisition of a given target among several others. It generalizes the work presented in [9] to a multi-target context.

2 EXPLOITATION OF HAPTIC AND/OR AUDITORY FEEDBACKS IN ACQUISITION OF A GIVEN TARGET AMONG $(n - 1)$ OTHERS

Our work aims at exploiting pure audio and/or haptic rendering for the acquisition of a given target amongst several others. For this purpose, we will define a method for each channel (haptic, audio) that will allow users to approach and to identify a target from $(n-1)$ others with no other rendered information. After a certain number of pre-trials showed that it was difficult for subjects to memorize a spatial scene with more than 4 haptic / audio signatures in a given timeframe, we have limited this study to an environment with 4 targets.

2.1 Haptic feedback rendering

For the completion of this task, the metaphor of an attraction space is exploited. For each target, T_i , we define a zone Z_i in the interior of which the target T_i attracts the user as a virtual magnet. We have therefore partitioned the workspace into a set of disjointed sub-spaces.

One manner for subdividing the space into separate zones would be to use a 3D Vorono partitioning algorithm [3]. However, using this approach every point of the workspace would be part of a target zone sub-area. As such, for any position in the workspace, the user would be in the attraction area of one target. To avoid this, it would be necessary to generate 'empty' targets in order to create smaller and more equally proportioned sub-spaces.

Due to these issues, the use of a simple attraction sphere is employed. The attraction sphere around a target T_i is centered on T_i and has a radius R such that R is the shortest distance between the T_i and the $(n - 1)$ other targets, ensuring no overlap between attraction areas. This provides some attraction free regions that can be used as references.

Concerning the attraction force, we exploited a model based on the haptic grid proposed in [12]. The magnitude of the feedback is adapted in the following way:

*e-mail: bob.menelas@limsi.fr

†e-mail: lorenzo@limsi.fr

‡e-mail: brian.katz@limsi.fr

§e-mail: patrick.bourdot@limsi.fr

With the radius R of the attraction sphere around target T , let $r < R$, let x be the distance to T , and F_{max} the maximum attraction force, with the magnitude of the attraction force F as defined in equation 1.

$$\left(\begin{array}{l} \|F\| = \|F_{max}\| * \sin(\pi * x / 2r) \quad x \in [0; r] \\ \|F\| = \|F_{max}\| * \sqrt{1 - (x - r)^2 / (R - r)} \quad x \in [r; R] \end{array} \right) \quad (1)$$

Starting from $x = R$, the user is attracted with a force proportional to x towards the target center until $x = r$, from where the force decreases sinusoidally. With such a model, the attractive force increases very quickly, so that entering into an attraction area can be quickly and easily perceived by the user. In contrast, as one closely approaches the target the attractive force decays smoothly, allowing the user to move freely and easily around the target region. In the present study, after the pre-trial tests, we defined $r = R/4$.

Furthermore, to minimize the impact of the distractors, as proposed in [10], the haptic feedback is adjusted according to the assumed intentions of the user. If the attraction force is towards an unwanted target (a distractor), it can be expected that the user will tend to oppose the attraction. To facilitate this change of intention, minimizing the impact of the distractors attraction force, the value of the attractive force is divided by a constant factor (4) when x is increasing, when the user is moving away from the target.

Finally, to let subjects identify haptically each target of the scene, we defined four haptic signatures (equation 2) using waveform amplitude modulation [1]. These haptic patterns, designed within the performance constraints of our device, were provided through rotational vibrations of the wrist of the 6 DoF haptic device around its transversal axis and are independent of the attraction force.

$$\left(\begin{array}{l} W_1 = a * \sin(2 * \pi * 121 * dt) \\ W_2 = a * \sin(2 * \pi * 0.5 * dt) * \sin(2 * \pi * 121 * dt) \\ W_3 = a * \sin(2 * \pi * 3 * dt) * \sin(2 * \pi * 121 * dt) \\ W_4 = a * \sin(2 * \pi * 31 * dt) * \sin(2 * \pi * 53 * dt) \end{array} \right) \quad (2)$$

W_1 defines a sinusoid, describing a continuous vibration of 121 Hz. W_2 is an amplitude modulation of W_1 by a 0.5 Hz sinusoid producing the sensation of a pulsing vibration. W_3 is a modulation of W_1 by a 3 Hz sinusoid, producing the sensation of rapid impulse vibration. W_4 is a 53 Hz sinusoid modulated by a 31 Hz whose combination resulted in a rather rough vibration sensation.

2.2 Audio feedback design

The audio rendering uses parameter mapping sonification combined with 3D audio spatialisation. The chosen syntax is an impact sound, spatially rendered at the target positions, whose repetition rate and level varies as a function of x . The metaphor proposes two different sonic cues, positional information via spatialisation, and distance via repetition rate and level variations.

Similar to the haptic experimental design, each target was attributed a unique audio signature. We chose four impact sounds from the freesound project¹: W_1 (a small bell, $f_0 = 2110$ Hz), W_2 (wood block, $f_0 = 840$ Hz), W_3 (table tap/click, $f_0 = 560$ Hz), and W_4 (window knock, $f_0 = 140$ Hz).

The repetition rate of the target sound was inversely proportional to the distance to each target, and were interpolated using a linear scaling from 1 Hz (workspace diagonal length, maximum possible distance) to 6 Hz (at the actual target position). The level of the signature sounds also varied inversely proportional with distance by 20 dB over the same distance range. This configuration allowed for the farthest target to remain still slightly audible and easily identifiable, while not being too loud when at the target position.

Each target sound signature was positioned at the target location using binaural spatialisation based on convolution of the signal and

the corresponding HRIR (Head Related Impulse Response,) of the position to be simulated. The spatialisation engine employed for the test allows for the individualization of Inter-aural Time Difference (ITD) applied to any selected HRTF (Head Related Transfer Function). While no room reverberation was used, distance attenuation was enhanced, and real-time correction of Inter-aural Level Difference (ILD) was implemented to aid in the binaural perception of target distance for both near and far targets.

As humans are capable of listening to several simultaneous audio streams, it was not necessary to apply the notion of activation area as was needed for the haptic feedback. The audio feedback for all targets was always present, with appropriate level dependence.

3 EXPERIMENTATION

This experiment aims at the evaluation of the proposed haptic and/or audio rendering for the selection of occluded targets. The effectiveness of haptic, auditory, and multimodal (haptic and audio) feedbacks in the completion of the task was computed and analyzed separately and together.

Three experimental conditions (A, H, M) are defined: in condition A (audio) only the auditory feedback (described in section 2.2) is available, in condition H (Haptic) only the haptic rendering (described in section 2.1) is provided, while in condition M (Multimodal) simultaneous haptic and auditory cues are provided.

Due to the limited workspace of the haptic device ($100 \times 90 \times 60$ cm) and to difficulties in rendering audio sources when these are very close to the head, the geometry of the haptic scene ($50 \times 50 \times 50$ cm) was scaled by a constant factor for the audio rendering so that the experimental cube maintained the same angular information (radius scaling of head-centered coordinate system). A pre-selection screening phase was used to help the users chose an optimal HRTF from an existing database. Selection, combined with individual ITD adaptation, improves the quality of spatial audio rendering for cases when non-individual HRTFs are used.

A total of 18 persons (14 male and 4 female), aged between 23 and 55, were asked to participate. The first six participants were evaluated under H and A conditions (3 using the order H then A and 3 others A then H). Six other participants were presented with the A and M conditions (3 with A then M, 3 with M then A). The final participants tested H and M conditions (3 with H then M and 3 with M then H).

The experimental setup consisted of the main application, which dealt with the data, and two other components related to the haptic and audio rendering. The haptic feedback was supplied via a Virtuose 6 DoF device by HAPTION. The audio rendering was provided through a wireless stereo headset (Sennheiser RS65). The auditory rendering was implemented in the MaxMSP environment. In all the condition setups (H, A, and M), the displacement of the probe in the virtual environment (VE) was performed via the haptic device.

The orientation of the head of the participant was tracked using an ARTrack infrared system, and sent to the audio rendering application in order to modify the spatial auditory rendering with respect to the head movements, maintaining the simulated 3D sound sources at their proper stable position in space, irrespective of head movements.

In the initial starting condition, the haptic device was physically placed at the center of its workspace, and the probe position was located at the origin of the VE.

3.1 Experimental plan

For each trial, one of the six arrangements shown in figure 1 was rendered. These six test configurations were chosen according to arrangements used in [6]. Four of these six configurations place the three distractors along or near the direct path leading to the target from the user's initial starting position.

¹ <http://www.freesound.org>

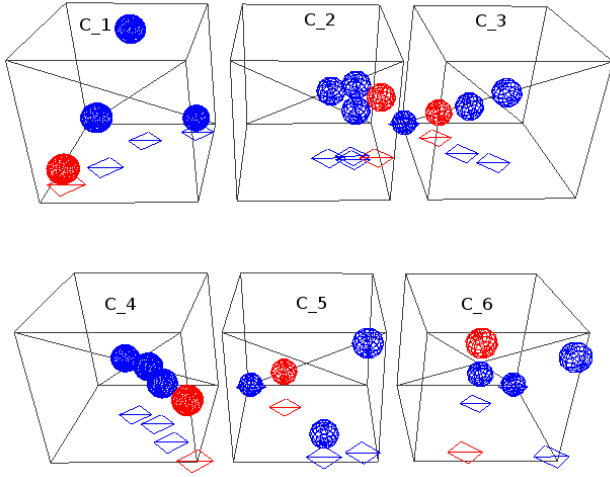


Figure 1: Test configurations: target (red) and distractor (blue) positions (spheres). 2D projections (diamonds) are shown for spatial reference.

For each session, each configuration was repeated 4 times. For each repetition, a different signature order assignment was used so that the desired target was never the same for the same configuration.

3.2 Procedure

A 3D acquisition task was used for the experiment. At the beginning, participants received a brief written and oral explanation about the goal of the experiment, after which followed a familiarization phase and the actual test. The test phase counts two stages.

Stage 1: In the first stage participants were presented with directional information (according to the current modality configuration) for each target sequentially in order to help construct a mental spatial map of the scene. Using this experimental design, both audio and haptic modalities present the same information and in roughly the same time frame. This specifically designed protocol allows for a haptic only condition to be feasible for the chosen task, enabling for the suitable comparison of results for each channel.

When the haptic feedback was active, the subject would be first attracted in the direction of the location of the target (for a duration of 1.5 s). Then, in addition to the attraction force, the haptic signature was rendered (duration 2 s). At the end of this rendering, the haptic device was pulled back towards the center of the virtual space in the event of any displacement of the user's hand. Once returned to the center, a pause of 2 s, then the process proceeded to the next target until all four targets were presented.

If the auditory modality was active, the auditory signature of the activated target was spatially rendered in the appropriate direction for 3 s at a fixed distance in space. After a pause of 2 s, the next target rendered.

This stage was repeated until subjects indicated that they understood the spatial configuration (directional only) of the different sources. The subject did not know at this stage which one of the sources would be the designated target.

Stage 2: In the second stage, the subject was first presented with the signature of the target to find (duration of 2 s) and was then instructed to exploit the available feedback in order to locate the specified target position as fast and as precisely as possible. The subject moved the haptic device (which also served as the position sensor) to find and select the target position. Timing for each trial starts when the participant began to move the haptic device and

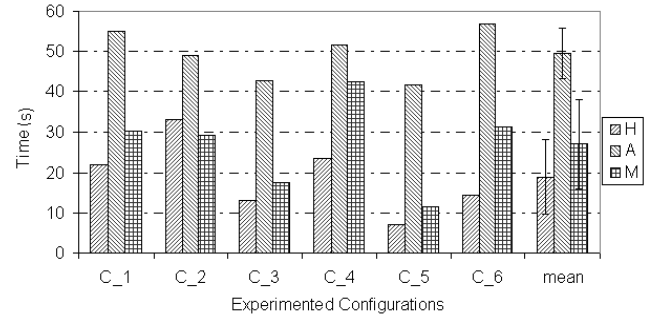


Figure 2: Average task completion time for each experimental condition and overall configuration mean and stdev results

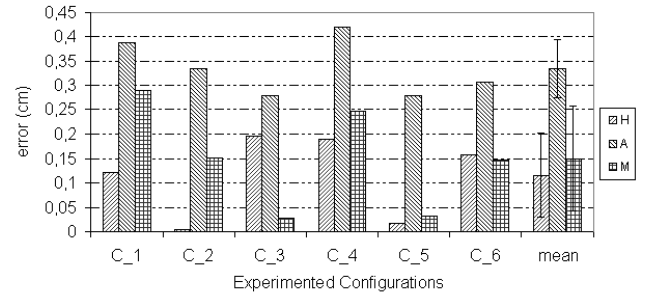


Figure 3: Average selection distance error for each experimental condition and overall configuration mean and stdev results

terminated when the right button of the haptic device was pressed, indicating that the current position had been identified (selected). The total experiment lasted 75 minutes on average.

3.3 Results and discussions

Comparisons are made between the effectiveness of the three tested conditions (H, A and M). The average time required by users and the average distance error between the selected point and the desired target is calculated (see figure 2 and figure 3 respectively).

Measurements made on the task completion time and distance error in the selection show that haptic and multimodal (haptic + audio) conditions offered much better performance than the audio alone. In the haptic condition, we observed an average of 18.89 s for the completion time and 0.115 cm error. For the multimodal condition, the average time is 26.96 s and the error 0.14 cm. In contrast, for the purely audio condition, a longer average completion time, 49.49 s, and distance error, 0.33 cm, are noted.

A one-way ANOVA, followed by a post hoc Tukey was performed on the mean results for each experimental configuration. Analysis of computed measures shows that H and M conditions are better suited to the task completion than the A condition where only audio feedback is available. Considering the average completion time and the distance error, a significant difference is observed between the three experimental condition ($F_{2,15} = 18.564, p = 0.008$) and ($F_{2,15} = 11.182, p = 0.001$): respectively. A post hoc Tukey highlights that both Haptic and Multimodal conditions are significantly better suited to the task than the Audio condition (H-A and M-A $p < 0.001$). However, no significant difference is noted between H and M conditions. For required time and the error in selection we note that $p = 0.29$ and $p = 0.77$ respectively.

With this experiment we observed many differences between the proposed auditory and haptic feedbacks. In post experimental dis-

cussions, many subjects noted a slight difficulty in the identification of haptic patterns, in contrast to the audio condition. More attention was required in order to identify haptic patterns. Another difference concerned the relevance of the first stage (sequentially presentation of direction). Subjects indicated that in haptic condition this stage was essential, while this was not the case in the audio condition. This was understandable as it was obvious that the completion of such a task would be difficult with the haptic only feedback. Furthermore, it was interesting to note that, in the haptic condition, while it was difficult to memorize the configuration, non-designated targets (distractors) could be used as points of reference and therefore may help in achieving the task.

In contrast, for the audio condition, since the second stage audio rendering also provided complete spatial information (all sources were always somewhat audible), subjects noted that the first stage seemed redundant and unnecessary. We note again that the two stage protocol was designed to allow for the haptic only condition to be feasible for the given task, and that at the beginning of the second stage the subject had the same directional mental map information, regardless of modality condition.

The main difficulty noted for the audio condition was the precise selection of the target. Compared to the H condition, in the A condition it was difficult for subjects to clearly determine when the audio feedback was at its highest level both in terms of repetition rate and level. Thus, it was difficult to precisely locate the designated target position, possibly due to the ranges and linear mapping used. For that, subjects exploited small displacements around the target while assessing changes in the audio feedback. Doing this, subjects employed a step-by-step approach to get closer to the target. This strategy explains the considerable time and possibly the observed errors in the selection task. However, in the H condition, subjects reported the simplicity of *letting themselves be guided* by the haptic device toward the designated target. They highly appreciated the assistance of this metaphor in the selection process and the benefit of an active feedback guide.

In addition to task time and selected position, the entire trajectory was stored. An inspection of the trajectories confirms the observations outlined above. Figure 4 represents the trajectories described by four subjects in C.2 configuration. One observes that in the first haptic condition (a) the subjects directly go towards the desired target while in (b) it appears that the user had difficulty identifying/recalling and locating the desired target. However in audio conditions one may appreciate two different strategies for approaching the target. In (c) the user directly goes toward the area where the target is located, while in (d) simple linear displacements are adopted. Moreover, we observe that in both audio conditions the precise location of the desired target was difficult to pinpoint. We also note that in general auditory localization in elevation is more difficult than in azimuth.

These observed differences explain why the multimodal condition appears as the most appreciated by subjects. Some participants remarked that the M condition contained benefits from the two modalities. Nevertheless, the presence of the haptic signature was moderately popular with users. For some participants, the haptic signature may be considered as a sort of confirmation of the audio signature of the target, while for others it appears as a surplus of information. This point can be addressed in a future work.

4 CONCLUSION AND FUTURE WORK

In a 3D virtual environment containing multiple and obscured targets, audio only, haptic only, and multimodal feedbacks have been investigated relative to the selection of a specified target. Through an analysis of results, comments of participants, and the various differences observed between Audio and Haptic conditions, we understood that these two channels (considered separately) do not seem to be completely adapted to the quick and precise selection of oc-

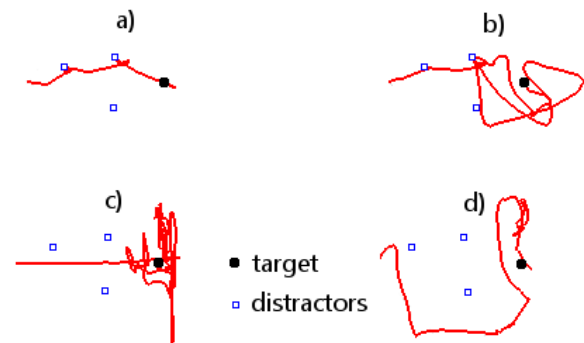


Figure 4: Vertical plane trajectory representation of four different users in the C.2 configuration. Cases a) and b) are realized in the Haptic condition while cases c) and d) are in the Audio condition.

cluded targets. Each of these two channels offers both advantages and drawbacks. While the auditory feedback presents a great potential for the spatial perception, the active attraction component offered by haptics may provide useful assistances for precise selections, even with multiple targets. Moreover, the tested combination of these channels shows a real potential for tackling the completion of the studied task. However, we believe that addition of both rendering (audio and haptic) while offering useful results may be improved. To confirm this hypothesis, we are now investigating optimized combinations of these channels.

REFERENCES

- [1] L. Brown, S. Brewster, and H. Purchase. A first investigation into the effectiveness of tactons. In *Proc. of World Haptics 2005*, pages 167–176, 2005.
- [2] A. Cockburn and S. Brewster. Multimodal feedback for the acquisition of small targets. In *Ergonomics*, Vol 48, No. 9:1129–1150., 2005.
- [3] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf. *Computational Geometry (2nd revised edition ed.) algorithm. Chapter 7: Voronoi Diagrams*. 2000.
- [4] J. Dennerlein and M. C. Yang. Haptic force feedback devices for the office computer: performance and musculoskeletal loading issues. *Human Factors*, 43, 2:278–286, 2001.
- [5] J. Flasar, , and J. Sochor. Manipulating objects behind obstacles. In: *Lecture Notes in Computer Science*, 4563:32–41, 2007.
- [6] F. Hwang, S. Keates, P. Langdon, and P. J. Clarkson. Multiple haptic targets for motion-impaired computer users, 2003.
- [7] S.-C. Kim and D.-S. Kwon. Haptic and sound grid for enhanced positioning in a 3-d virtual environment. *Haptic and Audio Interaction Design*, pages 98–109, 2007.
- [8] B. Menelas, M. Ammi, L. Pastur, and P. Bourdot. Haptical exploration of an unsteady flow. In *Proc. of World Haptics 2009. Third Joint*, pages 232–237, 2009.
- [9] B. Menelas, L. Picinali, B. F. G. Katz, P. Bourdot, and M. Ammi. Haptic audio guidance for target selection in a virtual environment. In *Proc. 4th International Haptic and Auditory Interaction Design Workshop (HAID'09), vol. II, Dresden*, Sept 2009.
- [10] I. Oakley, A. Adams, S. A. Brewster, and P. D. Gray. Guidelines for the design of haptic widgets. In *In Proc. BCS HCI'02, London, UK, Springer*, pages 195–212, 2002.
- [11] S. A. Wall, K. Paynter, A. M. Shillito, and S. Wright, M. & Scali. The effect of haptic feedback and stereo graphics in a 3d target acquisition. In *Proceedings of Eurohaptics, Edinburgh, UK*, page pp. 2329, 2002.
- [12] T. Yamada, D. Tsubouchi, T. Ogi, and M. Hirose. Desk-sized immersive workplace using force feedback grid interface. In *Proc. IEEE Virtual Reality*, pages 135–142, 2002.