AN EXPLORATORY STUDY ON THE EFFECT OF AUDITORY FEEDBACK ON GAZE BEHAVIOR IN A VIRTUAL THROWING TASK WITH AND WITHOUT HAPTIC FEEDBACK

Emma Frid, Roberto Bresin, Eva-Lotta Sallnäs Pysander
KTH Royal Institute of Technology
{emmafrid, roberto, evalotta}@kth.se

Jonas Moll
Uppsala University
jonas.moll@it.uu.se

ABSTRACT

This paper presents findings from an exploratory study on the effect of auditory feedback on gaze behavior. A total of 20 participants took part in an experiment where the task was to throw a virtual ball into a goal in different conditions: visual only, audiovisual, visuohaptic and audiovisuohaptic. Two different sound models were compared in the audio conditions. Analysis of eye tracking metrics indicated large inter-subject variability; difference between subjects was greater than difference between feedback conditions. No significant effect of condition could be observed, but clusters of similar behaviors were identified. Some of the participants' gaze behaviors appeared to have been affected by the presence of auditory feedback, but the effect of sound model was not consistent across subjects. We discuss individual behaviors and illustrate gaze behavior through sonification of gaze trajectories. Findings from this study raise intriguing questions that motivate future large-scale studies on the effect of auditory feedback on gaze behavior.

1. INTRODUCTION

Sound has been found to influence visual perception [1, 2] and affect attention [3]. It has been suggested that sound effects can be used to increase selective attention and inferential recognition among children [4] and that sound can facilitate visual learning [5]. The above-mentioned findings highlight the potential benefits of integrating auditory feedback in tasks requiring a certain level of visual focus. Despite the possible benefits of adding sound to such tasks, few studies have focused on the effect of auditory feedback on gaze behavior when interacting with a multimodal interface. The body of research related to this topic mainly involves studies on the effect of auditory feedback when watching videos. It has been found that different kinds of sounds influence gaze differently [6, 7]. In one study on soundtracks in videos, it was concluded that sound may influence eye position, fixation duration and saccade amplitude, but that the effect of sound is not necessarily constant over time [8]. In another study [9], it was found that the influence of audio depends on the consistency between the visual and audio signals: if the salient objects from the visual perspective are not consistent with the salient objects from the audio perspective, audio will influence visual attention. Otherwise, there is little influence on visual attention.

There has been some research on sonification of gaze behavior, e.g. [10], focusing on how sonification can be used to guide visual attention and [11], focusing on sonification for eye movement control. In [11], the effects of sonification on performance and learning were evaluated using an interface that sonified eye movements. Very heterogeneous behaviors were observed among subjects and the effect of movement sonification was therefore evaluated on an individual level. A similar approach, based on analysis on subject level, was used in [10].

In the studies referenced above, eye tracking technology was used to record gaze behavior. Eye tracking is a technique in which an individual's eye movements are measured in order to detect where a person is looking at a specific time [12]. Common eye tracking metrics include measures of fixations, such as e.g. total fixation duration and number of fixations. Fixations are moments when the eyes are relatively stationary due to information processing. Typically, fixations are correlated with attention, but interpretation of fixation metrics may differ depending on the specific nature of the study. Based on top-down cognitive theory, longer fixation durations in a specific region of an interface would reflect a participant's difficulty in interpretation for this particular area. Fixation duration is believed to relate to the cognitive effort expended [13, 14]. However, long fixation duration could also indicate that a specific area is more engaging in some way [15]. A high number of fixations in a particular area could also imply that the region is more noticeable or important than other regions [12].

In this paper we present main findings from a study on the distribution of visual attention in a multimodal single-user application comparing different modalities. We investigated the effect of auditory feedback on visual attention and gaze behavior. An experiment was performed with 20 participants. The participants were asked to throw a virtual ball into a goal in different feedback conditions. Eye tracking methodologies were used to investigate where the visual sensory information was acquired when performing the task. We evaluated the effect of auditory feedback on eye tracking metrics and gaze behavior, both when haptic force feedback was provided and when no haptic force feedback feedback was provided. The results presented in this paper are a subset of the results from the performed study. See [16] for a discussion about several other measures.

2. METHOD

One of the main aims of the study was to investigate if, and how, auditory feedback alters gaze behavior and distribution of visual attention in a nonhaptic versus haptic task. More dimensions are covered in [16]. The user was asked to throw a virtual ball into a goal, see Fig. 1. A throwing gesture was chosen since this is an intuitive and simple movement task that requires attention to shift between an object and a target. This gesture 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.

2.1 Hypothesis

We hypothesized that continuous auditory feedback effectively represents temporal information in such a manner that a shift in attention is enabled in a virtual (visual) throwing task. More specifically, our hypothesis was that *the use of interactive movement sonification will enable a user to shift visual focus from an actual interaction point (the location of the cursor) to a more distant target point in a virtual 3D environment.* Sonification can provide meaningful information about a movement. In conditions involving sonification, the user will not be required to focus as much on the ball, since information about the performed movement is communicated directly through sound. This will enable visual focus to shift from the ball, to the goal. Moreover, we hypothesized that *gaze behavior will be differently affected when haptic force feedback (adding a sense of weight and inertia) is simultaneously presented with movement sonification*.

2.2 Participants

A total of 20 participants took part in the experiment. All participants were first year students at KTH Royal Institute of Technology (11 M, 9 F; age=19-25 years, mean = 20.4 years). None of the participants reported having any hearing deficiencies or reduced touch sensitivity. All participants reported that they had normal or corrected-to-normal vision. Six of the participants normally wore glasses; two of them wore their glasses during the test sessions and two of them used contact lenses. None of the participants reported that they had any previous experience of using force feedback devices. All participants signed a consent form prior to taking part in the experiment. Six participants were disregarded from the analysis since the level of usable gaze data was below 80% [1]. Yet another participant was disregarded since the headphones accidentally switched off during the experiment. The final sample size for analysis was n=13 (8 M, 5 F; age=19-24 years, mean = 20.8 years).

[1] 100% in this context would imply that gaze data for both eyes were found throughout the entire recording.

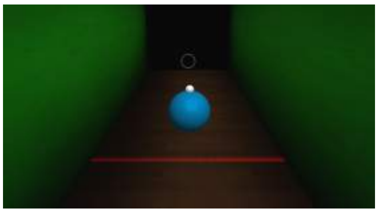

Figure 1. Screenshot of the graphical interface.

2.3 Apparatus

2.3.1 Technical Setup

The technical setup can be seen in Fig. 2. One desktop computer was used to run the haptic interface and the eye tracking application; one laptop computer was used for sound generation. The 3D interface was designed to allow the user to grasp a ball and to throw it into a circular goal area. A SensAbleTM Phantom® Desktop haptic device [2] was used to provide haptic feedback. This haptic device has a pen-like stylus that is attached to a robotic arm. The stylus can be used to haptically explore objects in virtual 3D environments and to provide force feedback. Weight and inertia can also be simulated. In our experiment, a button on the stylus was used to activate the function of grasping an object, enabling the user to lift and move the virtual ball. The 3D application, shown in Fig. 1, was based on the haptic software library Chai3D [17] and developed in C++.

Eye tracking data was captured at a sampling rate of 60 Hz using a commercial X2-60 eye-tracker from Tobii Technology [3] connected to an external processing unit. The software Tobii Studio was used for calibration and analysis of eye tracking data. Two interactive sonification models were implemented using Max/MSP [4]. Communication between Max/MSP and the 3D application was made possible via Open Sound Control (OSC) [18]. A pair of Sennheiser RS 220 headphones was used to provide auditory feedback during the experiment.

2.3.2 Sonification

For conditions involving auditory feedback, the movements of the virtual ball were continuously sonified. Two different sound models were used for sonification of the throwing gesture, i.e. when aiming towards the goal. The first sound model (from now on referred to as S1) was based on a physical model of friction, readily available from the Sound Design Toolkit [5] [19] (SDT). The second sound model (from now on referred to as S2) was obtained by filtering pink noise using the Max object `[biquad]` with a low-pass filter setting. We opted for these particular sound

[2] http://www.dentsable.com/haptic-phantom-desktop.htm
[3] www.tobiipro.com
[4] https://cycling74.com/
[5] http://soundobject.org/SDT/

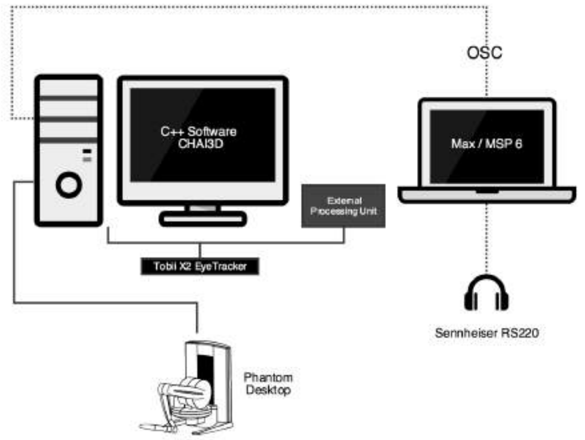

Figure 2. Technical setup.

models since they were hypothesized to be very perceptually different; S1 produced creaking sounds whereas S2 produced swishing sounds [6]. For sound model S1, x-,y- and z-positions were mapped to panning, frequency, and rubbing force, respectively. For S2, x-,y- and z-positions were mapped to panning, frequency and comb-filter characteristics, respectively. Velocity was mapped to amplitude for both sound models.

Different impact sounds from the Sound Design Toolkit (SDT) were used for the two sound models, in order to simulate different weights of the ball. These sounds were included to enable evaluation of the overall effect of auditory feedback (i.e. not only the effect of movement sonification). Parameters were chosen in such a manner that a sound reminiscent of a heavy rubber ball was used for bouncing events in S1, whereas the bouncing effect was tuned to remind more of the impact sound created by a very light table tennis ball (a ping pong ball) for S2. A set of interaction sounds were also included in the application in order to communicate occurrence of the following events: scoring a goal (MIDI sequence, increasing pitch), missing the goal (MIDI sequence, decreasing pitch), hitting a wall (SDT impact sound model used to create a sound reminding of a dissonant bell) and successfully grasping the ball (filtered noise with increasing frequency, producing a sweeping sound that finishes with a short "click").

2.4 Procedure

We opted for a within-participant design in which repeated measures were done on each participant. Participants performed the throwing task in six different feedback conditions. The following conditions were included in the experiment: visual only, i.e. visuo (V), audiovisual with a creaking sound (AV1), audiovisual with a swishing sound (AV2), visuohaptic (VH), audiovisuohaptic with a creaking sound (AVH1) and audiovisuohaptic with a swishing sound (AVH2). A randomized order of the six conditions was presented to each participant.

The experiment consisted of the following sub-parts: (1)

[6] Sound examples can be found at: `https://kth.box.com/s/2bhx2n2wrvq3na582h2qasfpqrxu0zod`

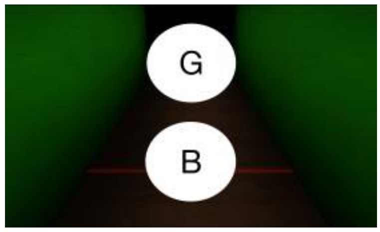

Figure 3. Definition of areas of interest. G stands for the goal AOI and B for the ball AOI.

general introduction and calibration of the system, (2) three condition sessions, (3) a 10 minute break, (4) three condition sessions and (5) concluding interview. Every condition was preceded by a practice trial (minimum 2 and maximum 4 minutes) in which participants got the chance to familiarize themselves with solving the task under the current feedback condition. The practice trials were included in order to reduce the risk of potential learning effects. The task was to throw the ball into the goal until 15 successful hits had been reached. Each experiment lasted about 1.5 hours. The concluding interview consisted of a discussion about aiming strategies and the visual focus on different areas of the screen.

2.5 Analysis

A common method in analysis of eye tracking data is to initially define areas that are of particular interest for the research hypothesis; Areas of Interest (AOIs). In this experiment, three AOIs were defined: one for the goal area (G), one for the ball area (B), and one for the entire screen. The G and B AOIs are displayed in Fig. 3. These elliptical AOIs were defined after visual inspection of heat maps and automatic generation of clusters of fixations, using the Tobii Studio software. The two AOIs had the same size.

The following metrics were computed for the AOIs: *total fixation duration* and *total fixation count*. All calculations were done on 15 throwing gestures. This was done in order to obtain a measure that was comparable despite the fact that task duration and performance varied substantially between participants (the total number of throwing gestures varied from 15 to 66). No consistent patterns in terms of correlations between error rates (number of unsuccessful attempts to score a goal for a total of 15 achieved hits) and eye tracking metrics could be observed [7]. For a more detailed discussion on task performance, see [16].

In the following section we provide descriptive statistics for the two eye tracking metrics, but without statistically significant conclusions when averaging the subjects' values [8]. Participants exhibited large inter-individual variability for both the total fixation duration and the total fixation count metric. Some general trends could however be observed, and these are explored through cluster analysis of total fixation duration, see Sec. 3.3. Finally, the effect of auditory feedback on individual gaze behaviors are dis-

[7] Analysis was done separately for respective condition.

[8] Continuous data was analyzed using a Friedman rank test for repeated measures and count data was analyzed using a mixed effect regression model with a Poisson error distribution.

cussed and explored through sonifications of gaze trajectories, see Sec. 3.4.

3. RESULTS

In order to test the hypotheses, an index of the eye tracking metric (fixation duration or fixation count) for the ball AOI divided by the metric for both the goal AOI and the ball AOI was defined, according to Eq. (1):

$$I = \frac{FixBall}{FixGoal + FixBall} \tag{1}$$

This index simplifies the interpretation of whether the participant was focusing mostly on the ball or the goal AOI without having to take data from both AOIs into consideration simultaneously.

3.1 Total Fixation Duration

Total fixation duration measures the total duration of all fixations within a defined area of interest. In the current study this measure represents the total time that a participant fixated on the goal, ball or the entire screen, respectively. We calculated total fixation duration on the entire screen for the first 15 throwing gestures. Descriptive statistics for this measure are presented in Tab. 1. We can observe that the highest median value was found for condition VH. Lowest median value was observed for condition AV2.

Filtering recordings to include gaze data only from when the participants were grasping the virtual ball, the total fixation duration index (according to Eq. 1) was computed for each participant and condition. As opposed to the measure presented in Tab. 1, this measure specifically evaluates the effect of movement sonification (not the effect of other auditory events caused by e.g. scoring a goal) on eye tracking metrics, since such data has been filtered out. A total fixation duration index higher than 0.5 indicates longer fixation on the ball AOI than on the goal AOI (when grasping the ball). Descriptive statistics for the measure is presented in Tab. 2. Highest median value was found for the haptic condition AVH2. Lowest median value was observed for the nonhaptic condition AV2. Total fixation duration index per participant and condition is visualized in Fig. 4.

Condition	Median	IQR
V	137.39	21.72
AV1	137.96	24.06
AV2	125.92	25.44
VH	142.73	44.49
AVH1	132.39	19.39
AVH2	131.92	31.97

Table 1. Total fixation duration on screen (seconds) for full duration of first 15 attempts to throw the ball into the goal. IQR=inter quartile range. Red cells : highest median value. Green cell: lowest median value.

Condition	Median	IQR
V	0.53	0.77
AV1	0.43	0.61
AV2	0.38	0.70
VH	0.49	0.69
AVH1	0.46	0.60
AVH2	0.66	0.70

Table 2. Total fixation duration index for first 15 throwing gestures, filtered to include data only when grasping the ball. IQR=inter quartile range. Red cell: highest median value. Green cell: lowest median value.

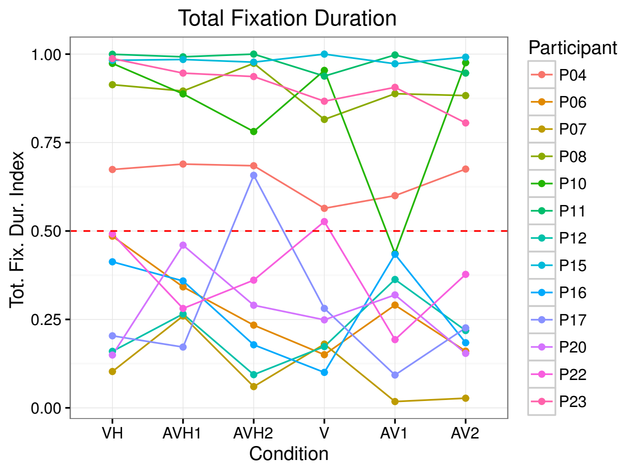

Figure 4. Total fixation duration index values for first 15 throwing gestures, filtered to include data only when grasping the ball. Participants above the dashed red line focused longer time on the ball than on the goal AOI.

3.2 Total Fixation Count

Total fixation count measures the total number of fixations within an area of interest. We calculated total fixation count on the entire screen for the first 15 throwing gestures. In the descriptive statistics table Tab. 3, we can observe that the highest median value was found for condition V. Lowest median value was observed for condition AV1.

Filtering data to include gaze data only from when the ball was grasped, total fixation count index (as described in Eq. 1) was computed for each participant and condition. A total fixation count index higher than 0.5 indicates more fixations on the ball AOI than on the goal AOI (when grasping the ball). Descriptive statistics for the measure is presented in Tab. 4. Highest median value was found for the haptic condition AVH2. Lowest medians were observed for the nonhaptic conditions AV1 and AV2.

3.3 Cluster Analysis - Total Fixation Duration

In order to explore similar behavioral patterns among participants, a K-means clustering analysis was performed. Appropriate number of clusters was decided after visually inspecting patterns in Fig. 4. The number of clusters was set to 5, and the initial cluster centers were evaluated based on the data, specifying random starting assignments to 20 and selecting the solution with the lowest within cluster

Condition	Median	IQR
V	211	88
AV1	183	80
AV2	200	74
VH	199	127
AVH1	188	102
AVH2	196	74

Table 3. Total fixation count on screen for full duration of first 15 attempts to throw the ball into the goal. IQR=inter quartile range. Red cell : highest median value. Green cell: lowest median value.

Condition	Median	IQR
V	0.59	0.31
AV1	0.53	0.23
AV2	0.53	0.31
VH	0.64	0.55
AVH1	0.64	0.33
AVH2	0.65	0.40

Table 4. Total fixation count index for first 15 throwing gestures. IQR=inter quartile range. Red cell: highest median value. Green cell: lowest median value.

variation. The total variance described by clusters was found to be 94.2 %. The cluster centers are given in Tab. 5. We could observe differences between conditions in the different cluster groups, but the effect of sound model was not consistent for all clusters. There were three clusters where several participants showed somewhat similar behaviors (2,3,5) and two single-participant clusters (1,3).

Overall, participants in cluster 3 had longer fixation durations on the goal AOI than on the ball AOI. Participants in cluster 5 showed the opposite behavior. As seen in Fig. 5, we can observe that 7 participants (belonging to clusters 2 and 3) generally had longer fixation durations on the goal AOI than on the ball AOI. A total of 6 participants (cluster 1,4 and 5) had longer fixation durations on the ball AOI.

Overall performance, measured as the sum of the error rates [9] for all conditions, was computed for each participant. In order to explore if there was a difference in performance between clusters, mean overall performance per cluster was computed. Mean error rate was 35.0 errors for cluster 1, 37.5 (median 37.5) for cluster 2, 32.0 (median 26.0) for cluster 3, 52.0 for cluster 4 and 95.0 (median 96.0) for cluster 5.

Cluster	VH	AVH1	AVH2	V	AV1	AV2
1	0.97	0.89	0.78	0.95	0.44	0.96
2	0.35	0.23	0.51	0.40	0.14	0.30
3	0.26	0.34	0.17	0.17	0.28	0.15
4	0.67	0.68	0.68	0.56	0.60	0.68
5	0.97	0.95	0.97	0.90	0.94	0.91

Table 5. Cluster centers per condition.

[9] For each condition, error rate was defined as the number of unsuccessful attempts to score a goal for a total of 15 achieved hits.

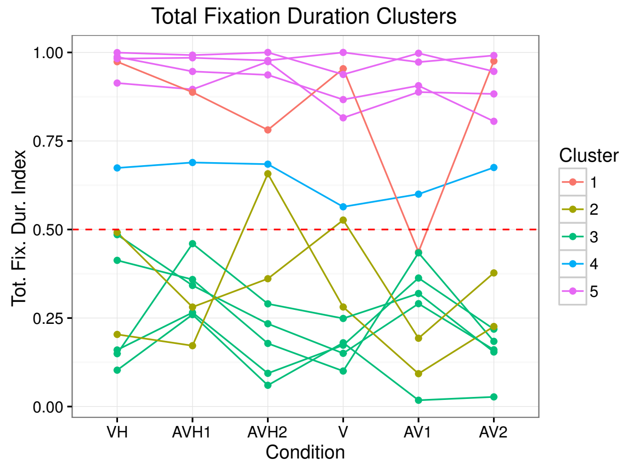

Figure 5. Cluster membership.

3.4 Individual Behaviors

Since very heterogeneous behaviors were observed among the participants we found it relevant to assess the effect of feedback condition for each participant individually. In this section, the gaze behavior of one participant from each cluster is discussed in detail. The metrics that are discussed in the following section come from the segmented data and only takes figures from when the ball was grasped into account. Total fixation duration- and total fixation count data for these individuals is presented in Tab. 6. Sonifications of these participants' behaviors when grasping the ball are also presented [10] in an attempt to explore temporal aspects and patterns in the gaze data. In these simple sonifications (based on sine waves with an applied reverb), the vertical coordinate of the averaged left- and right eye gaze points on the screen was mapped to pitch (increasing pitch for increasing vertical position) and the averaged horizontal coordinate was mapped to panning. Each throwing gesture was separated by a one second break. Interestingly, different behavioral patterns in cluster 3 and 5 are easily identified, for example by listening to the sonifications produced by gaze trajectories of participant 12 and 15.

3.4.1 *Cluster 1 - Participant 10*

This participant was classified separately in a solo cluster. The total error rate for all conditions for this participant was 35.0 errors. The participant performed the task in the following condition order: AV1, AVH2, AV2,

	VH	AVH1	AVH2	V	AV1	AV2
P04	21.59/28	31.4/48	26.77/45	15.98/30	18.02/30	19.82/29
P10	14.51/25	19.43/25	9.96/20	19.72/29	7.24/20	14.04/18
P12	3.05/13	5.04/28	2.10/14	3.13/10	7.44/28	5.43/19
P15	45.99/40	39.00/39	35.72/41	46.83/29	43.16/36	34.57/28
P22	16.26/57	5.81/25	6.65/31	29.67/77	4.29/22	12.25/35

Table 6. Total fixation duration (red) and total fixation count (blue) on ball AOI for participant P04, P10, P12, P15 and P22.

[10] Recordings of the sonified data are available at `https://kth.box.com/s/qgknhxjz7ul4sf3fi0oyqt7aipmun3v2`.

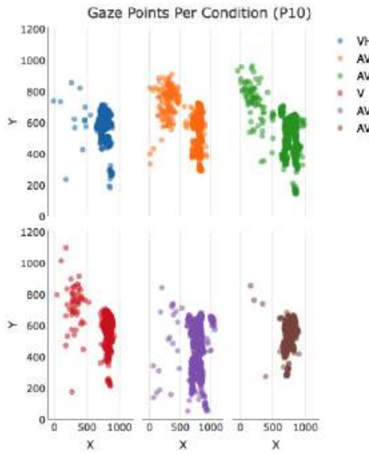

Figure 6. Gaze behavior of participant 10. Points correspond to the coordinates of the averaged left- and right eye gaze position on the screen.

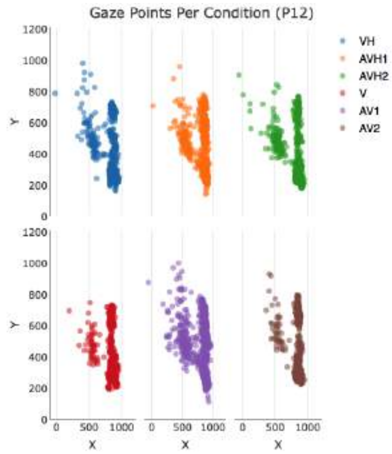

Figure 7. Gaze behavior of participant 12. Points correspond to the coordinates of the averaged left- and right eye gaze position on the screen.

AVH1, V and VH. The participant focused mostly on the ball, in most conditions (all apart from AV1). The coordinates of the averaged left- and right eye gaze positions are displayed in Fig. 6. The plot suggests that there might be a difference between conditions (the points are more or less scattered in different conditions). For the ball AOI, this participant had total fixation durations in the range between 7.24 and 19.72 seconds (lowest for condition AV1 and highest for condition V) and total fixation counts in the range between 18 and 29 counts (lowest for AV2 and highest for condition V).

3.4.2 Cluster 2 - Participant 22

Generally, participants belonging to cluster 2 focused longer on the ball AOI for the swishing sound model than for the creaking sound model. The total error rate for all conditions for participants in this cluster was 37.5 (median 37.5). Participant 22 performed the experiment in the following condition order: V, AV2, VH, AV1, AVH1 and AVH2. For most of the conditions, this particular participant focused longer time on the goal AOI than on the ball AOI. For the ball AOI, participant 22 had total fixation durations in the range between 4.29 and 29.67 seconds (lowest for AV1 and highest for V) and total fixation counts in the range between 22 and 77 counts (lowest for AV1 and highest for condition V). By listening to the sonified gaze trajectories, one can observe that condition V stands out from the other conditions; this condition had overall longer gaze durations, compared to other nonhaptic conditions. Interestingly, we can observe from the metrics in Tab. 6 that this participant focused more on the ball for conditions where no sound was present.

3.4.3 Cluster 3 - Participant 12

Participants belonging to cluster 3 generally focused longer on the goal AOI than on the ball AOI. The total error rate for all conditions for participants in this cluster was 32 (median 26), i.e. the best performance of all clusters. Contrary to the behavior of participants in cluster 2, most participants in cluster 3 focused longer on the ball AOI for the creaking sound model than for the swishing sound model. For participant 12, this tendency can be observed both for haptic and nonhaptic conditions. The participant performed the experiment in the following condition order: AVH1, AV1, AVH2, AV2, V and VH. The coordinates of the averaged left- and right eye gaze positions are displayed in Fig. 7. As seen in the figure, the gaze points are distributed in a similar pattern in all conditions. For the ball AOI, the participant had total fixation durations in the range between 3.05 and 7.44 seconds (lowest for VH and highest for AV1) and total fixation counts in the range between 10 and 28 counts (lowest for V and highest for AV1). As indicated by the plots presented in Fig. 7, the sonifications of this participant's gaze behavior do sound different in different conditions.

3.4.4 Cluster 4 - Participant 04

This participant was classified separately in a solo cluster. The total error rate for all conditions for this participant was 52.0. The participant performed the experiment in the following condition order: AVH1, VH, V, AVH1, AV2 and AV1. As indicated by the metrics in Tab. 5, this participant focused approximately as much on the goal AOI as on the ball AOI in all conditions. For the ball AOI, participant 4 had total fixation durations in the range between 15.98 and 31.40 seconds (lowest for V and highest for AVH1) and total fixation counts in the range between 28 and 48 counts (lowest for VH and highest for AVH1). The V condition stands out compared to other conditions.

3.4.5 Cluster 5 - Participant 15

Participants in cluster 5 showed the opposite behavior of participants in cluster 3: they generally focused more on the ball AOI than on the goal AOI. The total error rate for all conditions for participants in this cluster was 95.0 (median 96.0), i.e. the highest error rate of all clusters. Participant 15 performed the conditions in the following order: V, AV1, VH, AVH1, AV2 and AVH2. The coordinates of the averaged left- and right eye gaze positions are displayed in Fig. 8. As seen in the figure, gaze trajectories appear to

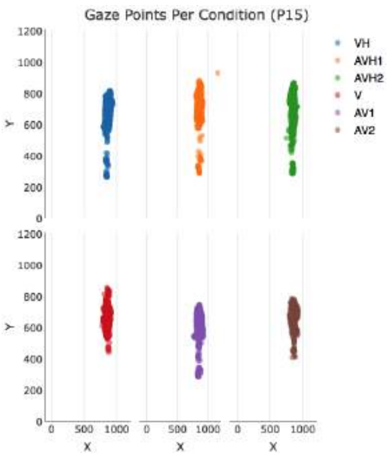

Figure 8. Gaze behavior of participant 15. Points correspond to the coordinates of the averaged left- and right eye gaze position on the screen.

have been very focused for this participant. Sonifications reveal that the participant showed similar gaze trajectory patterns for all conditions. This is also clearly shown in Fig. 4. In general, the participant had rather long total fixation durations, in the range between 34.57 and 46.83 seconds (lowest for AV2 and highest for V) and total fixation counts in the range between 28 and 40 counts (lowest for AV2 and highest for AVH1), for the ball AOI.

4. DISCUSSION

In this paper, we explored the effect of auditory feedback on gaze behavior in haptic versus nonhaptic conditions when manipulating an object in a virtual environment. Since there can be substantial differences in gaze behavior between participants even for identical tasks, it is good practice to use a within-participant design in these contexts, in order to make valid comparisons of performance [20]. As in [11], we observed very heterogeneous behaviors among participants. The effect of movement sonification was thus evaluated on an individual level, similarly to previous studies on the effect of sonification on gaze behavior [10, 11]. Possibly due to the limited sample size of this study (and the fact that a substantial part of the eye tracking data had to be discarded since the percentage of usable gaze data was too low), no significant differences between feedback conditions could be observed. We can therefore not make any general assumptions about the hypotheses defined in section 2.1. Some general trends could, however, be identified.

For the full duration of the first 15 attempts to throw the ball into the goal, the total fixation count median value was highest for condition V, in which no auditory or haptic feedback was provided. For total fixation duration, the median was highest for condition VH (followed by condition V). For fixation count, the V condition also had higher median value than the AV1 and AV2 conditions, and the VH condition had higher median value than the AVH1 and AVH2 conditions. Although no significant difference between conditions could be observed, these results indicate that it would be worth investigating if the presence of auditory feedback affects gaze behavior in future studies.

Cluster analysis revealed that the participants could be divided into five different clusters in which similar gaze behaviors were found. Within only 13 participants, 5 different sorts of behavior were observed. In total, 6 participants had longer fixation duration on the ball AOI than the goal AOI. A total of 7 participants showed the opposite behavior. On an individual level, we could observe that some participants appeared to have been affected by the presented auditory feedback (for example, differences appeared for participants in cluster 2 and 3), whereas others showed very similar gaze behavior across all conditions. Even when an effect of auditory models seemed to have been present, the behaviors were not consistent. For example, participants in cluster 2 showed the opposite behavior of participants in cluster 3.

The performed interviews shed some light on the effects of auditory feedback. A few participants stated that it was easier to aim in the conditions with auditory feedback. A possible explanation, given by one of the participants, was: "*It was easier to get an understanding of where the ball was located [in auditory conditions]. I could make use of how much sound that I needed [...] You could find a rhythm in the sound that you could follow.*". This quote indicates that the sonification of the movement gave some information that otherwise had to be provided graphically. Another participant further emphasized this by stating: "*I liked the sound because I could hear which velocity I had. Then I could also focus more on the goal and the direction of the ball, instead of just on the ball.*".

Mean performance varied between the clusters. Sorting cluster by best to worst mean performance, we obtain: cluster 3, cluster 1, cluster 2, cluster 4 and cluster 5. We can observe that best performance was observed for participants belonging to clusters where the longer fixation duration was on the goal AOI, rather than on the ball AOI. The cluster with participants who focused longer on the ball AOI than on the goal AOI (cluster 5) had worst overall performance.

Large differences in eye movements between participants encourage large datasets. To address the issue of heterogeneous behaviors, we propose a future more controlled experiment with larger sample size to evaluate the effect of sonification on gaze behavior. The current study focused specifically on movement sonification (based on position and velocity). Possibly, auditory feedback might have been more powerful in terms of affecting gaze behavior if the sonification had focused on guiding the user in terms of signaling if the performed movement was "correct" or not (i.e. error sonification).

Finally, we encourage the use of sonification for rapid exploration of large sets of gaze data. Sonifying gaze trajectories enables the listener to identify behaviors not easily discovered in 2D plots. For example, patterns such as rhythm and overall duration of a gaze trajectory are easily identified using sonification. In this particular study, sonification enabled us to confirm different gaze behaviors identified in the cluster analysis, such as the distinction between focusing mostly on the ball versus goal AOI for cluster 3 and 5.

5. CONCLUSIONS

This exploratory study focused on the effect of auditory feedback on gaze behavior. Analysis of eye tracking metrics revealed several clusters of different gaze behaviors among only 13 participants. Some of the participants appeared to have been affected by the presence of auditory feedback, whereas other showed similar gaze trajectory patterns across conditions. More thorough studies are required in order to be able to draw any general conclusions regarding the effect of auditory feedback in this context. This exploratory study raises intriguing questions that motivate future large-scale studies on the effect of auditory feedback on gaze behavior.

Acknowledgments

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.

6. REFERENCES

[1] J. Vroomen and B. d. Gelder, “Sound enhances visual perception: cross-modal effects of auditory organization on vision.” *Journal of experimental psychology: Human perception and performance*, vol. 26, no. 5, p. 1583, 2000.

[2] L. Shams, Y. Kamitani, and S. Shimojo, “Illusions: What you see is what you hear,” *Nature*, vol. 408, no. 6814, pp. 788–788, 2000.

[3] J. Mishra, A. Martínez, and S. A. Hillyard, “Effect of attention on early cortical processes associated with the sound-induced extra flash illusion,” *Journal of cognitive neuroscience*, vol. 22, no. 8, pp. 1714–1729, 2010.

[4] S. L. Calvert and T. L. Gersh, “The selective use of sound effects and visual inserts for children’s television story comprehension,” *Journal of Applied Developmental Psychology*, vol. 8, no. 4, pp. 363–375, 1987.

[5] A. R. Seitz, R. Kim, and L. Shams, “Sound facilitates visual learning,” *Current Biology*, vol. 16, no. 14, pp. 1422–1427, 2006.

[6] G. Song, D. Pellerin, and L. Granjon, “Different types of sounds influence gaze differently in videos,” *Journal of Eye Movement Research*, vol. 6, no. 4, pp. 1–13, 2013.

[7] G. Song, D. Pellerin, and L. Granjon, “Sound effect on visual gaze when looking at videos,” in *Signal Processing Conference, 2011 19th European*. IEEE, 2011, pp. 2034–2038.

[8] A. Coutrot, N. Guyader, G. Ionescu, and A. Caplier, “Influence of soundtrack on eye movements during video exploration,” *Journal of Eye Movement Research*, vol. 5, no. 4, p. 2, 2012.

[9] X. Min, G. Zhai, Z. Gao, C. Hu, and X. Yang, “Sound influences visual attention discriminately in videos,” in *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*. IEEE, 2014, pp. 153–158.

[10] V. Losing, L. Rottkamp, M. Zeunert, and T. Pfeiffer, “Guiding visual search tasks using gaze-contingent auditory feedback,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 1093–1102.

[11] E. Boyer, “Continuous auditory feedback for sensorimotor learning,” Ph.D. dissertation, Université Pierre et Marie Curie-Paris VI, 2015.

[12] A. Poole and L. J. Ball, “Eye tracking in hci and usability research,” *Encyclopedia of human computer interaction*, vol. 1, pp. 211–219.

[13] “Computer interface evaluation using eye movements: methods and constructs,” *International Journal of Industrial Ergonomics*, vol. 24, no. 6, pp. 631 – 645, 1999.

[14] A. Duchowski, *Eye tracking methodology: Theory and practice*. Springer Science & Business Media, 2007, vol. 373.

[15] M. A. Just and P. A. Carpenter, “Eye fixations and cognitive processes,” *Cognitive Psychology*, vol. 8, no. 4, pp. 441 – 480, 1976.

[16] E. Frid, J. Moll, R. Bresin, and E.-L. Sallnäs-Pysander, “Influence of movement sonification on performance in a virtual throwing task performed with and without haptic feedback,” *Manuscript submitted for publication.*

[17] F. Conti, F. Barbagli, D. Morris, and C. Sewell, “{CHAI} an open-source library for the rapid development of haptic scenes,” in *Proceedings of the IEEE World Haptics Conference*, 2005. [Online]. Available: http://www.chai3d.org/

[18] M. Wright, “Open sound control: an enabling technology for musical networking,” *Organised Sound*, vol. 10, no. 03, pp. 193–200, 2005.

[19] S. D. Monache, P. Polotti, and D. Rocchesso, “A toolkit for explorations in sonic interaction design,” in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*. ACM, 2010, p. 1.

[20] J. H. Goldberg and A. M. Wichansky, “Chapter 23 - eye tracking in usability evaluation: A practitioner’s guide,” in *The Mind’s Eye*, J. Hyn, R. Radach, and H. Deubel, Eds. Amsterdam: North-Holland, 2003, pp. 493 – 516.