

Non-visual identification, localization, and selection of entities of interest in a 3D environment

Bob-Antoine J. Menelas · Lorenzo Picinali ·
Patrick Bourdot · Brian F. G. Katz

Received: 26 October 2012 / Accepted: 9 January 2014 / Published online: 17 February 2014
© OpenInterface Association 2014

Abstract This paper addresses the use of audio and haptics as a mean 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 interactions for the acquisition of a target of interest in an environment containing multiple and obscured distractors. A first study compares means for identifying and locating a specified target among others employing either audio, haptic, or both sensori-motor channels activated simultaneously. Following an analysis of the results and subject comments, an improved multimodal approach is proposed and evaluated in a second study, combining advantages offered by each sensory channel. Results confirm the efficiency and effectiveness of the proposed multimodal approach.

Keywords Audio · Haptics · Non-visual interaction · Multimodal interaction · Multi-target · Haptic identification · Haptic selection · Sonification · Haptic audio interaction

B.-A. J. Menelas (✉)
Mathematics and Computer Science Department,
University of Quebec at Chicoutimi, 555, boulevard de l'Université,
Chicoutimi, Canada
e-mail: Bob-Antoine-Jerry_Menelas@uqac.ca

L. Picinali
Faculty of Technology, De Montfort University,
The Gateway, Leicester LE1 9BH, UK
e-mail: lpicinali@dmu.ac.uk

B. F. G. Katz · P. Bourdot
LIMSI-CNRS, BP 133, Orsay, France
e-mail: brian.katz@limsi.fr

P. Bourdot
e-mail: Patrick.Bourdot@limsi.fr

1 Introduction

In everyday life, we interact and perceive the surrounding environment through our various senses. While reading a book, without any noticeable difficulty and by taking advantage of haptic and auditory channels, we may grasp the ringing phone located outside of our field of vision without having to turn away from the book. However, common user interface designs tend to focus on visual feedback. Nevertheless, when it comes to 3D environments, in the presence of multiple objects, some of these might be hidden by others. While various studies have been carried out to address this situation by means of visual feedback, the use of haptic or audio has rarely been explored for such a task. Among the techniques based on the visual channel, two main approaches can be found: modifying the rendering mode and interaction-based methods. Considering a polygonal scene, transparency or wireframe rendering can be used to reduce visual masking or occlusions between objects within the scene. Other rendering techniques, such as space distortion, have also been exploited [8]. For example, Elmqvist and Tsigas have used the metaphor of a spherical force field that can be inflated or deflated like a balloon, having the vproperty of moving objects of the scene with which it collides during its expansion [12]. Regarding interaction methods, some techniques employ virtual navigation paradigms while others are based on interaction metaphors such as a virtual mirror [16]. The work of [14] can be referred to for a detailed survey of such techniques.

However, to take full advantage of the sensory capabilities of the human perceptual system, a benefit would certainly come from the exploitation of other sensory channels such as haptics and audio [28,29]. As illustrated previously, haptic and audio channels can indeed be relevant for the perception and acquisition of occluded targets in everyday life.

In this paper, we investigate whether haptic and/or auditory interactions can be used for identification, localization, and selection of a given occluded target among several others in a 3D environment. Using the proposed method, the visual channel is left free and can thus be exploited for other tasks as shown in our previous works regarding the analysis of large data sets [28,29,33].

After a brief overview of related works in Sects. 2, 3 describes the exploitation of haptic and/or auditory interactions in the acquisition of an occluded target. Proposed methods are evaluated in Sect. 4. A combined multimodal approach is proposed and evaluated in Sect. 5.

2 Related work

This section surveys works related to exploitation of haptics and audio in target acquisition. After a brief review concerning cases where a single target is presented in the environment, we discuss works which deal with multi-target tasks.

2.1 Single target condition

Regarding single target acquisition, various studies have highlighted the benefits of adding force and auditory feedback to a purely visual system. It was shown in [17] that haptic feedback can improve the speed of task completion. In addition to speed increasing, Dennerlein et al. [10] underlined that the use of an attractive force-feedback may reduce the muscle skeletal load during computer mouse use. In the same way, Oakley et al. [37] used haptic feedback in interactions with a conventional graphic user interface in order to reduce error rates while in [24] a haptic and audio grid has been introduced in order to enhance recognition of ambiguous visual depth cues for position selection. In [9], audio and haptic feedback have been used in conjunction with a graphical interface for a single target task: the research outlined how multimodal interaction could provide substantial improvements when compared with visual only feedback. Spatial auditory rendering has been used in virtual environments for spatial studies with visually impaired individuals [1] and similarly for the development of guidance and navigation systems for the blind [23]. In the same way, haptic rendering was combined with audio feedback in order to investigate object locations in a non-visual spatial environment for blind and visually impaired children in [27]. A target selection approach has been used in [43] in the scope of more general research on multimodal rendering, outlining that the addition of multimodal feedback was always preferred by the user, and that in many cases it could also speed up task completion. In earlier works related to the current study, audio and haptic renderings have been investigated in a single target 3D exploration task, outlining that while audio

feedback showed relevant benefits with the addition of haptic, the opposite was not true. In fact, there was always a significant difference, in favor of haptics, between audio-only and haptic-only condition [31,38]. These results were a driving force for the definition of the current study, which includes the additional complexity of multiple active potential targets.

2.2 Multiple target condition: haptic distractors

In contrast to previous works, few studies conducted in the context of environments presenting multiple targets have been carried out with the inclusion of haptic and auditory feedback.

Vanacken et al. [42] investigated a multimodal feedback system for target acquisition in densely populated environments. In the tested system, haptic as well as auditory feedback were employed in order to inform the user about the existence of a target in their immediate vicinity. No directional information was provided to the user by either haptic or audio channels, and results did not show any significant improvement of such a multimodal system when compared to a purely visual system.

Evaluations in [11] of a point-and-click task showed that performances were better with the addition of force feedback even with distracting haptic force fields. However it was noted that speed tended to decrease with the addition of distractors. On the other hand, Wall et al. [44] 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. [20] have investigated the impact of multiple haptic distractors on the performance of motion impaired users. Their studies showed that positioning the distractors along the route of the target was detrimental to performance.

In contrast, Oakley et al. [36] highlighted the benefits of an adjusted visual-haptic condition relative to visual-only and direct visual-haptic conditions. In the adjusted condition, in order to reduce the impact of distractors which are in fact non-desired attracters; their attractive effects are decreased whenever they do not seem to interest the user. Results indicated that target selection errors were reduced to the same level as in the haptic condition, while speed was not compromised when compared to the visual condition.

Regarding audio feedback, multi-source environments have often been employed for multi-talker speech intelligibility, stream segregation, and complicated sound source localization tasks (see [18] and [40]). Studies concerning audio-haptic coupling have examined the exploration of 1D and 2D scalar data fields using common cross-modal stimuli design [3]. To our knowledge, no work has addressed the usage of a purely audio/haptic system in a multi-target/distractor context in a 3D environment. Our work aims at filling this gap.

Following findings of previous studies, the attractive force feedback metaphor seems well-suited to target acquisition, with the necessary addition of an adjusted condition as proposed in [36,39]. Moreover, configurations which are deemed to be difficult [20], such as ones with occluded targets, should be included in any evaluation.

3 Exploitation of haptic and/or auditory feedback in the acquisition of an occluded target

Let us consider the case of an environment that contains a set of n targets which may be occluded. This section presents a system that takes advantage of haptic and auditory feedback for identification, localization, and selection of a specified target among the others. Preliminary analysis of the results of this particular study have been presented in [30] with additional protocol details and results presented here.

3.1 Proposed haptic feedback

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 to the interior of which the target T_i attracts the user as a virtual magnet. A unique haptic signature is used to identify each target. The attraction area, the attraction force and the haptic signature of each target are defined in the following sections.

3.1.1 Attraction area

Defining an attraction area for each target results in a partitioning of the workspace into a set of disjointed sub-spaces. As such, each target has its own singular non-overlapping area of influence. One manner to subdivide the space into separate zones would be to use a 3D Voronoi space partitioning algorithm [5]. However, with such an approach every point of the workspace will be part of a target zone sub-area. As such, for any position in the workspace, the user will always be in the attraction area of a target, regardless of the distance to the target. To avoid this, it would be necessary to create “empty” targets allowing smaller, empty and more equally proportioned sub-spaces with the Voronoi approach. In addition, as Voronoi sub-space shapes are not radially symmetric (by definition), the large variations between such defined attraction areas could difficult for the user to conceptualize and interpret such spatial divisions without visual cues.

To avoid these problems, the use of an attraction sphere has been preferred, defined as follows. The attraction sphere around a target T_i is centred on T_i and has a radius R such that R is the smallest distance between two distinct targets of the group, ensuring no overlap of attraction areas.

3.1.2 Attraction force: target localization

Yamada et al. [45] presented different force models which may be used for a haptic grid. It was shown that models with an attractive function followed by a spring function appeared better suited to target acquisition tasks. Following those results, the proposed method uses the following model:

With R defined as the radius of the attraction sphere around target T , let r be a scalar less than R , x the distance to T and F_{max} the maximum attraction force, the magnitude of the attraction force F is defined as follows (see Eq. 1):

$$\vec{\|F\|} = \begin{cases} \|F_{max}\| \times \sin(\pi/2 \times x/r) & ; x \in [0; r] \\ \|F_{max}\| \times [1 - ((x - r)/(R - r))^2] & ; x \in [r; R] \end{cases} \tag{1}$$

Starting from the border of the attraction sphere, the user is attracted with a force proportional to the squared distance toward the target centre until $x = r$, from which point the force decreases sinusoidally (see Fig. 1). With such a model the attractive force increases very quickly, such 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 freely and easily move around the target region.

Furthermore, to minimize the impact of the distractors, as proposed in [36], the haptic feedback is adjusted according to the requirements of the user. As mentioned previously, because of the quadratic part of our force model, entering in an attraction area is quickly perceived by the user. If this attraction is towards an unwanted target (a distractor), it is expected that the user will tend to oppose the attraction force.

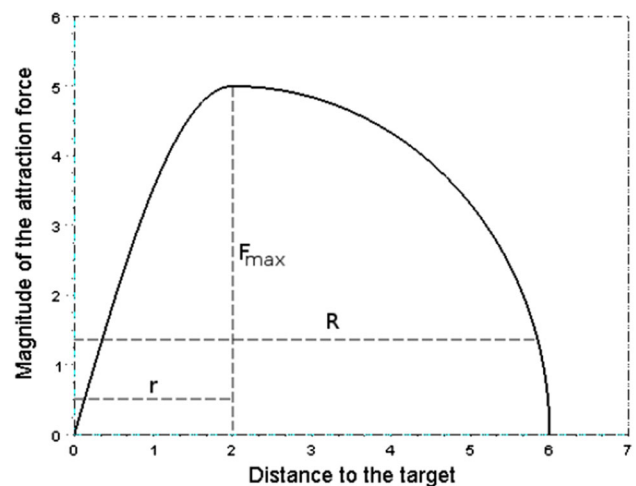


Fig. 1 Representation of the magnitude of the attraction force according to the distance to the target

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 (2) when x is increasing, such as the case when the user moves away from the target for the current attraction sphere.

3.1.3 Haptic signatures: target identification

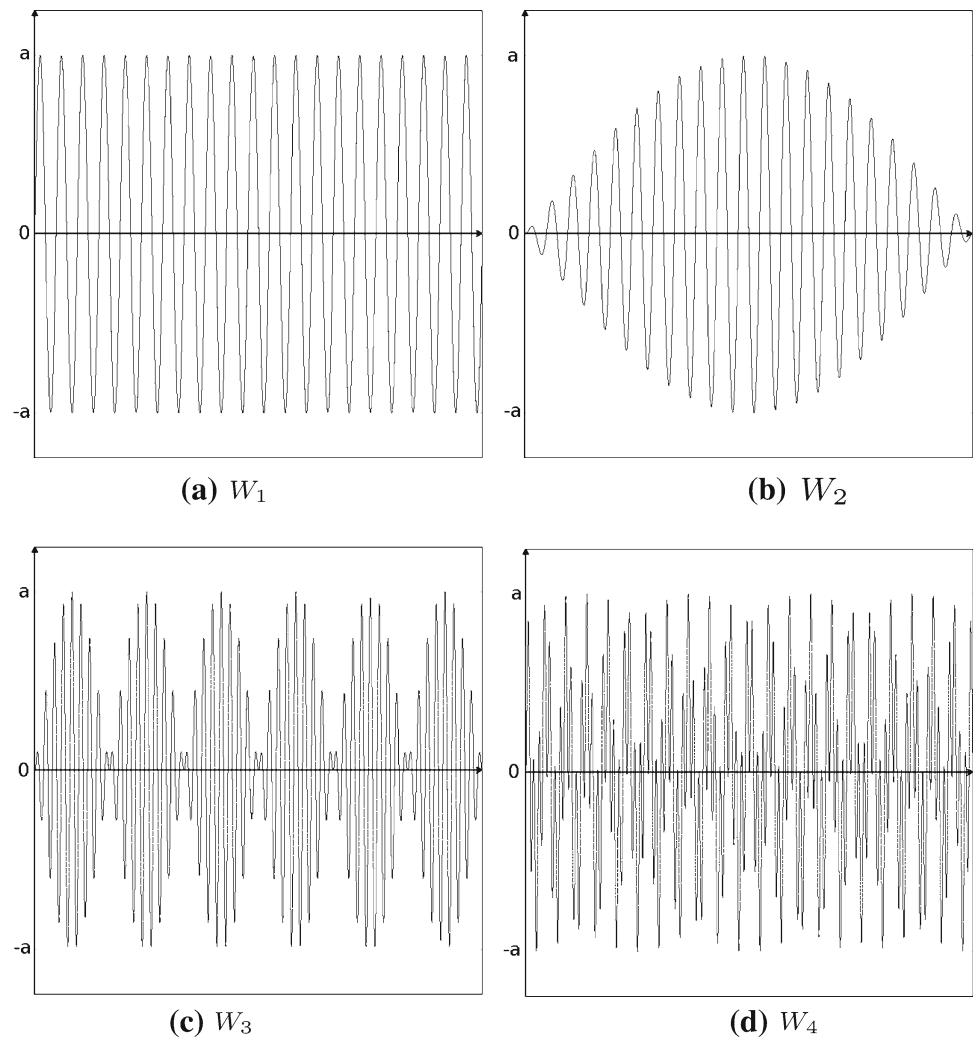
Let us consider an environment containing four targets. In a visual context, each target may be represented by a coloured sphere (red, green, blue, gray for example). Therefore, at any time a user would be able to identify, without any ambiguity, one of the four targets (the red example). The idea of a haptic signature is essential to implement haptic feedback that would serve for identifying (similarly to the visual channel) clearly and accurately each of the four targets. One may see a haptic signature as a *haptic icon* [26]. Further, it is known that humans have an accurate tactile perception. Since vibro-tactile feedback can be an effective memory aid for users with impaired memory [25], a set of vibro-tactile

signatures were created. Based on the guidelines for creation of haptic icons [13], and preliminary evaluations, we define four distinct and clearly identifiable haptic signatures (see Eq. 2). Related works such as waveform amplitude modulation [2, 7, 34] were used.

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

W_1 defines a pure sinusoidal wave, resulting in 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 results in a somewhat rough vibration sensation (see Fig. 2).

Fig. 2 Temporal waveforms of the haptic signatures.



3.2 Proposed auditory feedback

The audio rendering uses a parameter mapping sonification [19] combined with 3D audio spatialisation. The audio signature for each target, an impact sound described in Sect. 3.2.2, is spatially rendered at the target positions. The sonification is applied to the distance using a repetition rate metaphor where repetition rate as well as sound level varies as a function of the scalar distance to the target position. The metaphor proposes two different sonic cues, positional information via spatialization, and distance via repetition variations and sound level distance attenuation.

The repetition rate of the target sound is determined from the distance to each target. The repetition rates are interpolated using a linear scale between 1 Hz, for distances larger than 1 m, to 6 Hz at the actual target position. The level of the signature sounds varied inversely proportional with distance by 20 dB over the same distance range.

3.2.1 Audio spatialization: target localization

Each target sound signature is positioned at the target location using binaural spatialization based on convolution of the signal and the corresponding Head Related Impulse Response (HRIR, see [35]) of the position to be simulated. The spatialization engine used [22] allows for the individualization of Inter-aural Time Difference (ITD) applied to a general Head Related Transfer Function (HRTF) [4] to aid binaural localization performance. While no room reverberation is used, distance attenuation is enhanced to aid in the perception of target distance. The orientation of the user's head is continuously tracked and the relative sound source position is modified accordingly in order to maintain the simulated 3D sound source at its proper stable position in space, irrespective of head movements. To ensure that only movements of the haptic device allow reaching a target, only the orientation of the users head is taken into account in the rendering of audio stereoscopic feedback. Indeed, if the position of the user's head is considered, the user would be able to locate the target just by actually moving in the physical environment of experimentation space. Therefore, the implemented metaphor simulates the fact that the users head is placed at the position of the haptic probe

As humans are capable of listening to several simultaneous audio streams, it is not necessary to apply the notion of activation areas as with haptic feedback for only four concurrent sources [40]. The audio feedback for all targets is always active.

3.2.2 Audio signatures: target identification

Similar to the haptic experimental design, each target is attributed a unique audio signature. We chose four brief

impact sounds, inspired by the haptic signatures, taken from the *freesound* project¹ audio database. Repeating impulsive sounds were chosen due to their ease of localisation as well as to minimise the actual occurrence of concurrent sounds despite have four continuously active signatures. Target W_1 , with a haptic pure tone signature, was paired with the sound of a small bell ($f_0 = 2110$ Hz). W_2 , haptically a pulsating vibration, was paired with the sound of a wood block ($f_0 = 840$ Hz). W_3 , a haptic signature of rapid impulse vibrations, was paired with the impact sound of tapping on a table ($f_0 = 560$ Hz). Finally, W_4 , haptically a rough vibration, was paired with the sound of someone knocking on a window ($f_0 = 140$ Hz). Each sound was chosen so as to be clearly identifiable and distinct from the other audio signatures by considering their fundamental frequency, harmonic response, and timbre. The frequency spectra of the four audio signatures are shown in Fig. 3.

3.3 Use of multimodal redundant rendering

For this evaluation, we want to assess whether the association of both haptic and auditory feedback (bimodal rendering) can improve task performance. As a starting point, both modalities were rendered simultaneously and compared to each modality independently. It is known that several approaches can be considered for the design of a multimodal feedback. The associated feedback can be complementary, redundant, equivalent, specialized, or concurrent [6]. As the goal is to allow users to compare the usefulness of each channel, a redundant bimodal rendering was chosen. As a result, in this multimodal rendering method, the two proposed feedback (haptic and audio) are always available simultaneously.

4 Audio/haptic/multimodal evaluation

This experiment aims at evaluating the proposed haptic and audio rendering for identification, localization, and selection of occluded targets. The effectiveness of haptic, auditory, and the multimodal redundant (haptic and audio) feedback in the completion of the task is computed and analysed.

Three experimental conditions (A , H , M_r) are defined: in condition A (audio) only the auditory feedback (described in Sect. 3.2) is available, in condition H (haptic) only the haptic rendering (described in Sect. 3.1) is provided, while in condition M_r (multimodal redundant) simultaneous haptic and auditory cues are provided.

Due to difficulties in rendering audio sources when they are very close to the head, and because of the limited workspace of the haptic device ($100 \times 90 \times 60$ cm), the

¹ <http://www.freesound.org>.

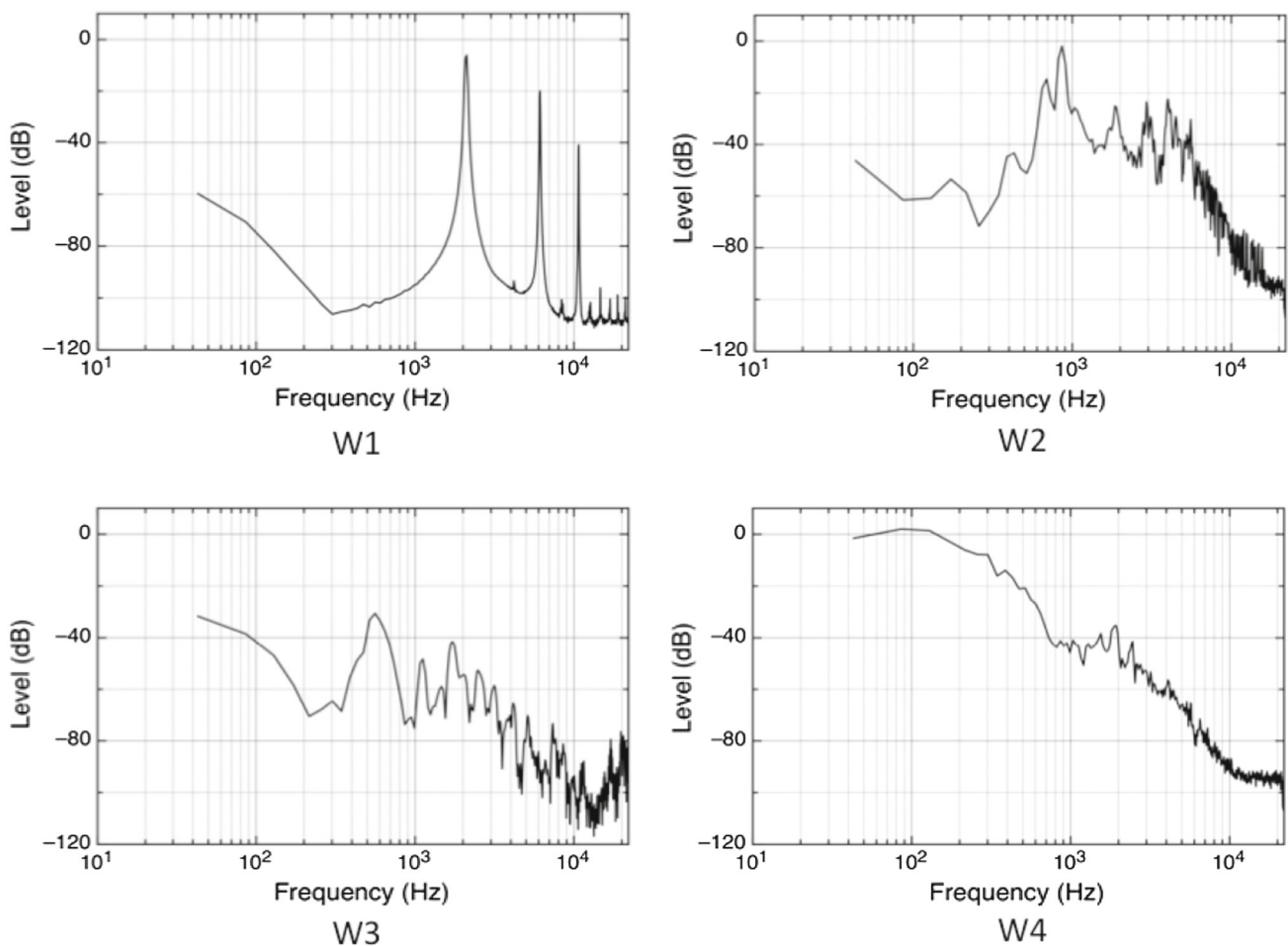


Fig. 3 Spectral representations of the auditory signatures

geometry of the haptic scene ($50 \times 50 \times 50$ cm) was scaled by a factor of $\times 5$ 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 [21]. Selection, combined with individual ITD adaptation, improved the quality of spatial audio rendering for cases when non-individual HRTFs are used.

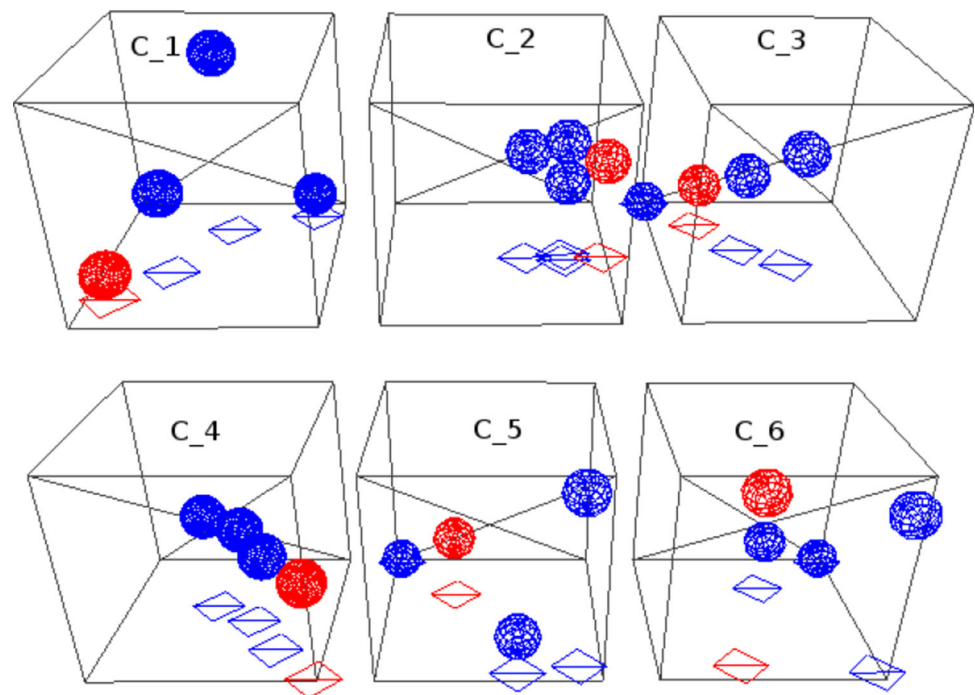
A total of 18 persons (14 male), aged between 23 and 55 took part in this study. Among the subjects, three were researchers in haptics and six were researchers in acoustics. Because of their background, these nine participants are reported as *experts* in what follows. Half of the participants were graduate students. Within the population, two did not have previous experience with haptic devices, and only one did not have past experience with spatialized audio rendering. Six participants were evaluated under *H* and *A* conditions (3 using the order *H* then *A*, 3 others *A* then *H*). Six other participants were presented with the *A* and *M_r* conditions (3 with *A* then *M_r*, 3 with *M_r* then *A*). The final participants

tested *H* and *M_r* conditions (3 with *H* then *M_r*, 3 with *M_r* then *H*).

The experimental setup consisted of the main application, which dealt with the data, and of two other parts related to the haptic and audio rendering. The haptic feedback was rendered via a HAPTION Virtuose 6 DoF device. This haptic interface was selected due to its large working area, well suited to immersive VR systems. The audio rendering was rendered through a wireless stereo headset (Sennheiser RS65). The auditory rendering was implemented in the MaxMSP environment. In all three conditions (*H*, *A*, *M_r*), the displacement of the probe in the 3D 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 update the spatial auditory rendering with respect to head movements, maintaining the simulated 3D sound sources at their proper stable position in space, irrespective of head movements. This information has been added in order to provide users with a natural audio feedback.

Fig. 4 Test configurations: attraction sphere of target (*red*) and attraction sphere of distractors (*blue*) are represented. 2D projections (*diamonds*) are shown for spatial reference. One may note that some targets are located in the projection plane (C_3 , C_5 and C_6)



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

4.1 Experimental design

For each trial, one of the six arrangement configurations shown in Fig. 4 was rendered. The number of targets was limited to four after pre-trials showed memorization of additional targets too difficult. Six test configurations were chosen according to arrangements used in [20]. Four of these six configurations place the three distractors along the path leading to the target from the user's initial starting position, or around it.

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

4.2 Procedure

A 3D acquisition task was used for the experiment. Prior to the experiment, participants received a brief written and oral explanation about the goal of the experiment; a calibration and familiarization phase then followed, after which the test was initiated.

4.2.1 Familiarization phase

In the familiarization phase, subjects were instructed to familiarize themselves with the four haptic and audio sig-



Fig. 5 A user during the experimentation. The virtual workspaces for the familiarization phase are represented by coloured squares. In this case the user was exploring the third signature

natures that will be exploited. At this step, four virtual zones (Z_1 , Z_2 , Z_3 , Z_4) were defined in the middle of the workspace of the haptic device (see Fig. 5). Throughout this step, by freely moving the haptic device, the subject explored the virtual workspace. Whenever the user was in an area Z_i , both haptic and audio signatures corresponding to the zone Z_i were rendered by the system. Once the subjects indicated that they were familiar with the different signatures, the two sample learning configurations were presented (see Fig. 6) in order to let participants familiarize themselves with the test. For this step, participants were asked to identify, locate and select the red target represented in each configuration

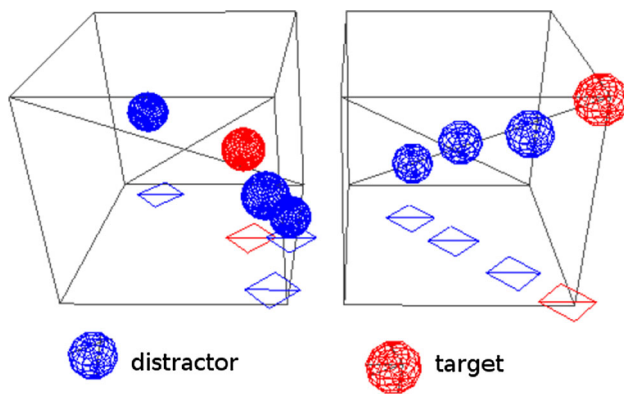


Fig. 6 Learning configurations

of Fig. 6. There was no explicit time limit for the familiarization phase. Participants repeated this task until they felt comfortable in their understanding of the different signatures and how the system functioned. An average of 20 min was observed per subject.

4.2.2 The test phase

The test was divided into three steps, represented in Table 1. In step (i) participants were presented with directional information for each target sequentially in order to help in the construction of a mental spatial map (see [1]). Using this experimental design, both the audio and haptic rendering channels presented the same information and in the same amount of time. This protocol was developed to allow for a haptic only condition to be feasible for such a task, allowing for suitable comparison of results by modality.

On the haptics side, the subject was first attracted in the direction of the location of the target (for a duration of 1.5 s). Thereafter, in addition to the attraction force the haptic signature was also rendered (duration 2 s). At the end of this rendering, the haptic device was pulled back towards the centre of the virtual space in the case of any displacement of the user's hand. Once returned to the centre, there was a pause of 2 s and the process continued to the next target until all four targets were reviewed.

Table 1 Auditory and haptic impulsion feedback schemes for the different test phase steps for A , H , M_r , and M_c conditions

	Audio	Haptic
Step (i) orientation sequential presentation of four targets	Spatialized signature (3 s)	Directional attraction (1.5 s) + directional attraction with haptic signature (2 s)
Step (ii) Target indication	Single non-spatialized target signature	Single haptic signature
Step (iii) Target search	All signatures active and spatialized	Directional attraction + haptic signature when within activation zone
Multimodal redundant M_r	Audio (signature & spatialization) + haptic (signature & attraction)	
Multimodal complementary M_c	Audio (signature & spatialization) + haptic (attraction only)	

On the auditory side, the auditory signature of the activated target was spatially rendered for 3.5 s at a fixed distance in space. After a pause of 2 s, the next target was rendered. Since the position of the user's head was tracked, and the 3D sound field was rendered accordingly, the user could rotate his/her head in order to better detect the direction of target location.

This phase was repeated until the subject indicated that they understood the spatial configuration (directional only) of the different sources. The subject did not know at this stage which of the targets was the true target to find for the task, nor those considered distractors.

In step (ii), the subject was presented with the signature of the target to find (duration of 2 s), without directional information in either modality.

In step (iii) the subject was instructed to exploit the available feedback in order to locate the indicated target's position from step (ii) as fast and as precisely as possible. The subject explored the space using the haptic device (which also serves as the position sensor) until target position selection was achieved. The timing for each trial started when the participant began to move the haptic device, and stopped when the right button of the haptic device was pressed, indicating that the current position had been selected by the subject as that of the indicated target. The experiment lasted 75 min on average for each subject.

4.3 Results and key lessons drawn from this experiment

For each trial, in addition to the task time and final selected position, the entire trajectory was stored. Comparisons were made between the effectiveness of the three tested conditions (H , A , M_r) in terms of completion time (see Table 2) and accuracy (see Table 3) of the selection. In the H condition, we observed an average duration of 18.89 s and 0.115 cm of error. For the M_r condition, the average duration was 26.96 s with an error of 0.14 cm. In contrast, for the A condition, the average duration was 49 s with an average positional error of 0.33 cm.

Table 2 Average task completion time (seconds) for each experimental condition and overall results

	A	H	M_r
C_1	55.02	22.13	30.21
C_2	49.03	33.03	28.96
C_3	42.75	13.11	17.49
C_4	51.53	23.51	42.36
C_5	41.71	7.11	11.48
C_6	56.9	14.47	31.28
Mean	49.49	18.89	26.96
σ	6.25	9.20	10.95

Table 3 Average selection distance error (cm) for each experimental condition and overall results

	A	H	M_r
C_1	0.38	0.12	0.29
C_2	0.33	0.01	0.15
C_3	0.27	0.19	0.02
C_4	0.42	0.19	0.24
C_5	0.27	0.01	0.03
C_6	0.3	0.15	0.14
Mean	0.33	0.11	0.14
σ	0.05	0.08	0.10

After having verified the normality and homogeneity of the variance distribution, the ANOVA and post hoc Tukey were applied to the results of each experimental configuration. For the completion time and the distance error in the selection, a significant difference, ($F_{2,15} = 18.564, p < 0.008$) and ($F_{2,15} = 11.182, p < 0.001$) respectively, was found between the three conditions. A post hoc Tukey highlights that both haptic and multimodal conditions are significantly better than the audio condition ($H-A$ and $M-A$ $p < 0.001$). However, no significant difference was noted between H and M conditions. For the required time and the selection error it can be noted that $p < 0.29$ and $p < 0.77$, respectively. For a more complete discussion of these results, one can refer to [30]. In Table 2, for each experimented condition [haptic (H), audio (A), and multimodal redundant (M_r)] the results of the experiment are summarized.

4.3.1 Haptic condition

Some users noted a slight difficulty in the identification of haptic signatures. They noted that, in contrast to the audio condition, more concentration was required in order to identify haptic signatures (described in subject. 3.1.3). This could be due to the fact that only one haptic signature can be rendered at a time; users must recognize the current signature

and compare it to their haptic memory. This task was simpler in the audio condition, where all signatures were rendered simultaneously.

Moreover, subjects noted that the first step (sequential target direction presentation) of the experimentation was essential for creating a mental map of the scene, allowing them to mentally visualize the spatial arrangement of the different targets. In the case represented in Fig. 4(C_3), following this first step the subject would be informed that one target (T_2) was located at the back right whereas the three others (T_1, T_3, T_4) were at the front left along the same direction (on a diagonal of the cube). Thus in the second step, if the target of interest was located on the right (T_1), the subject did not only know the direction where to go, but he/she could also anticipate that two other targets (T_3 and T_4) might be encountered in the same direction. It is thus understood that the haptic condition, as proposed in this study through the use of arbitrary signatures, necessitated some memorization effort. It is interesting to note that with the presence of the mental map of the environment, non-designated targets (distractors) can appear as points of reference for navigation and therefore they may help in achieving the task.

Finally, subjects reported the simplicity of letting themselves be guided by the haptic device toward the designated target. Before performing the actual task, they could estimate that this interaction could assist them better in their selection process.

Two examples that highlight these observations are illustrated in Fig. 7. On the left the subject went directly towards the desired target. On the right, it seems that the subject has forgotten the configuration, a possible explanation for why he/she seems to be randomly searching for the target of interest.

4.3.2 Audio condition

In the audio condition A, in contrast to the haptic condition, subjects noted that the first stage was not deemed necessary. This is understandable, since in the exploration stage they had enough information for exploring the environment. This

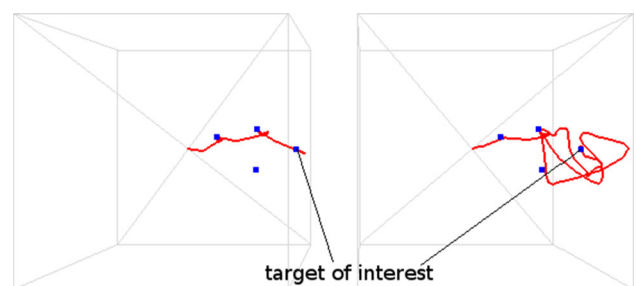


Fig. 7 Two trajectories described by the same user in the same configuration (C_2) in the haptic condition, H

phase was only included in order to have a common protocol for all modality conditions. Subjects did not feel the need to remember the spatial configuration. Moreover, when analysing subjects' comments, it is possible to divide the exploration step into two distinct stages: the approach to the area where the target is located followed by the acquisition of the target position.

1. Approaching the target area

To approach a target, exploration strategies seem to differ according to participant's familiarity with 3D audio rendering. While expert users simply listened to the four sources (without moving their head very much) in order to determine with little difficulty where the target source was located, non-expert users had to take advantage of the head tracking system. It is only through the use of head movements that they were able to approach the area of the target of interest.

2. Precise target selection

Unlike the step of approaching the target, the selection of the target was difficult in the A condition for all subjects. It was not easy for subjects to clearly determine when the audio feedback was at its highest level both in terms of frequency and amplitude. The difference in scale factors between the positional input and the audio feedback environment may also have contributed to confusions. Thus, it was difficult to precisely locate the position of the target of interest. For that, subjects exploited small displacements around the target while assessing changes of the auditory feedback. Doing this, subjects employed a step by step approach to get closer to the target. This strategy explains the considerable time, and observed errors in the selection.

The experimental design ensured that all target signatures remained at least slightly audible throughout the workspace, while not becoming excessively loud at the target position. Furthermore, the design allowed for the repetition rates to not become excessively fast or slow. These experimental choices explain the users comments that have been reported above, and all these comments are confirmed by an analysis of the subjects' trajectories. Figure 8 represents the trajectories described by two subjects with different backgrounds in 3D audio rendering. In the first part of the trajectory the expert was able to move directly toward the target while the naive subject followed a very different path. Moreover, it can be noted that both subjects had difficulty in pinpointing the position of the target of interest.

4.3.3 Multimodal redundant condition

The multimodal condition appears to be the most appreciated by the subjects. This choice can be explained considering that

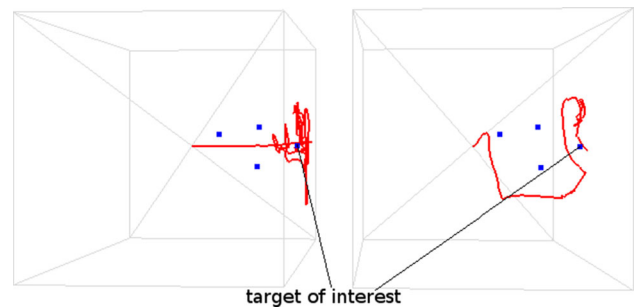


Fig. 8 Two trajectories described by two users with different backgrounds in the same configuration (C_2) in the auditory condition, A

the multimodal condition gathers benefits from the two other conditions. It was noted that the first step seemed unnecessary with the presence of the audio feedback. On the other hand, thanks to the effect of the virtual magnet, the desired target was easily selected. The haptic signature was moderately popular with users. For some, it was used as a sort of confirmation of the audio signature of the target, while it appeared a little disturbing for others.

4.4 Discussion

From the previous evaluation, it appears that the audio channel provided a useful means for memorization and identification of specified signatures. Moreover, thanks to the audio feedback, subjects (through different strategies) were able to approach the area containing the target of interest. Whenever the audio feedback was available there was no need for the two-step design in the exploration process. On the other hand, the attractive haptic feedback provided a clear advantage for precise selection of the target position. However, haptic signatures were not as relevant as audio ones. Finally, the multimodal approach, although giving similar results to the haptic-only condition (any significant statistical difference has been noted) was the most appreciated by users. Furthermore, these results show that the multimodal condition has a real potential for acquisition of targets in 3D environments. However, they also suggest that the direct superposition of haptic and audio feedback may not be the best suited option for such an association.

Such observations suggest the need for an optimized multimodal approach than can tackle the acquisition of a desired target among multiple ones.

5 Proposed complementary audio-haptic interaction

Based on the previous experiment, a combined multimodal feedback which associates previously described haptic and auditory feedback is proposed. In addition, the experimental platform has been reduced in complexity to better suit desktop use versus the previous immersive VR architecture.

Although the previous experiment did not allow us to establish a significant difference between haptic and—multimodal redundant—conditions, it clearly pointed out advantages and disadvantages of each channel. Indeed, the auditory feedback is more suited to localization and identification tasks, whereas the attractive force feedback seems to provide effective assistance for precise selections, even in a multi-target context. From this, we are able to formulate the following hypothesis.

For non-visual localization and selection of a given entity of interest among several others in a 3D environment, it appears that:

- The audio rendering should be used for identification and localization of each target.
- The haptic attraction feedback can help at facilitating an accurate selection of a given target.

Therefore, in the proposed multimodal condition we exploit the audio channel for identification and localization of the targets in the 3D environment. The audio feedback is hence enabled throughout the duration of the task. To complement this feedback, whenever the user enters the attraction zone, the attractive feedback described in Sect. 3 becomes active. The haptic signature is no longer employed.

This experiment aims at evaluating the effectiveness of this combined multimodal method. For this, the effectiveness of the multimodal complementary feedback (M_c) is compared to the multimodal redundant feedback (M_r) (presented in Sect. 3).

5.1 Participants

A total of 24 participants (20 male, 4 female) aged between 20 and 40, were asked to participate. No participants from the first experiment participated in the second experiment. Among the subjects, 14 were college and undergraduate students in computer science or in video game programming. Based on our pre-experimental questionnaire, five participants claimed being familiar with spatialized audio rendering and only two had previous experience with haptic devices. Participants received no compensation.

As in the first study, a between subject design was used for the experiment. 12 subjects performed the experiment in the M_c condition while the other half was evaluated in the M_r condition.

5.2 Experimental setup

In the previous experiment, in order to provide a natural interaction, we integrated the tracking of the user's head in the rendering of the binaural sound rather than the orientation of the haptic interface. As such, the rendered sound depended

on two data: the position of the haptic probe interface and the orientation of the user's head. Therefore, the exploitation of auditory cues involved the integration of two non-collocated movements; one with the hand and another with the head. In such a case, a reduction in performance could occur [41]. It is therefore possible that this interaction complexity may increase task completion time as displacement and rotation actions may be performed sequentially rather than as integrated movements. Because of this, in the current experiment position and orientation are both controlled by the position and orientation of haptic device. Therefore, the implemented metaphor mimics a directional microphone at the position of the haptic probe. This protocol allows for direction manipulation of the auditory reference frame using only the hand, which is simpler with regards to implementation, but may require additional cognitive load as natural head movements are no longer taken into account in the spatial auditory perception feedback loop.

In adapting the experimental platform to a desktop style configuration, we have selected a Sensable Phantom Omni haptic device, which has a smaller working area than the Virtuouse arm used in the first study. The audio rendering was provided via stereo headphones (Sennheiser HD429). No tracking system was used.

5.3 Experimental plan and procedure

The same target arrangement configurations from the first experiment (see Fig. 4) were used. As previously, for each session, each configuration was repeated four times while changing the signature assignment for each target.

The same two-phase protocol was employed. During the familiarization phase, the four auditory signals were presented to the user, and when necessary (for half of the participants) the haptic signatures were also rendered. 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.

As opposed to the previous experiment, the test phase was made of only two steps. There was no more need for a sequential presentation of the direction of each target. In other words, step (i) of Table 1 was omitted. Hence, to present the target of interest, the auditory signal, and if necessary the haptic signature (in the M_c condition), of the target of interest was presented (for a duration of 2 s), then the participant had to locate and select the designated target by using auditory and haptic feedback. The subject moved the haptic device (which also served as the position and orientation sensor) to find and select the target position. Timing for each trial started when the participant began to move the haptic device, and terminated when the right button of the haptic device was pressed, indicating that the current position has been identified (selected). The total experiment lasted 35 min on average.

5.4 Results and discussion

Both the initial redundant and combined multimodal methods rely on haptic attraction in order to achieve a precise selection of the target. Because of this similarity, it is not relevant to analyse the selection precision error (the distance between the actual position of the target and that designated by the participant). Therefore, time completion and trajectories carried out by the users have been used as means of comparison.

The subjects unanimously emphasized that the multimodal complementary method was suitable for the completion of the task. Their comments confirmed initial observations collected during the first study. They not only emphasized the benefits of using haptics in a precise selection of targets, but they also highlighted the potential of audio feedback in the process of locating and identifying the chosen target among a number of distractors with occlusion. Some subjects particularly emphasized the fact that, at any moment, they were able to ignore other signals (other targets) presented in the 3D environment. This is illustrated in the trajectory example shown in Fig. 9 where one can observe that distractors had little impact on the task, since only a small deviation is present along the route. This is in contrast to the trajectories represented in Fig. 7. These observations were supported by analysing the duration of the task performance. With the multimodal complementary method (M_c), we indeed observed a significant decrease ($p < 0.04$) in the time required for the selection task when compared to the multimodal redundant method (M_r). A significant improvement in speed was observed, with a mean task completion time difference of 4.72 s, or a gain of 18 %. Table 4 provides a summary of the time to task completion results.

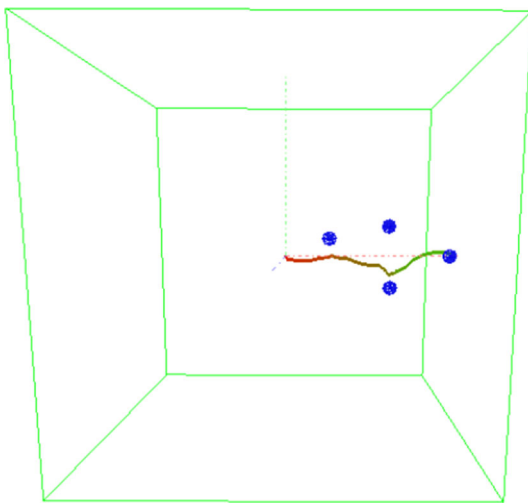


Fig. 9 Trajectory described by a user in the C_2 configuration in the multimodal complementary condition, M_c . The colour code of the trajectory represents the evolution of the displacement through time

Table 4 Average task time for each experimental condition

	M_r	M_c
C_1	32.21	21.09
C_2	29.56	26.10
C_3	19.49	14.02
C_4	37.39	30.50
C_5	12.08	12.98
C_6	26.28	24.02
Mean	26.17	21.45
σ	9.13	6.88

Some of the participants who performed the experiment in the multimodal redundant condition noticed that having both haptic and auditory signatures was sometime useful as it offered two ways for identifying the target of interest. Based on such a comment, this suggests that in such a condition, more cognitive effort was required since two mental representations should be maintained. These observations support the idea that the multimodal complementary method was more suitable for the proposed task.

In a more general way, it is interesting to note that these conclusions are in line with the psychophysical characteristics of the human being in regards to both channels. Regarding the physiological characteristics of haptic perception, various works have indeed supported the idea that haptics is more suited to this particular selection task than to the perception of spatial properties, which is more appropriate for the audio modality [15,32]. Moreover, we note that the design choice concerning the use of the audio modality for identification and localisation of targets of interest is also reinforced by the study presented in [40], which shows that even in an environment presenting 19 sound sources, with appropriate training, listeners were able to get a thorough knowledge of the spatial arrangement of the sources.

6 Conclusion

Many situations of interaction within a 3D numerical environment may necessitate identification, localization, and selection of an entity of interest than may be occluded by one or more distractors. In the literature we observed that many works have addressed such a situation by mean of the visual feedback. Considering the multimodal capabilities of humans, we believe that haptic and auditory feedback can offer a more natural means of interaction for such a task. In this regard, several audio and/or haptic methods have been proposed and evaluated. The originality of our work resides in the fact that the strengths and weaknesses of each channel have been analysed.

In a non-visual 3D environment, audio, haptic, and combined multimodal feedback have been investigated in the

context of the selection of a specified target among a number of distractors in a virtual spatial environment. Following the analysis of results and the comments of participants of the first study, a modified multimodal approach which combines the advantages offered by each channel was proposed and evaluated. The audio feedback allowed subjects to easily locate and distinguish each target in the 3D space and therefore to approach the area of the desired target as desired. On the other hand, haptic feedback provided a useful and effortless guidance towards the precise target position for selection. Analysis of objective and subjective results indicated that the proposed optimized multimodal approach was the best suited for the completion of the task in terms of required time as well as the spatial precision of target selection.

The goal of this study was to evaluate the use of haptic and audio modalities to accomplish spatial identification, localisation, and selection tasks, thereby reducing the workload of the visual channel from complex situations. As a future work, experimental studies will be carried out in order to determine how good an audio/haptic rendering can perform when compared to a visual condition. To achieve this goal, we are investigating an audio/haptic association that can offer the best condition for localization and selection tasks. In particular, the audio and haptic cues will be further developed and evaluated in isolation before the multimodal association.

At the same time, based on the experimental results reported here, we are now developing a *Social Soccer Game* which takes advantage of multimodal feedback to complement the visual feedback in three different situations. Firstly, when the ball is called by a team-mate, the sound of the running player is rendered in 3D to the player having the ball. The binaural audio aspect is exploited in order to let the player who has the ball know in which direction the ball should be kicked. Secondly, each interaction with the ball is perceived via vibrotactile rendering. Hence, controlling the ball is conveyed via a continuous haptic feedback while a kick is haptically rendered throughout a sawtooth wave. In contrast to the study reported here, the vibrotactile feedback will be presented to the sole of the foot of the user. We are currently running preliminary evaluations in order to design a set of easily identifiable tactons in this multimodal (visual, audio, haptic) rendering condition.

References

- Afonso A, Blum A, Katz B, Tarroux P, Borst G, Denis M (2010) Structural properties of spatial representations in blind people: scanning images constructed from haptic exploration or from locomotion in a 3-D audio virtual environment. *Memory Cognit* 38:591–604. doi:[10.3758/MC.38.5.591](https://doi.org/10.3758/MC.38.5.591)
- Ammi M, Katz B (2011) Design of haptic stimuli for audio-haptic concurrent coupling. In: IEEE international symposium on haptic audio-visual environments and games. Hebei, China, pp 74–80
- Ammi M, Katz B (2012) Audio-haptic intermodal coupling for comparative search tasks. In: IEEE haptics symposium (HAPTICS 2012). Vancouver, Canada, pp 307–313
- Aussal M, Alouges F, Katz BF (2012) ITD interpolation and personalization for binaural synthesis using spherical harmonics. In: Audio engineering society UK conference. New York, UK, pp 04
- de Berg M, van Kreveld M, Overmars M, Schwarzkopf O (2000) Computational geometry. In: Algorithm. Chapter 7: Voronoi diagrams, 2nd revised edn. Springer
- Bowman D, Kruijff E, LaViola J, Poupyrev I (2004) 3D User interfaces: theory and practice. Addison-Wesley
- Brown L, Brewster S, Purchase H (2005) A first investigation into the effectiveness of tactons. In: Proceedings first joint Eurohaptics conference and symposium on haptic interfaces for virtual environment and teleoperator systems world haptics 2005, pp 167–176. doi:[10.1109/WHC.2005.6](https://doi.org/10.1109/WHC.2005.6)
- Carpendale MST, Cowperthwaite DJ, Fracchia FD (1997) Extending distortion viewing from 2d to 3d. *IEEE Compu Graph Appl* 17(4):42–51
- Cockburn A, Brewster S (2005) Multimodal feedback for the acquisition of small targets. *Ergonomics* 48(9):1129–1150
- Dennerlein J, Yang M (1999) Perceived musculoskeletal loading during use of a force-feedback computer mouse. In: Human factors and ergonomics society annual conference, Houston
- Dennerlein J, Yang MC (2001) Haptic force feedback devices for the office computer: performance and musculoskeletal loading issues. *Hum Factors* 43(2):278–286
- Elmqvist N, Tsigas P (2006) View projection animation for occlusion reduction. In: AVI 06: Proceedings of the working conference on advanced visual interfaces. ACM, NY, pp 471–475. doi:[10.1145/1133265.1133361](https://doi.org/10.1145/1133265.1133361)
- Enriquez MJ, MacLean KE (2003) The hapticon editor: a tool in support of haptic communication research. In: HAPTICS' 03. IEEE Computer Society, Los Angeles, p 356
- Flasar J, Sochor J (2007) Manipulating objects behind obstacles. In: Lecture notes in computer science, vol 4563. pp 32–41
- Friedes D (1974) Human information processing and sensory modality: cross-modal functions, information complexity, memory and deficit. *Psychol Bull* 81:284–310
- Grosjean J, Coquillart S (1999) The magic mirror: a metaphor for assisting the exploration of virtual worlds. In: Proceedings spring conference on computer graphics
- Hasser C, Goldenberg A (1998) User performance in a gui pointing task with a low-cost force-feedback computer mouse. In: Seventh annual symposium on haptic interfaces. International mechanical engineering congress and exposition, Anaheim, CA
- Hawley LM, Litovsky RY, Colburn HS (1999) Speech intelligibility and localization in a multi-source environment. *J Acoust Soc Am* 105(6):3236–3448
- Hermann T, Ritter H (1999) Listen to your data: model-based sonification for data analysis. In: Advances in intelligent computing and multimedia systems, G. E. Lasker, Baden-Baden, pp 189–194
- Hwang F, Keates S, Langdon P, Clarkson PJ (2003) Multiple haptic targets for motion-impaired computer users. In: CHI '03: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 41–48
- Katz B, Parseihian G (2012) Perceptually based head-related transfer function database optimization. *J Acoust Soc Am* 131(2):EL99-EL105
- Katz B, Rio E, Picinali L (2010) LIMSI spatialisation engine. In: International deposit digital number IDN.FR.001.340014.000.S.P.2010.000.31235.
- Katz BF, Kammoun S, Parseihian G, Gutierrez O, Brilhault A, Auvray M, Truillet P, Denis M, Thorpe S, Jouffrais C (2012) NAVIG: augmented reality guidance system for the visually

- impaired. *Virtual Real* 16(4):253–269. doi:[10.1007/s10055-012-0213-6](https://doi.org/10.1007/s10055-012-0213-6)
24. Kim SC, Kwon DS (2007) Haptic and sound grid for enhanced positioning in a 3-d virtual environment. In: Proceedings of haptic and audio interaction design. Second international workshop, HAID vol 2007 no. 1. pp 98–109
 25. Kuznetsov S, Dey AK, Hudson SE (2009) The effectiveness of haptic cues as an assistive technology for human memory. In: Proceedings of the 7th international conference on pervasive computing. Pervasive '09, Springer, Berlin, pp 168–175
 26. Maclean K, Enriquez M (2003) Perceptual design of haptic icons. In: Proceedings of Eurohaptics, pp 351–363
 27. Magnusson C, Danielsson H, Rasmus-Gröhn K (2006) Non visual haptic audio tools for virtual environments. In: Proceedings of the first international conference on haptic and audio interaction design, HAID'06, pp 111–120
 28. Menelas B, Ammi M, Bourdot P (2008) A flexible method for haptic rendering of isosurface from volumetric data. In: Proceedings of the 6th international conference on Haptics: perception. Devices and scenarios, EuroHaptics '08, Springer, Berlin, pp 687–693
 29. Menelas B, Ammi M, Pastur L, Bourdot P (2009) Haptical exploration of an unsteady flow. In: Proceedings and symposium on haptic interfaces for virtual environment and teleoperator systems EuroHaptics conference world haptics 2009. Third Joint, pp 232–237
 30. Menelas B, Picinali L, Katz B, Bourdot P (2010) Audio haptic feedbacks in a task of targets acquisition. In: IEEE symposium on 3D user interfaces (3DUI 2010). Waltham, pp 51–54. doi:[10.1109/3DUI.2010.5444722](https://doi.org/10.1109/3DUI.2010.5444722)
 31. Menelas B, Picinali L, Katz B, Bourdot P, Ammi M (2009) Haptic audio guidance for target selection in a virtual environment. In: 4th Intl. haptic and auditory interaction design workshop (HAID'09), Dresden, pp 12–13
 32. Menelas BAJ (2013) Innovative approaches of data visualization and visual analytics, chap. In: *Virtual reality technologies (visual, haptics and audio) in large datasets analysis*. IGI Global, pp 225–250
 33. Menelas BAJ (2013) Interactive analysis of cavity-flows in a virtual environment. In: Proceedings of the 28th spring conference on computer graphics. SCCG '12, ACM, NY, pp 31–37, doi:[10.1145/2448531.2448535](https://doi.org/10.1145/2448531.2448535)
 34. Menelas, BAJ, Otis MJD (2012) Design of a serious game for learning vibrotactile messages. In: *Haptic audio visual environments and games*. IEEE, pp 124–129
 35. Moller H, Sorensen M, Hammershoi D, Jensen C (1995) Head-related transfer functions of human subjects. *J Audio Eng Soc* 45(5):300–321
 36. Oakley I, Brewster S, Gray P (2001) Solving multi-target haptic problems in menu interaction. In: CHI '01: CHI '01 extended abstracts on human factors in computing systems, pp 357–358
 37. Oakley I, McGee MR, Brewster S, Gray P (2000) Putting the feel in 'look and feel'. In: CHI '00: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, NY, pp 415–422
 38. Picinali L, Menelas B, Katz B, Bourdot P (2010) Evaluation of a haptic / audio system for 3-D targeting tasks. In: 128th Conv. audio engineering society. London, pp 1–11
 39. Picon F, Ammi M, Bourdot P (2008) Force model for CAD selection. In: Feiner S, Thalmann D, Guitton P, Frhlich B, Kruijff E, Hachet M (eds) VRST, ACM, pp 283–284. doi:[10.1145/1450579.1450654](https://doi.org/10.1145/1450579.1450654)
 40. Simpson BD, Brungart DS, Nandini I (2009) Sound localization in multisource environments. In: Tech. rep., Pentagon report, human effectiveness directorate Wright-Patterson AFB OH 711 human performance wing
 41. Swapp D, Pawar V, Loscos C (2006) Interaction with co-located haptic feedback in virtual reality. *Virtual Real* 10(1):24–30
 42. Vanacken L, Grossman T, Coninx K (2009) Multimodal selection techniques for dense and occluded 3d virtual environments. *Int J Hum Comput Stud* 67(3):237–255
 43. Vanacken L, Raymaekers C, Coninx K (2006) Evaluating the influence of multimodal feedback on egocentric selection metaphors in virtual environments. In: Proceedings of the first international conference on haptic and audio interaction design. HAID'06, Springer, Berlin, pp 12–23
 44. Wall SA, Paynter K, Shillito AM, Wright M, Scali S (2002) The effect of haptic feedback and stereo graphics in a 3d target acquisition. In: Proceedings of eurohaptics, Edinburgh
 45. Yamada T, Tsubouchi D, Ogi T, Hirose M (2002) Desk-sized immersive workplace using force feedback grid interface. In: Proceedings IEEE. Virtual Reality, pp 135–142