



Haptic feedback combined with movement sonification using a friction sound improves task performance in a virtual throwing task

Emma Frid¹ · Jonas Moll² · Roberto Bresin¹ · Eva-Lotta Sallnäs Pysander¹

Received: 8 July 2017 / Accepted: 18 April 2018 / Published online: 9 May 2018
© The Author(s) 2018 corrected publication September/2018

Abstract

In this paper we present a study on the effects of auditory- and haptic feedback in a virtual throwing task performed with a point-based haptic device. The main research objective was to investigate if and how task performance and perceived intuitiveness is affected when interactive sonification and/or haptic feedback is used to provide real-time feedback about a movement performed in a 3D virtual environment. Emphasis was put on task solving efficiency and subjective accounts of participants' experiences of the multimodal interaction in different conditions. The experiment used a within-subjects design in which the participants solved the same task in different conditions: visual-only, visuohaptic, audiovisual and audiovisuohaptic. Two different sound models were implemented and compared. Significantly lower error rates were obtained in the audiovisuohaptic condition involving movement sonification based on a physical model of friction, compared to the visual-only condition. Moreover, a significant increase in perceived intuitiveness was observed for most conditions involving haptic and/or auditory feedback, compared to the visual-only condition. The main finding of this study is that multimodal feedback can not only improve perceived intuitiveness of an interface but that certain combinations of haptic feedback and movement sonification can also contribute with performance-enhancing properties. This highlights the importance of carefully designing feedback combinations for interactive applications.

Keywords Sonification · Haptic feedback · Multimodal interaction · Task performance

This work was supported by the Swedish Research Council (Grant No. D0511301). It was also partially funded by a grant to Roberto Bresin from KTH Royal Institute of Technology.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12193-018-0264-4>) contains supplementary material, which is available to authorized users.

✉ Emma Frid
emmafrid@kth.se

Jonas Moll
jonas.moll@it.uu.se

Roberto Bresin
roberto@kth.se

Eva-Lotta Sallnäs Pysander
evalotta@kth.se

¹ KTH Royal Institute of Technology, Lindstedsvägen 3, Stockholm, Sweden

² Uppsala University, Lägerhyddsvägen 2, Uppsala, Sweden

1 Introduction

Modern interfaces are becoming increasingly interactive and multimodal. Multimodal interfaces enable information exploration through novel and rich ways, for example through interactive sonification, interactive visualization or haptic exploration. Considerable research has already been devoted to multimodal interfaces, see e.g. [4] for an overview of the effect of visual-auditory and visual-tactile feedback on user performance and [50] for a review on unimodal visual-, auditory-, haptic- feedback in motor learning. However, few studies have focused on interfaces providing concurrent trimodal feedback involving visual feedback, 3D haptics and movement sonification, or comparisons between modalities in such contexts. To our knowledge, studies focusing on multimodal interfaces providing both auditory feedback and haptic feedback rendered with a point-based haptic device have largely focused on simple sound models or discrete feedback sounds (e.g. sine waves [18], beeps or clicks [38], speech synthesis and short MIDI notes [58], animal sounds [41]) or on sonification of position [6,18,24].

Accordingly, little attention has been paid to the use of continuous movement sonification and more complex sound models (e.g. physical modeling or granular synthesis) or the effect of such auditory feedback on task performance and perceived interaction.

In this paper we present an exploratory study on the effect of auditory- and haptic feedback in a virtual throwing task. Emphasis was put on task solving efficiency and subjective accounts of participants' experiences of the multimodal interaction in different conditions. More dimensions are covered in [39]. The novelty of the current study lies in its systematic comparison between, and added value of, complex and continuous multimodal feedback. The current study builds on our previous work on haptic interfaces involving simple auditory icons [37,38], indicating that continuous feedback could possibly provide added value to haptic interfaces designed for dynamic tasks. In the following sections we present the relevant background to our research, followed by a description of our experiment.

2 Background

2.1 Haptic feedback

Haptic interfaces provide feedback to our sense of touch. The human haptic system uses sensory information derived from mechanoreceptors and thermoreceptors embedded in the skin (cutaneous inputs) together with mechanoreceptors embedded in muscles, tendons, and joints (kinesthetic inputs) [29]. The cutaneous inputs contribute to the human perception of various sensations such as e.g. pressure, vibration, skin deformation and temperature, whereas the kinesthetic inputs contribute to the human perception of limb position and limb movement in space [29]. The relatively modern research field *computer haptics* is concerned with generating and rendering haptic stimuli to the user, just as computer graphics deals with generating and rendering visual images [53]. The term *haptic feedback* relates to touch sensations that are produced as a response to user actions. With modern haptic feedback equipment such as the point-based PHANTOM-device used in the study presented in this paper, it is possible to simulate complex shapes, texture, friction, forces, weight and inertia. Studies have shown that task performance, as well as the sense of presence and awareness, can improve significantly when haptic feedback is added to a visual interface [30,36,45].

2.2 Auditory feedback

The term *auditory feedback* relates to sounds that are produced in response to user actions. There are several means of incorporating auditory feedback in computer interfaces; soni-

fication [21,27], auditory icons [16] and earcons [3]. *Sonification* is defined as “the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation” [27]. This definition was further expanded by Hermann [20] to “*Sonification is the data dependent generation of sound, if the transformation is systematic, objective and reproducible, so that it can be used as scientific method*”. Sonification has been extensively applied to map physical dimensions into auditory ones (see [11] for an overview). Interactive sonification is an emerging field of research that focuses on the interactive representation of data by means of sound. It can be particularly useful for exploration of data that changes over time, for example in inspection of body movement data, and has proved to have positive effects on motor learning [12,13,49].

2.3 Cross-modal perception and interaction

Multimodal feedback refers to feedback presented to two or more modalities, i.e. feedback that stimulates several senses. The term *cross-modal interaction* can be defined as the process of coordinating such information from different sensory channels into a final percept [10] and relates to the continuous integration of information from e.g. vision, hearing and touch. In a meta-analysis of 43 studies on the effects of multimodal feedback on user performance, adding an additional modality (auditory- or tactile feedback) to visual feedback was found to improve performance overall [4].

When processing multisensory information, the brain is faced with the challenge of correctly categorizing and interpreting many sensory stimulations simultaneously. Incongruities between sensory inputs can result in unexpected perceptual effects and illusions (see e.g. the McGurk [33] or the size-weight haptic and visual illusions [26]). In fact, studies suggest that cross-modal interactions are the rule and not the exception; cortical pathways are modulated by signals from other modalities, they are not sensory specific [48]. The neural processes that are involved in synthesizing information from cross-modal stimuli are usually referred to as *multisensory integration* [54]. The response to a cross-modal stimulus can be greater than the response to the most effective of its component stimuli, resulting in multisensory enhancement, or smaller than the response to the most effective of its component stimuli, resulting in multisensory depression. Multisensory integration can thus alter the salience of cross-modal events. [54]. If correctly designed, crossmodal stimuli can result in enhancement of certain sensorial aspects, creating experiences beyond what is possible using a unimodal stimulus.

Numerous studies on multimodal feedback have focused on interactions between the auditory and tactile modality, see e.g [25] for an overview. Findings show that auditory

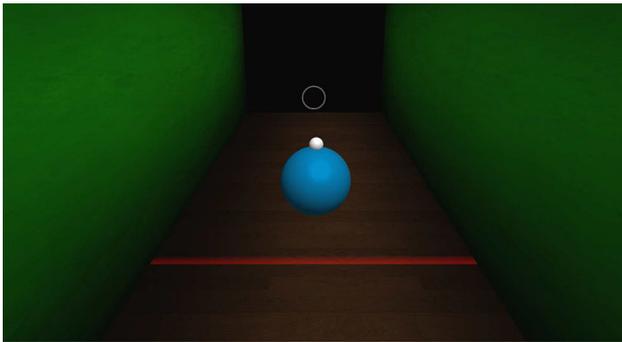


Fig. 1 Screenshot of the graphical user interface. The blue sphere is the haptic object, i.e. the ball, the white dot is the cursor and the red-line is an indication of the position where one should release the ball. The gray circle represents the target area, i.e. the goal

information may influence touch perception in systematic ways [17,22,23,43,47]. The tactile sense has been found to be able to activate the human auditory cortex [46] and several studies report generation of haptic sensations by manipulation of auditory feedback [1,15,19,23,31].

A term related to the topic of cross-modal interaction is *semantic cross-modal congruence*, i.e. that information from one sensory modality is consistent with information from a different sensory modality. Semantic congruence usually refers to pairs of auditory- and visual stimuli that vary (match versus mismatch) in terms of their identity and/or meaning [52]. For example, a semantically congruent bisensory stimulus could be a picture of a particular animal, paired with the sound of the same animal, as in [34]. Semantically congruent cross-modal stimulation has been shown to improve behavioral performance [28].

3 Method

3.1 Experimental setup

A haptic- and graphical interface providing real-time auditory feedback was developed. The interface was designed to allow the user to grasp a ball (a graphical sphere) in a 3D environment and throw it into a circular goal situated on the opposite side of a virtual room (see Fig. 1). The circular goal changed from a grey to green color when a successful hit was achieved. When the goal was not successfully hit, the circle changed from a grey to a red color. We opted for a throwing gesture since this provided us with an intuitive and simple movement task. This movement could also be intuitively sonified. We define the movement performed in this study as a “virtual throwing gesture”, since the haptic device that was used has limited possibilities in terms of affording a “real” throwing gesture.

The experimental setup is illustrated in Fig. 2. A desktop computer (Dell Precision Workstation T3500, Intel® Xeon®

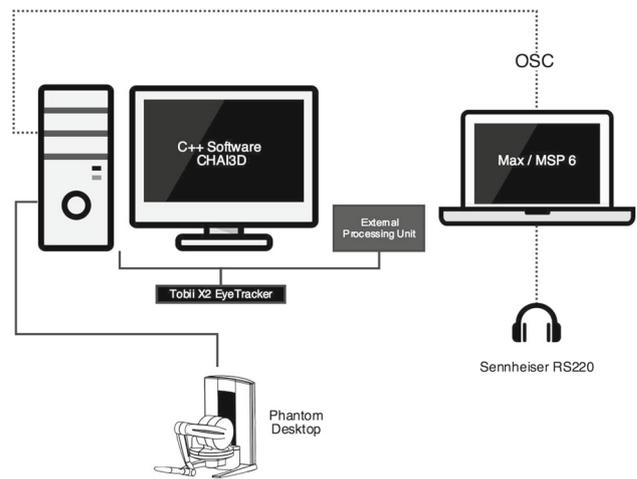


Fig. 2 Experimental setup

CPU E5507 2.27 GHz, 4 CPUs) was used to run the haptic interface. A laptop computer (Macbook Pro, 2.8 GHz Intel Core i5) was used for sound generation and logging of task performance measures. The graphic and haptic application was written in C++ and based on the open source Chai3D library [5]. Communication between the desktop and laptop computer, enabling real-time sonification of the user’s interaction with the ball, was established using the Open Sound Control protocol (OSC) [57]. Sound was synthesized in Max/MSP [8] and a pair of Sennheiser RS220 headphones were used to provide auditory feedback. A SensAble™ PHANToM® Desktop haptic device was used for haptic rendering. The PHANToM-device is a one-point-interaction haptic tool with three degrees of freedom that has a pen-like stylus attached to a robotic arm. When the proxy, associated with the pen’s tip (hence the term ‘point-based haptic device’), coincides with a virtual object, forces are generated so as to simulate touch. When the user picks up an object by touching it and pushing a button, weight and inertia, as well as collisions with other objects, can be simulated. The experiment was recorded using a video camera. A Tobii Pro X2-60 eye tracking device was also included in the setup in order to capture gaze data (a detailed discussion on gaze behavior in this context is presented in [39]).

3.2 Pilot experiments

Two preparatory experiments were carried out prior to the study in order to detect potential pitfalls in the experimental design and improve the design of multimodal stimuli. The first experiment was a vocal sketching experiment that was carried out as an initial step in the design process, with the aim of identifying two perceptually different sonification models. Vocal sketching is an extension of *vocal prototyping*, a process in which users vocalize gesture-sound interactions [2].

It has proved to be a feasible method for producing rich outcomes in terms of sonic designs [14]. Five participants (age = 23–28, $M = 26.6$; 2 female) interacted in the haptic environment described in Sect. 3.1, sketching interactions with the haptic tool using their voice. We wanted to design two sonification models that would be semantically congruent with a light versus a heavy ball, for the nonhaptic versus haptic condition. Participants were therefore instructed to vocalize sounds that they associated with the following categories: “high effort and weight” versus “low effort and no weight”. Results were used as design implications for the design of two perceptually different sonification models: one model reminiscent of a swishing sound and one model reminiscent of the sound of creaking wood. Several participants commented that the interaction would be a lot more intuitive if the auditory feedback, apart from sonifying the movement, also would indicate whether you scored a goal or if the ball bounced somewhere. Sounds indicating these events were therefore also included in the final sound design.

A second pilot experiment was carried out in order to ensure that the ball was perceived as weightless in the nonhaptic conditions and that the weight of the ball was set to an appropriate level in the haptic conditions. Without looking at the screen, three participants (age = 26–58, $M = 38$; 1 female) interacted with the ball in the application described in Sect. 3.1 in three different conditions presented in a randomized order: one haptic, one nonhaptic and one condition in which the application and the PHANToM-device was switched off (i.e. when only the weight of the physical stylus could be perceived). Participants rated the perceived weight of the ball in all conditions (scale 0–100). Results confirmed that there was no difference between the nonhaptic condition and when the device was switched off. The weight of the ball (in haptic conditions) was slightly reduced for the final experiment, since participants had commented that the ball was too heavy.

3.3 Experimental design

Using a within groups design, we evaluated the effect of auditory- and haptic feedback on a set of parameters in six different conditions: visual only, i.e. visuo (V), audiovisual with a creaking sound (AV^1), audiovisual with a swishing sound (AV^2), visuohaptic (VH), audiovisuohaptic with a creaking sound (AVH^1) and audiovisuohaptic with a swishing sound (AVH^2). The effect of feedback type was evaluated in terms of task performance and perceived intuitiveness. The experiment is described in detail below.

3.3.1 Participants

A total of 20 participants took part in the experiment (age = 19–25 years, $M = 20.40$ years; 9 female). None of

them had used a haptic device before or participated in any of the pilot experiments. None reported having any hearing deficiencies or reduced touch sensitivity.

3.3.2 Compliance with ethical standards

All participants gave written consent for participation. Participants in the pilot experiments were also recruited from the students and staff at KTH. They did not receive any monetary compensation. Participants who took part in the final experiment were also recruited from the students at KTH (invitations were sent out to first-year students enrolled in the Computer Science Program). They were compensated with cinema tickets. All participants consented to their data being collected. At the time of the experiment, no Ethics Approval was required from KTH for behavioral studies such as the one reported in this paper. For the management of participants' personal data, we followed regulations according to the KTH Royal Institute of Technology's Ethics Officer.

3.3.3 Procedure

The experiment was conducted by two instructors. Instructions were presented in written format. The experiment consisted of the following parts: (1) general introduction, (2) three test sessions (conditions), (3) 10 min break, (4) three test sessions, (5) a concluding semi-structured interview. The presentation order of conditions was randomized for each participant. Each test session began with a practice trial (minimum duration 2 min, maximum duration 4 min). The practice trials were included in order to reduce the risk of potential learning effects. These trials enabled participants to familiarize themselves with the task in the current feedback condition. All participants practiced less than 4 min.

For each condition, participants were allowed to begin with the actual task (the test session) once they had indicated that they had understood how the task should be performed in the current condition. The application was then restarted and participants were instructed to continue performing the throwing gesture until the instructors announced that the task was completed. Task completion was defined as the time when the goal had been successfully hit 15 times, but the participant was not aware of this definition. After each condition, the participant filled out a questionnaire, rating the intuitiveness of the interaction in the particular condition. In order for us to gather qualitative data on perceived interaction in respective condition, a subset of the participants ($n = 13$) were also instructed to *think aloud* [51] when filling out the questionnaire. During the concluding interview, performed at the end of the experimental session, questions were asked about the following themes:

- Strategies used when aiming towards the target
- Perceived effects of auditory- versus haptic feedback

3.3.4 Auditory feedback

The sound design was guided by findings from the vocal sketching pilot experiment, see Sect. 3.2. Two types of auditory feedback were implemented: (a) feedback concerning the gesture and aiming movement, i.e. movement sonification, as well as sounds indicating a successful grasp of the ball, and (b) feedback from the point when the user is no longer in control of the ball, i.e. bouncing sounds and sounds indicating if you score versus miss the goal. Feedback (a) relates to the human control of the ball, whereas feedback (b) relates more to functions and overall experience of the interface. These sounds were included in order to provide a more realistic sensory experience and to motivate the participant by adding gamification elements, i.e. sounds effects similar to those present in an arcade game. The bouncing sounds also provide a more ecological consistent context to the whole experiment.

No new information was added in conditions with auditory feedback, rather, provided visual feedback was augmented with sound. A summary of actions/movement features and the corresponding auditory feedback is presented in

Table 1 Auditory feedback

Event	Auditory feedback
	<i>Movement sonification</i>
Aiming movement 1)	velZ → amplitude posX → stereo panning posY → frequency posZ → rubbing force
Aiming movement 2)	velZ → amplitude posX → stereo panning posY → comb-filter delay time posZ → comb-filter feedback gain
Ball bouncing on:	<i>Impact sounds</i>
Floor 1)	Heavy wooden impact sound
Floor 2)	Light impact sound
Wall 1), 2)	Metallic impact sound
	<i>Earcons and auditory icons</i>
Scoring a goal 1), 2)	MIDI sequence, increasing pitch
Missing the goal 1), 2)	MIDI sequence, decreasing pitch
Grasping the ball 1), 2)	Band-pass-filtered noise with Sweeping f0 frequency + click

Auditory feedback and mapping for sound model 1) (creaking sound) and 2) (swishing sound).

The following abbreviations are used: *velZ* velocity along the z-axis, *posX* position along the x-axis, and so on

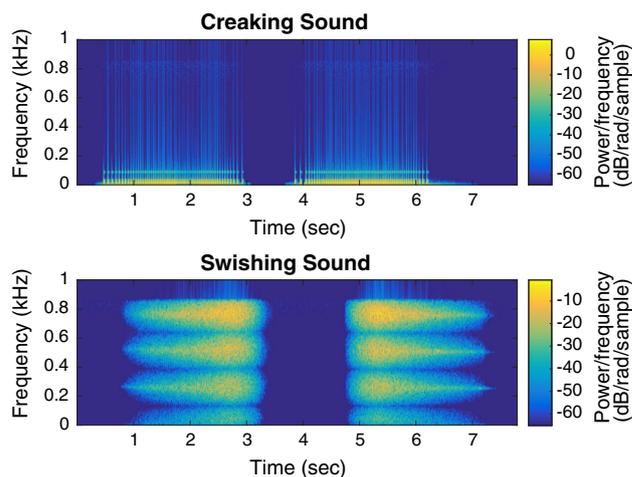


Fig. 3 Excerpts of two sonified throwing gestures using the two sound models. Window size set to 256 sample points, 50 % overlap and minimum threshold – 65 dB

Table 1. For the movement sonification of the interaction with the virtual ball, i.e. the gesture in which the user aims towards the target, we opted for two different complex sound models. The first model (creaking sound) used the friction preset from the Sound Design Toolkit [9] (SDT). SDT consists of a library of ecologically-founded, physics-based, sound synthesis algorithms. The second model (swishing sound) was based on filtering pink noise. Spectrograms of recordings generated using the two sonification models in two throwing gestures are displayed in Fig. 3. For the creaking sound model, the rubbing force variable of the SDT friction preset was scaled to z-position and frequency was scaled to y-position. For the swishing sound model, pink noise was filtered using a band-pass filter (cutoff frequency was set to 18,500 Hz and Q value to 0.7), producing a source signal that was, in turn, filtered using a comb filter with varying feedback (0.2–1) and delay time (5–25 ms), depending on an estimation¹ of velocity and y-position, respectively. For both models, stereo panning was scaled depending on x-position (left or right) and amplitude was scaled logarithmically depending on velocity along the z-axis; sound could only be heard when the ball was moving in the direction towards the goal, not when it was moving side-ways. The bouncing sounds were synthesized in SDT, using two different impact sounds associated with objects of different sizes. For the creaking sound model we used an impact sound reminiscent of a heavy rubber ball impact, for the swishing sound model we used an impact sound reminiscent of a table tennis ball impact. Since many of the events in the application had no intuitive mappings to auditory events (such as grasping the ball, scoring a goal or missing the target), earcons were used to signal these events.

¹ Positional difference between subsequent samples.

3.3.5 Haptic feedback

When the user grabs the virtual ball, an invisible “rubber band” is created between the center of the proxy and the center of the ball. The same function has been used successfully in many earlier studies [35,37,44] and it provides a smooth and realistic feeling of controlling an object, by conveying sensation of both weight and inertia. When the user drops the ball by releasing the button on the PHANToM-device, the rubber band is deleted and the ball bounces away. For the user, the haptic experience of throwing the ball corresponds somewhat to performing a pendulum movement and releasing the object at the end-point. The task of throwing the ball into the goal involves (1) grasping the ball by pressing a button on the PHANToM-stylus when in contact with the ball, (2) lifting the ball by moving the stylus upwards (continuously pressing the button) (3) swinging the ball back and forth and (4) launching the ball by releasing the button, resulting in a flicking upwards ball movement directed towards the goal.²

We used a ball with minimum weight in the nonhaptic conditions, verifying with a dynamometer attached to the PHANToM-stylus that no difference could be observed between the case when the proxy was switched off and when a ball with minimum weight was bound to the proxy. The ball with minimum weight weighed 28 g (this is actually the weight of the haptic stylus), and the ball in the haptic condition weighed 127 g (the weight of the haptic stylus included). In order to make sure that the weight would not influence the behavior of the ball after it had been thrown, the weight was always set to 127 g at the time of release. Thus, when bouncing towards the target, the ball always behaves according to the same rules of physics as in the haptic conditions, regardless of if the ball weighs 28 or 127 g when it is grasped.

3.4 Quantitative data

3.4.1 Performance measures

Two continuous performance variables were measured for each condition: error rate and time to task completion, i.e. task duration. These dependent variables are defined below.

- *Error rate* The total number of failed attempts to score a goal for 15 successful hits, i.e. the number of misses for a total of 15 goals.
- *Task duration* The time, in seconds, required to successfully score 15 times.

² A video of the graphical representation of the gesture is provided as Supplementary Material.

3.4.2 Perceived intuitiveness

One subjective dependent measure was used: perceived intuitiveness. Participants rated the intuitiveness of the interaction in respective condition on a visual analog scale (VAS) of length 10 cm, labeled “not intuitive” versus “intuitive” at the endpoints. These ratings were normalized on participant level based on minimum and maximum values of ratings in all conditions. Normalization on participant level was used since all participants used very different ranges on the VAS.

3.5 Qualitative data

Apart from performance measures and perceptual ratings, qualitative data from interviews and think-aloud was collected. Both the interviews and the participants’ verbalized reasons behind their ratings of intuitiveness were transcribed verbatim. The transcribed material was analysed using content analysis with an emergent themes approach. The material was coded and sorted into themes. The interview results presented in Sect. 5 discuss those themes that relate to the hypothesis of the study.

3.6 Hypotheses

Based on previous findings indicating that adding an additional modality to a visual interface can improve overall performance [4] and that movement sonification can have positive effects on motor learning [12,13,49], we derived the following research hypothesis:

- H1** Task performance, measured as the number of errors performed and the time to complete the task, will be better when auditory feedback is provided (better performance in AVH^1 and AVH^2 compared to VH ; better performance in AV^1 and AV^2 compared to V).

Building on previous findings suggesting that haptic feedback rendered by a point-based haptic device can improve task performance [30,36,45], we also hypothesized that:

- H2** Task performance will be better in the visuohaptic condition (VH) than in the visual-only (nonhaptic) condition (V).

Based on previous research suggesting that semantically congruent cross-modal stimuli can improve behavioral performance [28], we also hypothesized that:

- H3** Task performance will be better when semantically congruent haptic- and auditory feedback is presented (better performance in AVH^1 than in AVH^2 , better performance in AV^2 than in AV^1).

The AVH^1 condition, combining haptic feedback with the creaking movement sonification and a heavy bouncing sound, and the AV^2 condition, a nonhaptic condition with the swishing movement sonification and a light bouncing sound, are considered semantically congruent feedback combinations. Finally, we also hypothesized that:

H4 Perceived intuitiveness of the interaction will be better in conditions in which haptic- and/or auditory feedback is presented (lowest intuitiveness in V , compared to the other conditions).

4 Results

Since the data did not meet the assumption of normality for any of the dependent variables (this was evaluated through visual inspection of histograms and the Shapiro–Wilk test of normality), non-parametric statistical methods for repeated measures were used. We performed a Friedman test to examine main effects, using the `friedman.test` function provided in R [40]. The Friedman test is a multi-sample extension of the sign test. If the Friedman test was significant, we proceeded with post-hoc analysis using multiple sign tests for two-sample paired data to compare specific feedback conditions (using `SIGN.test` from the `BSDA` package in R). Paired sign tests were used since the assumption of symmetry required for the paired Wilcoxon signed rank test was violated. Instead of using the very conservative Bonferroni correction of family-wise error for pairwise comparisons, we report estimates based on confidence intervals for the median of the difference (along with uncorrected p values). This procedure was used to avoid interpreting results solely based on null-hypothesis significance testing, since confidence intervals are much more informative than p values as they indicate the extent of uncertainty, as suggested by Cumming [7].

4.1 Performance measures

4.1.1 Error rate

The total number of participants included in the analysis was 18. One participant was disregarded from the analysis since the headphones accidentally switched off in one of the conditions. Another participant's data was disregarded since the subject stated that he had forgotten to solve the task in some of the conditions and had randomly played around with the haptic device. Mean error rate, when collapsing data across participants and conditions, was 9.046 ($\sigma = 8.913$, $SEM = .858$). Mean and median errors per condition are visualized in Fig. 4. A Friedman ANOVA rendered a χ^2 ($df = 5$, $N = 18$) of 11.364, which was

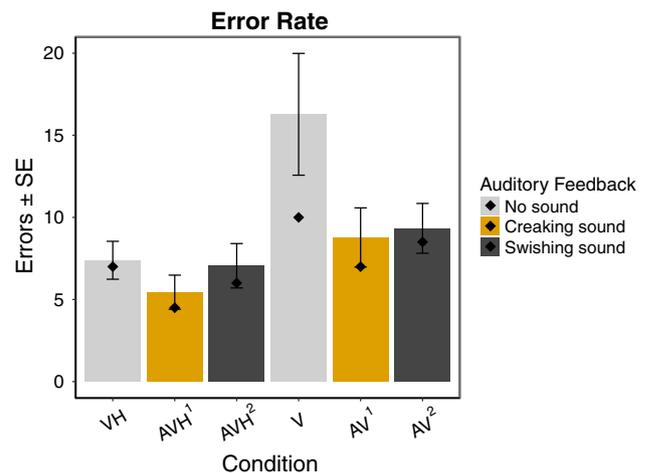


Fig. 4 Mean error rate per condition. Diamonds represent median values for respective condition. V = visuo, AV^1 = audiovisual with a creaking sound, AV^2 = audiovisual with a swishing sound, VH = visuohaptic, AVH^2 = audiovisuohaptic with a creaking sound and AVH^2 = audiovisuohaptic with a swishing sound. SE = Standard error of the mean

significant ($p = .045$), providing evidence against the null hypothesis that the conditions were equivalent with respect to feedback condition. A summary of uncorrected pairwise sign tests, with confidence intervals for the median differences for the pairwise comparisons, is shown in Table 2. Mean ranks are presented in Table 3. Considering both confidence intervals and significance values for each pair-wise comparison, we can conclude that the visual-only condition (V) had significantly higher error rates than the AVH^1 condition (audiovisuohaptic with a creaking sound)³.

4.1.2 Task duration

Data from the same participants as the ones used for the error rate measure was analyzed. Mean task duration, when collapsing data across all participants and conditions, was 226.407s ($\sigma = 108.715$, $SEM = 10.461$). Mean and median task duration for all conditions are visualized in Fig. 5. A Friedman ANOVA rendered a χ^2 ($df = 5$, $N = 18$) of 5.843, which was not significant ($p = .322$). A Pearson product-moment correlation coefficient was computed to assess the relationship between the averaged task duration and error rate for respective participant across all conditions. There was a strong positive correlation between the two variables ($\rho(16) = .898$, $p = 4.380 \cdot 10^{-7}$). In other words, an increase in task duration was correlated with an increase in error rate.

³ The p value was significant also for comparison with condition AV^1 , but the confidence interval passed through zero.

Table 2 Error rate—pairwise comparisons

	<i>VH</i>	<i>AVH</i> ¹	<i>AVH</i> ²	<i>V</i>	<i>AV</i> ¹
<i>AVH</i> ¹	.332 [− 1.000, 3.000]				
<i>AVH</i> ²	1.000 [− 4.708, 3.708]	.302 [− 5.415, .708]			
<i>V</i>	.144 [− 15.491, .708]	.013 [− 17.368, − 1.000]	.238 [− 21.613, 1.000]		
<i>AV</i> ¹	1.000 [− 4.123, 1.708]	.144 [− 7.538, .708]	.804 [− 3.000, 2.415]	.035 [.000, 10.830]	
<i>AV</i> ²	1.000 [− 8.123, 3.708]	.481 [− 8.708, 1.000]	.629 [− 4.000, 3.123]	.096 [− 0.415, 10.123]	.629 [− 6.123, 4.706]

Sign test for two-sample paired data, pairwise comparisons with uncorrected *p* values (first row) and confidence intervals (second row)

V visuo, *AV*¹ audiovisual with a creaking sound, *AV*² audiovisual with a swishing sound, *VH* visuohaptic, *AVH*¹ audiovisuohaptic with a creaking sound and *AVH*² audiovisuohaptic with a swishing sound

Table 3 Error rate ranks

Cond.	<i>VH</i>	<i>AVH</i> ¹	<i>AVH</i> ²	<i>V</i>	<i>AV</i> ¹	<i>AV</i> ²
Err.	3.389	2.611	3.333	4.638	3.611	3.417

Mean error rate ranks for respective condition.

Cond. condition, Err. error rate rank, *V* visuo, *AV*¹ audiovisual with a creaking sound, *AV*² audiovisual with a swishing sound, *VH* visuohaptic, *AVH*¹ audiovisuohaptic with a creaking sound and *AVH*² audiovisuohaptic with a swishing sound

4.2 Perceived intuitiveness

Data from all 20 participants was analyzed. A Friedman ANOVA rendered a χ^2 value ($df = 5, N = 20$) of $F = 17.562$, which was highly significant ($p = .004$), providing evidence against the null hypothesis that the conditions were equivalent with respect to feedback condition. A summary of uncorrected pairwise sign tests, with confidence intervals for the median differences for the pairwise comparisons, is shown in Table 4. Mean ranks are presented in Table 5. Box plots of perceived intuitiveness per condition are shown in Fig. 6. Considering both confidence intervals and significance values for each pair-wise comparison, we can conclude that the visual-only condition (*V*) was significantly less intuitive than most of the other conditions (the only non-significant comparison was observed for condition *AVH*², audiovisuohaptic with a swishing sound). A Pearson product-moment correlation coefficient was computed to assess the relationship between the averaged intuitiveness and error rate of respective participant across all conditions. No strong correlation between the two variables could be observed ($\rho(16) = .092, p = .716$).

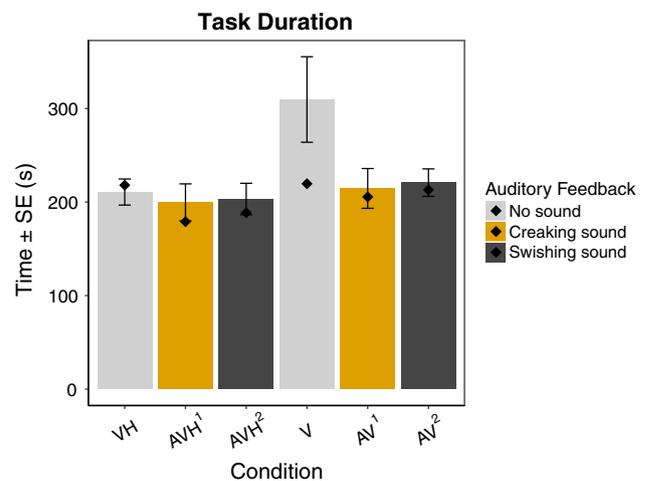


Fig. 5 Mean task duration per condition. *V* = visuo, *AV*¹ = audiovisual with a creaking sound, *AV*² = audiovisual with a swishing sound, *VH* = visuohaptic, *AVH*¹ = audiovisuohaptic with a creaking sound and *AVH*² = audiovisuohaptic with a swishing sound

5 Interview results

5.1 Effects of auditory feedback

Although the quantitative results mainly indicated lower error rates for the *AVH*¹ condition compared to condition *V*, several participants stated that it was easier to control the ball when movement sonification was present. Several participants stated that this was especially true for non-haptic conditions. Some participants commented that the influence of sound was most important in the beginning of each condition, and that they stopped attending to the sound after learning how to perform the particular movement. Several participants described how the movement sonifica-

Table 4 Perceived intuitiveness—pairwise comparisons

	VH	AVH ¹	AVH ²	V	AV ¹
AVH ¹	.167 [–.518, .018]				
AVH ²	.167 [–.282, .088]	1.000 [–.241, .205]			
V	.031 [.010, .708]	.019 [.154, .969]	.064 [.028, .867]		
AV ¹	.115 [–.438, .065]	.815 [–.087, .314]	.263 [–.130, .089]	.004 [–.810, –.102]	
AV ²	1.000 [–.237, .310]	.210 [.000, .389]	.263 [–.099, .469]	.041 [–.679, –.103]	.359 [–.093, .413]

Sign test for two-sample paired data, pairwise comparisons with uncorrected *p* values (first row) and confidence intervals (second row)

V visuo, AV¹ audiovisual with a creaking sound, AV² audiovisual with a swishing sound, VH visuohaptic, AVH¹ audiovisuohaptic with a creaking sound and AVH² audiovisuohaptic with a swishing sound

Table 5 Perceived intuitiveness ranks

Cond.	VH	AVH ¹	AVH ²	V	AV ¹	AV ²
Int.	3.200	4.175	4.175	2.175	3.950	3.325

Mean perceived intuitiveness ranks for respective condition

Cond condition, Int. Intuitiveness, V visuo, AV¹ audiovisual with a creaking sound, AV² audiovisual with a swishing sound, VH visuohaptic, AVH¹ audiovisuohaptic with a creaking sound and AVH² audiovisuohaptic with a swishing sound

tion provided useful information about the throwing gesture, mentioning aspects such as position, depth perception, velocity, timing, or how much the ball was moving overall. The following quote sheds light on the potential benefits of adding movement sonification to this particular task: “(...) it was easier to get an understanding of where the ball was located [in auditory conditions] (...) you could find a rhythm of the sound that you could follow.” This strategy, to find a rhythm in the pendulum movement of the throwing gesture, was mentioned by several participants.

5.2 Effects of haptic feedback

Even if the quantitative results could not confirm that the presence of haptic feedback improved task performance, several participants stated there were benefits of “feeling how the ball was swinging”. One particular benefit of haptic feedback brought up by the participants was the higher degree of control: many participants stated that it was easier to control the ball in the haptic conditions since the movements were somewhat constrained by the resisting force provided by the haptic device. One strategy that was brought up was to use the force as guidance when trying to move the ball in a straight line towards the target.

5.3 Effects of feedback combinations and semantic congruence

Effects produced by presenting certain auditory feedback in combination with, or without, haptic feedback were also

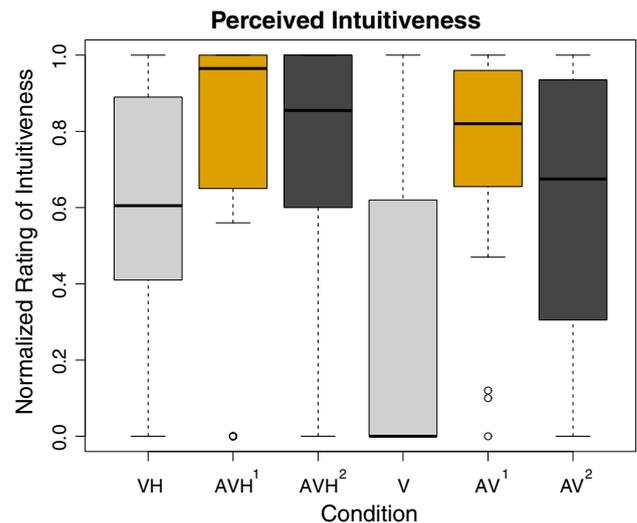


Fig. 6 Perceived intuitiveness per condition. Diamonds represent median values for respective condition. Abbreviations are the same as in Fig. 4. Whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. V visuo, AV¹ audiovisual with a creaking sound, AV² audiovisual with a swishing sound, VH visuohaptic, AVH¹ audiovisuohaptic with a creaking sound and AVH² audiovisuohaptic with a swishing sound

brought up by the interviewees. Some participants mentioned that the auditory feedback gave them an awareness of the weight of the ball, which in turn affected their behavior and the throwing gesture. A few participants discussed the sounds’ ability to compensate for absence of haptic feedback. Some participants described that they felt like they were interacting with a table tennis ball in the auditory conditions AV² and AVH². There were even descriptions of the ball in terms of a “a balloon losing it’s air”. Some participants also described conflicts between what was heard and felt, and that these conflicts led to confusion, forcing them to switch attention between different modalities. One participant stated that the movement sonification tricked her into misjudging the effort required to throw the ball in the semantically incongruent nonhaptic condition with the creaking movement sonification (AV¹).

6 Discussion

Interestingly, error rates in condition V were significantly higher than error rates in condition AVH^1 . These results suggest that the presence of haptic and auditory feedback in terms of a creaking sound helped reducing errors. The visual-only condition V had the highest error rate rank and condition AVH^1 had the lowest error rate rank. Possibly, the more realistic feedback combination in condition AVH^1 might have led to improvements in terms of reduced error rates. However, we could not conclude that auditory feedback in general helped improving task performance (hypothesis H1), as the effect was only present for the AVH^1 condition. Nevertheless, the interview results suggested that the sound design and mapping decisions were successful; users could get an understanding of the ball's position and speed, simply by listening to the movement sonification. The sound appeared to help some of the participants in finding a rhythm supporting the movement in the repetitive throwing task. This is supported by previous findings [42], suggesting that it is easier to synchronize to auditory rhythms than visual ones when performing a rhythmic movement. The creaking sound model, which uses a somewhat cartoonified physical model of friction, is a perhaps more natural sound than the synthetic swishing sound model. Previous studies have shown that there is a difference in processing times for natural versus artificial sounds [55]. Natural sounds with causal mappings have also been found to be rated as more pleasant than arbitrary mappings [56].

As opposed to previous findings [30,36,45] we could not conclude that task performance was significantly improved when haptic feedback alone was added to the visual interface, thus rejecting hypothesis H2. However, when it came to the perception of control and precision, the haptic feedback seems to have made a difference, as clearly indicated in the interviews.

Regarding hypothesis H3, we hypothesized that task performance would be better in semantically congruent conditions than in semantically incongruent conditions. The fact that the creaking sound model was ecologically congruent with the haptic feedback may have affected performance since error rates were significantly lower in the AVH^1 condition than in the visual only condition V . However, we could not conclude that semantically congruent feedback in general improved task performance compared to semantically incongruent feedback. No significant difference could be observed between incongruent- and congruent stimuli for the intuitiveness rating either. Nevertheless, results from the interviews did shed some light on the effect of semantic incongruence; several participants stated that confusion caused by semantically incongruent feedback forced them to switch attention between different modalities.

Finally, we can conclude that having access to more senses appeared to add to perceived intuitiveness; the V condition was significantly different from all other conditions, with the exception of condition AVH^2 . In other words, most conditions that involved haptic and/or auditory feedback were rated as more intuitive than the visual-only condition. The conditions that were rated as most intuitive (in terms of mean ranks) were the ones including both haptic and auditory feedback, i.e. AVH^1 and AVH^2 .

6.1 Methodological concerns

In general, the inter-subject variability observed in this study motivates follow-up studies involving a larger number of participants. Subjects used different strategies (i.e. throwing gestures) when solving the task; perhaps an even more controlled experimental setting would have facilitated interpretation of the results. In any case, a larger data set would be advantageous for this type of experiment. Moreover, it has previously been found that learning effects can be observed in both coordination mode and variability of various parameters of limb motion [32]. To a certain extent, one may assume that learning effects may have influenced the performance measures described in this paper. Nevertheless, the order of conditions was randomized and even so, a significant effect of feedback type could be observed for the error rate measure.

7 Conclusions

In this paper we present a study on the effect of auditory- and haptic feedback in a virtual throwing task performed with a point-based haptic device. We could not conclude that the addition of haptic feedback in general significantly improved task performance. Moreover, the addition of auditory feedback alone could not be shown to significantly improve task performance. However, significantly lower error rates were observed when a sonification model based on a physical model of friction was used *together with haptic feedback*, compared to a visual-only condition. Despite these results, we could not conclude that conditions with semantically congruent auditory- and haptic feedback in general resulted in significantly improved performance. Lastly, we hypothesized that perceived intuitiveness of the interaction would be better in conditions in which haptic- and/or auditory feedback was present. All conditions apart from one were found to be significantly more intuitive than the visual-only condition. The conditions rated as most intuitive were the ones involving both haptic- and auditory feedback.

The main contribution of this study lies in its finding that multimodal feedback can not only increase perceived

intuitiveness of an interactive interface, but also that certain combinations of haptic- and auditory feedback, e.g. movement sonification simulating the sound of friction presented together with haptic feedback simulating inertia, can contribute with performance-enhancing properties. These findings highlight the importance of carefully designing feedback combinations for interactive applications.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Avanzini F, Rocchesso D, Serafin S (2004) Friction sounds for sensory substitution. In: Proceedings of the 10th meeting of the international conference on auditory display (ICAD 2004), Sydney, pp 1–8
- Bencina R, Wilde D, Langley S (2008) Gesture \approx sound experiments: process and mappings. In: Proceedings of the 2008 conference on new interfaces for musical expression (NIME08), pp 197–202
- Brewster SA, Wright PC, Edwards AD (1993) An evaluation of earcons for use in auditory human–computer interfaces. In: Proceedings of the INTERACT’93 and CHI’93 conference on human factors in computing systems, pp 222–227
- Burke JL, Prewett MS, Gray AA, Yang L, Stilson FR, Covert MD, Elliot LR, Redden E (2006) Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. In: Proceedings of the 8th international conference on multimodal interfaces, pp 108–117
- Conti F, Barbagli F, Balaniuk R, Halg M, Lu C, Morris D, Sentis L, Warren J, Khatib O, Salisbury K (2003) The chai libraries. In: Proceedings of Eurohaptics 2003, Dublin, pp 496–500
- Crommentuijn K, Winberg F (2006) Designing auditory displays to facilitate object localization in virtual haptic 3d environments. In: Proceedings of the 8th international ACM SIGACCESS conference on computers and accessibility. ACM, pp 255–256
- Cumming G (2014) The new statistics—why and how. *Psychol Sci* 25:7–29
- Cycling ’74: Max. <https://cycling74.com/>. Accessed 25 Apr 2018
- Delle Monache S, Polotti P, Rocchesso D (2010) A toolkit for explorations in sonic interaction design. In: Proceedings of the 5th audio mostly conference: a conference on interaction with sound, p. 1
- Driver J, Spence C (1998) Crossmodal attention. *Curr Opin Neurobiol* 8(2):245–253
- Dubus G, Bresin R (2013) A systematic review of mapping strategies for the sonification of physical quantities. *PloS one* 8(12):e82491
- Dyer J, Rodger M, Stapleton P (2016) Transposing musical skill: sonification of movement as concurrent augmented feedback enhances learning in a bimanual task. *Psychol Res* 81(4):1–13
- Effenberg AO, Fehse U, Schmitz G, Krueger B, Mechling H (2016) Movement sonification: effects on motor learning beyond rhythmic adjustments. *Front Neurosci* 10:219
- Ekman I, Rinott M (2010) Using vocal sketching for designing sonic interactions. In: Proceedings of the 8th ACM conference on designing interactive systems (DIS), pp 123–131
- Fernström M, McNamara C (2005) After direct manipulation—direct sonification. *ACM Trans Appl Percept (TAP)* 2(4):495–499
- Gaver WW (1993) Synthesizing auditory icons. In: Proceedings of the INTERACT’93 and CHI’93 conference on human factors in computing systems. ACM, pp 228–235
- Gescheider GA, Barton WG, Bruce MR, Goldberg JH, Greenspan MJ (1969) Effects of simultaneous auditory stimulation on the detection of tactile stimuli. *J Exp Psychol* 81(1):120
- Grabowski NA, Barner KE (1998) Data visualisation methods for the blind using force feedback and sonification. In: Proceedings of Spie—the international society for optical engineering, vol 3524. SPIE-Int. Soc. Opt. Eng., pp 131–139
- Guest S, Catmur C, Lloyd D, Spence C (2002) Audiotactile interactions in roughness perception. *Exp Brain Res* 146(2):161–171
- Hermann T Sonification: a definition. <http://sonification.de/son/definition>. Accessed 25 Apr 2018
- Hermann T, Hunt A, Neuhoff JG (2011) The sonification handbook. Logos Verlag, Berlin
- Hötting K, Röder B (2004) Hearing cheats touch, but less in congenitally blind than in sighted individuals. *Psychol Sci* 15(1):60–64
- Jousmäki V, Hari R (1998) Parchment-skin illusion: sound-biased touch. *Curr Biol* 8(6):R190–R191
- Kaklanis N, Votis K, Tzavaras D (2013) Open touch/sound maps: a system to convey street data through haptic and auditory feedback. *Comput Geosci* 57:59–67
- Kitagawa N, Spence C (2006) Audiotactile multisensory interactions in human information processing. *Jpn Psychol Res* 48(3):158–173
- Koseleff P (1957) Studies in the perception of heaviness. i. 1.2: some relevant facts concerning the size-weight-effect. *Acta Psychologica* 13:242–252
- Kramer G, Walker B, Bonebright T, Cook P, Flowers JH, Miner N, Neuhoff J (2010) Sonification report: status of the field and research agenda. Tech. rep
- Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT (2004) Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res* 158(4):405–414
- Lederman SJ, Klatzky RL (2009) Haptic perception: a tutorial. *Atten Percept Psychophys* 71(7):1439–1459
- Lee S, Kim GJ (2008) Effects of haptic feedback, stereoscopy, and image resolution on performance and presence in remote navigation. *Int J Hum Comput Stud* 66(10):701–717
- Magnusson C, Rasmussen-Gröhn K (2008) A pilot study on audio induced pseudo-haptics. In: HAID’08, pp 6–7
- McDonald P, Van Emmerik R, Newell K (1989) The effects of practice on limb kinematics in a throwing task. *J Mot Behav* 21(3):245–264
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748
- Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb Cortex* 14(4):452–465
- Moll J, Sallnäs Pysander EL (2013) A haptic tool for group work on geometrical concepts engaging blind and sighted pupils. *ACM Trans Access Comput (TACCESS)* 4(4):14
- Nam CS, Shu J, Chung D (2008) The roles of sensory modalities in collaborative virtual environments (CVES). *Comput Hum Behav* 24(4):1404–1417
- Moll J, Huang Y, Sallnäs EL (2010) Audio makes a difference in haptic collaborative virtual environments. *Interact Comput* 22(6):544–555

38. Moll J, Pysander ELS, Eklundh KS, Hellström SO (2014) The effects of audio and haptic feedback on collaborative scanning and placing. *Interact Comput* 26(3):177–195
39. Frid E, Bresin R, Moll J, Pysander ELS (2017) An exploratory study on the effect of auditory feedback on gaze behavior in a virtual throwing task with and without haptic feedback. In: *Sound and Music Computing Conference 2017 (SMC2017)*
40. R Development Core Team (2008) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. Accessed 25 Apr 2018 ISBN 3-900051-07-0
41. Rassmus-Gröhn CMK (2005) Audio haptic tools for navigation in non visual environments. In: *ENACTIVE 2005, the 2nd international conference on enactive interfaces, Genoa-Italy, 2005*, pp 17–18
42. Repp BH, Penel A (2004) Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychol Res* 68(4):252–270
43. Ro T, Hsu J, Yasar NE, Elmore LC, Beauchamp MS (2009) Sound enhances touch perception. *Exp Brain Res* 195(1):135–143
44. Sallnäs EL (2004) The effect of modality on social presence, presence and performance in collaborative virtual environments. PhD Thesis. KTH Royal Institute of Technology, Stockholm
45. Sallnäs EL, Rassmus-Gröhn K, Sjöström C (2000) Supporting presence in collaborative environments by haptic force feedback. *ACM Trans Comput Hum Interact (TOCHI)* 7(4):461–476
46. Schürmann M, Caetano G, Hlushchuk Y, Jousmäki V, Hari R (2006) Touch activates human auditory cortex. *NeuroImage* 30(4):1325–1331
47. Sherrick CE (1976) The antagonisms of hearing and touch. In: Hirsh SK, Eldredge DH, Hirsh IJ, Silverman SR (eds) *Hearing and Davis: essays honoring Hallowell Davis*, St. Washington University Press, Louis, pp 149–158
48. Shimojo S, Shams L (2001) Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol* 11(4):505–509
49. Sigrist R, Rauter G, Marchal-Crespo L, Riener R, Wolf P (2015) Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning. *Exp Brain Res* 233(3):909–925
50. Sigrist R, Rauter G, Riener R, Wolf P (2013) Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon Bull Rev* 20(1):21–53
51. Someren Mv, Barnard YF, Sandberg JA et al (1994) *The think aloud method: a practical approach to modelling cognitive processes*. Academic Press, Cambridge
52. Spence C (2011) Crossmodal correspondences: a tutorial review. *Atten Percept Psychophys* 73(4):971–995
53. Srinivasan MA, Basdogan C (1997) Haptics in virtual environments: taxonomy, research status, and challenges. *Comput Gr* 21(4):393–404
54. Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 9(4):255–266
55. Suied C, Susini P, McAdams S, Patterson RD (2010) Why are natural sounds detected faster than pips? *J Acoust Soc Am* 127(3):EL105–EL110
56. Susini P, Misdariis N, Lemaitre G, Houix O (2012) Naturalness influences the perceived usability and pleasantness of an interfaces sonic feedback. *J Multimodal User Interfaces* 5(3–4):175–186
57. Wright M, Freed A et al (1997) Open Sound Control: A new protocol for communicating with sound synthesizers. In: *Proceedings of the 1997 international computer music conference (ICMC1997)*, vol 2013, p 10
58. Yu W, Brewster S (2003) Evaluation of multimodal graphs for blind people. *Univ Access Inf Soc* 2(2):105–124

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.