The Effect of Pitch in Auditory Error Feedback for Fitts' Tasks in Virtual Reality Training Systems

Anil Ufuk Batmaz*
Simon Fraser University
BC, Canada

Wolfgang Stuerzlinger†
Simon Fraser University
BC, Canada

ABSTRACT

Fitts' law and the associated throughput measure characterize user pointing performance in virtual reality (VR) training systems and simulators well. Yet, pointing performance can be affected by the feedback users receive from a VR application. This work examines the effect of the pitch of auditory error feedback on user performance in a Fitts' task through a distributed experiment. In our first study, we used middle- and high-frequency sound feedback and demonstrated that high-pitch error feedback significantly decreases user performance in terms of time and throughput. In the second study, we used adaptive sound feedback, where we increased the frequency with the error rate, while asking subjects to execute the task "as fast/as precise/as fast and precise as possible". Results showed that adaptive sound feedback decreases the error rate for "as fast as possible" task execution without affecting the time. The results can be used to enhance and design various VR systems.

Index Terms: Human-centered computing—Human Computer Interaction (HCI); Human-centered computing—Virtual Reality; Human-centered computing—Pointing;

1 INTRODUCTION

In virtual environments (VEs), typically presented in virtual reality (VR) and augmented reality (AR) systems, software developers and practitioners can easily collect detailed 3D user movement data, with lower cost and higher performance compared to conventional methods [45]. The controllability of a VE, such as the ability to change the lighting or wind conditions, allows users to practice a task as much as required while still affording different degrees of challenge. Similarly, the ability to perfectly repeat a scenario, like a ball following a certain trajectory, which is hard to do in the real world, makes such training systems even more attractive for trainees [17]. Further, trainees can receive such training without environmental challenges or causing environmental harm and even train remotely, e.g., in their home, saving money, time, and increasing their personal productivity.

Previous work investigated training systems and simulators in VR in a variety of fields, including but not limited to sports [23, 33, 59, 62], surgery [16, 44, 52], aircraft maintenance and piloting [27, 70], firefighter [22, 25, 49], and marine training [36, 66]. Our paper focuses on training scenarios where the trainees aim to improve their eye-hand coordination performance through VR training systems and simulators, by decreasing their reaction time while improving precision and accuracy. Especially for fields that require fast and accurate movements, such as catching a ball in sports, reaching a button inside the cockpit of an aircraft, or positioning the tool-tip into the correct location in surgery, eye-hand coordination training systems are already deployed widely.

*e-mail: abatmaz@sfu.ca
†e-mail: w.s@sfu.ca

Previous work showed that the Fitts' task and throughput measure are useful to assess user performance in VR [68] and to reduce reaction time [13]. This latter study asked participants to perform the experiment as fast and as precise as possible, with a single kind of auditory error feedback. Yet, they did not investigate how user performance changes with different forms of auditory feedback and task execution conditions in VR. A recent study extended throughput assessment to VE eye-hand coordination training systems [8].

Yet, other work has identified that the pitch of audio feedback can alter user performance in steering tasks for VR medical training systems and simulators [4, 7]. Researchers were able to "control" user performance by changing the frequency of the error feedback. We believe that these outcomes with a steering task mean that it is possible to control the trainee performance in Fitts' law pointing tasks, i.e., through the task-inherent speed-accuracy trade-off. Other research has shown that arbitrarily designed sound feedback can even have detrimental effects on the users' motivation and interpretation, and thus, on their performance [61].

While trainees practice with a training systems or simulator, they can use different learning and task execution strategies [5]. Research has identified that it is better for novices to focus on precision (rather than speed) to increase their efficiency [4, 5]. Beyond this criterion, throughput incorporates speed, precision, and accuracy of the users into one measure and thus combines different learning criteria in a single value [8, 13]. The combination of speed, error, and throughput assessment for eye-hand coordination training allows trainers to monitor trainee performance and to inform trainees on how to best improve their skills. Thus, to increase the efficiency of VR training systems and simulators for eye-hand coordination performance it is crucial to understand how to control trainee task execution strategies.

This paper investigates how different sounds as error feedback during a Fitts' task affect user performance, also with different task execution strategies. Also, we identify design recommendations for VR-based training systems and simulators targeted at improving eye-hand coordination motor performance.

Our contributions are:

- High-frequency error feedback significantly decreases user performance in VR training systems in terms of time and throughput while decreasing the error rate.
- An identification of the effect of different error sound frequencies on task execution strategies in a Fitts' task.
- Suggestions for appropriate auditory error feedback for designers of VR training systems and simulators.

In this study, we investigate the following research questions: How does user performance change with different sound frequencies for error feedback in Fitts' task for time, error rate and throughput? Which trade-offs appear with high-frequency auditory feedback? And how can sound feedback be used to improve user performance by encouraging different task execution strategies?

To answer these questions, we extend previous work on the effect that pitch has in auditory feedback for steering tasks [7] to Fitts' law pointing tasks through two different, remotely conducted, user studies with the same participant pool. In the first study, we focused on three forms of sound feedback: no feedback, a middle C pitch (C4), and a high C (C8), for both virtual hand and ray casting selection techniques. In our second study, we introduced adaptive feedback and increased the pitch when subjects made more errors. We investigated adaptive feedback and constant pitch (C4) error feedback for three task execution strategies: "as fast/as precise/as fast and precise as possible". We measured participants' time, error rate, and throughput performance during the task execution. We show that our findings better inform trainees as to how to deal with the fundamental speed-accuracy trade-off. We believe that our results can be applied to VR simulators and training systems to ultimately give trainees better options to train themselves, especially for situations where real-world scenarios are hard to replicate.

2 Previous Work

2.1 Fitts' law & Effective Throughput

Fitts' law is a mathematical equation used to model the movement time for rapid human movements [26] in Human-Computer Interaction studies. Equation 1 shows the Shannon capacity formulation of Fitts' law, which we use in our work here:

$$Movement\,Time\,(MT) = a + b * log_2\left(\frac{A}{W} + 1\right) = a + b * ID \quad (1)$$

In Equation 1, A is the distance between target centers and W the size of each target. The *log* term in the equation represents the task difficulty or the index of difficulty *ID*. The coefficients a and b are empirically derived via linear regression. In this work, we also use ISO 9241-411:2012's throughput [34] as assessment criterion:

$$Throughput = \left(\frac{ID_e}{MT}\right) \quad (2)$$

In Equation 2 ID_e is the effective index of difficulty. According to ISO 9241-411:2012 [34], ID_e is the "user **precision** achieved in accomplishing a task" and calculated as:

$$ID_e = \log_2\left(\frac{A_e}{W_e} + 1\right) \quad (3)$$

In Equation 3, the effective distance A_e is the real distance traversed to execute the task, i.e., the distance between two selection points, and W_e is the effective target width, which is calculated as $W_e = 4.133 \times SD_x$, where SD_x is the standard deviation of selection coordinates along the task axis. SD_x represents the **accuracy** of the task [40, 41]. Thus, Equation 2 combines time, precision, and accuracy into one equation, which enables us to better understand the trade-offs in user motor performance [34].

2.2 Auditory feedback

Any change in perceptual information can potentially change user motor performance [61]. For instance, previous work has shown that auditory feedback affects the reaction time of subjects [51] and reduces the target acquisition time [50, 54, 56].

Sound feedback in Fitts' tasks has been studied as uni-modal and multi-modal feedback [2, 15, 73]. Akamatsu et al. [1] used 2kHz sounds and compared haptic, auditory, and visual feedback. They showed that the combination of haptic, visual, and auditory feedback does not increase user performance as much as haptic feedback alone. On the other hand, Brent et al. [31] used a 1kHz sine wave and showed that subjects were faster with confirmatory auditory feedback. Sterkenburg et al. increased the frequency of the sine wave, i.e., the pitch, when participants hit closer to targets and showed that the combination of visual and continuous auditory feedback increases throughput performance [64, 65].

Previous work also used auditory *error* feedback to signal to trainees if they needed to improve their performance [38, 39, 60]. For instance, Konttinen et al. [39] mapped user movements to pitch in a shooting task, and showed that, on average, the deviation of the subjects decreased with higher frequencies. Similarly, Sigrist et al. [60] mapped deviations in rowing-type movements to auditory pitch and showed that subjects can follow a desired trajectory. Other similar work, e.g., [19, 28, 35], also focused on mapping the error in the movement trajectory to auditory feedback for motor performance improvement. Still, the effect of auditory error feedback on 3D pointing tasks has not been investigated before.

In some 2D Fitts' law studies, the auditory component of the selection action is provided through the 2D mouse click or the tap on a tablet, e.g., [1]. In contrast, researchers often provide auditory error feedback to notify participants if they failed to successfully point to a target in VR pointing studies, e.g., [55, 69]. In our study, we used auditory error feedback in a similar manner, where we played a sound when subjects did not hit (missed) the target. Previous work has shown that to improve subjects' motor performance, key events should be correlated to changes in volume [61]. Yet, in pointing tasks, it is common to observe low error rates. By (frequently) giving positive auditory feedback for hits, the awareness of a miss could be reduced, which could negatively impact user performance. Thus, it is better to give *auditory feedback for errors* in pointing tasks.

2.3 Sensory performance assessment

Previous work on different learning strategies showed that time alone is not a good assessment criterion since individuals can follow different learning strategies [5, 6]. Especially for novice trainees, research suggests that they should focus on precision in their initial task execution to improve their motor skill acquisition in training systems [4, 5]. In VR training systems and simulators, user performance can be easily monitored by trainers and then used to give feedback to the trainee to facilitate the learning process. Such an active feedback training method can improve trainee performance and increase training efficiency [21, 23, 57].

2.4 Fitts' law and throughput as assessment criterion

Fitts' law, as shown in Equation 1, is a well-known performance assessment criteria to compare user pointing performance with different input methods in HCI studies. Previous studies showed that throughput is useful to assess human performance. For instance, Teather and Stuerzlinger [68] examined pointing tasks in VR through time, error, and throughput. Throughput was also used as an assessment criterion in various VR applications to monitor the user performance. Scheme and Englehart used 3D Fitts' task and throughput in a clinical evaluation of prosthetics as training systems [58]. Kim et al. used Fitts' law to analyze user performance for individuals with chronic strokes [37]. A recent series of VR/AR studies on eye-hand coordination training systems showed that the Fitts' task and throughput measure are suitable assessment criteria for tasks designed to improve the reaction time of athletes [8, 13, 14, 48].

3 User Study 1

3.1 Motivation & Hypotheses

Here, our goal is to investigate if user motor performance is affected negatively with high-pitch auditory feedback in a Fitts' task.

H1.1. **User performance is altered with different auditory error feedback in Fitts' tasks:** As shown in previous studies in other domains, we expect that participants' performance changes with different feedback sounds in Fitts' tasks [1, 31].

H1.2. **High-pitch error feedback increases task execution time, while also decreasing the error rate and increasing throughput for Fitts' tasks:** Previous work on the steering task showed that task execution time decreased significantly with higher pitch, while subjects made fewer errors [4, 7]. We expect similar results for pointing tasks. Since throughput is correlated to task execution time, we also expect a decrease in throughput.

3.2 Subjects

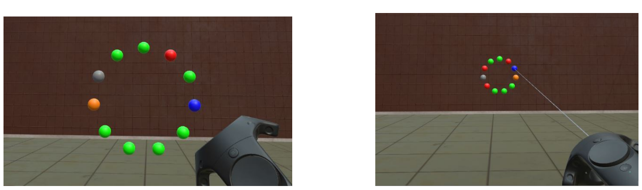
We recruited 17 right-handed and one left-handed participants (10 male and 8 female) on average 29.31±4.29 years old.

3.3 Procedure

We conducted the user study remotely. The participants started the experiment by filling a demographic questionnaire. Then, the experimenter explained the study to the participant. In the VE, subjects were sitting in the middle of an empty room with depth cues, as shown in the Fig. 1. During the experiment, participants used their dominant hand to control a cursor via the VR controller to point at targets. They used the space bar on a keyboard with their non-dominant hand to select a target. We made this choice to avoid the adverse consequences of the Heisenberg spatial effect [18].

We used an ISO 9241-411:2012 [34] task with 11 gray targets placed in a circle at the participants' eye level. Participants selected each target with the cursor attached to the VR controller with two different **selection techniques**, virtual hand (Fig. 1(a)) and ray casting (Fig. 1(b)). Targets were placed 0.4 m and 1.5 m away from the subjects for the virtual hand and ray casting pointing conditions, respectively. These distances were chosen based on the outcomes of previous mid-air selection work [9]. To eliminate diplopia, we placed the 1 cm cursor sphere 3 cm above the VR controller for the virtual hand condition. For the ray casting method, the cursor was displayed at the intersection between the ray and the 2D target plane, i.e., always 1.5 m away from the user.

During a round of trials, we asked subjects to select targets with the spacebar when the cursor was inside a target. While the cursor was inside the orange-coloured target, we changed its colour to blue as visual feedback, i.e., highlighting. If the cursor was inside that target when the user hit the space bar, we changed the target's colour to green to provide positive feedback to the participants. On the other hand, when the cursor was outside of the target, the target's colour was changed to red, and we played a sound as error feedback. The first target was randomly chosen by the software and participants executed the task either in a clockwise or counter-clockwise direction. For the first study, we asked subjects to be "as fast and as precise as possible" while selecting targets.



(a) (b)

Figure 1: Study 1 selection techniques (a) virtual hand (b) ray casting.

For **auditory error feedback**, we used either no sound or two different forms of auditory feedback. In the control condition, we did not give auditory error feedback, which is equivalent to a 0Hz tone. For the two other conditions, we played a middle C (C4 - 262Hz) or the highest C on a piano keyboard (C8 - 4186Hz) for 0.25 seconds for errors. We made this choice so that our work would be comparable to previous work [4,7]. Moreover, the duration is similar to the "adequate response time" of 243ms suggested by Brungart et al. [20]. Participants heard this sound either through the HMD's built-in headphones or external audio headsets.

To vary the index of difficulty (ID), we used different **target distances** and **target sizes**, see below.

At the end of the study, subjects filled a questionnaire about their observations on the different forms of auditory feedback. Overall, the first study took around 10 minutes, and we counterbalanced the **auditory feedback** and **selection technique** with a Latin square to eliminate potential learning effects.

3.4 Experimental Design

In this first study, we used a two-factor within-subjects design with three **auditory error feedback** (3_{AEF} = no auditory feedback, C4 (262Hz) and C8 (4186Hz)) conditions and two **selection techniques** (2_{ST} = virtual hand and ray casting) conditions, comprising a $3_{AEF} \times 2_{ST}$ design. We measured task execution time (seconds), error rate (%), and effective throughput (bits/s) of the subjects. We varied the index of difficulty *ID*, by using three target sizes (3_{TS} = 1.5, 2.5 and 3.5 cm) and two target distances (2_{TD} = 12.5 and 25 cm), which created 6 unique IDs between 2.19 and 4.14. Target sizes and distances were randomly selected for each round of trials. Each subject performed $3_{AEF} \times 2_{ST} \times 6_{ID} \times$ 11 repetitions = 396 trials.

Table 1: Study 1 Experimental Conditions, AEF = Auditory error feedback, NONE = No auditory feedback, C4 = 262 Hz, C8 = 4186 Hz, ST = Selection technique, VH = Virtual hand, RC = Ray casting, A = Target distance, W = Target width.

$3_{AEF} \times 2_{ST}$ conditions counterbalanced with Latin Square

AEF	NONE	NONE	C4	C4	C8	C8
ST	VH	RC	VH	RC	VH	RC

Randomized task variables for each condition

A	12.5	12.5	12.5	25	25	25
W	1.5	2.5	3.5	1.5	2.5	3.5

3.5 Results

The data for this first user study were analyzed using Repeated Measures (RM) ANOVA in SPSS 24. We used Skewness (S) and Kurtosis (K) for normality analysis and considered data as normally distributed when the S and K values were within ± 1.5 [30, 42]. When the data was not normally distributed, we used ART [71]. We only report significant results and used the Bonferroni method for post-hoc analyses. Results are illustrated as means and standard error of means in figures. We applied Huynh-Feldt correction when the ε was less than 0.75.

Time results: The time dependent variable was normally distributed, S = 0.99 and K = 1.36. According to the one-way RM ANOVA results in Table 2, subjects were slower with high-pitch feedback (Fig. 2(a)) and with the ray casting condition (Fig. 2(b)). We did not find a significant interaction between auditory feedback and selection technique ($F(2,34) = 0.047$, $p = 0.293$, $\eta^2 = 0.141$). We also did not identify a significant three-way interaction between auditory feedback, selection technique, and ID ($F(7.482,126.275) = 0.021$, $p = 0.229$, $\eta^2 = 0.066$).

Error rate results: The error rate dependent variable was not normally distributed, S = 1.3, K = 2.5, so we used ART. The one-way RM ANOVA results for error rate are shown in Table 2, Fig. 2(c)), and Fig. 2(d)). Subjects made fewer errors with the high-pitch feedback (C8) and the virtual hand condition. We also found a significant two-way interaction between auditory feedback and selection technique $F(2,34) = 9.919$, $p < 0.001$, $\eta^2 = 0.368$. The results are shown in Fig. 2(g), where the error rate of the subjects decreased with the virtual hand condition compared to ray casting for C4 and no auditory feedback. For the C8 tone, even though the error rate decreased for the virtual hand condition, the interaction was not significant. We did not find a significant three-way interaction between auditory feedback, selection technique, and ID ($F(10,170) = 1.664$, $p = 0.093$, $\eta^2 = 0.089$).

Throughput results: Throughput was normally distributed, S = 0.43, K = 0.39. According to the one-way RM ANOVA results shown in Table 2, Fig. 2(c)) and (Fig. 2(d), the throughput performance of the subjects decreased with C8 feedback and with ray casting. We did not observe a significant two-way interaction in the RM ANOVA between auditory feedback and selection technique for throughput (F(2,34) = 0.331, p = 0.720, $\eta^2 = 0.019$), nor a three-way interaction between auditory feedback, selection technique, and ID (F(10,170) = 1.195, p = 0.297, $\eta^2 = 0.125$).

Table 2: Study 1 One-Way RM ANOVA results

	Auditory Feedback	Selection Technique	ID
Time	F(2, 34)= 9.679, p<0.001, η^2= 0.363	F(1, 17)= 19.631, p<0.001, η^2= 0.536	F(2.617, 44.492)= 88.725, p<0.001, η^2= 0.839
Error rate	F(2,34)= 33.372, p<0.001, η^2= 0.663	F(1,17)= 23.531, p<0.001, η^2= 0.581	F(5,85)= 6.160, p<0.001, η^2= 0.266
Throughput	F(2,34)= 8.639, p<0.001, η^2= 0.337	F(1,17)= 83.904, p<0.001, η^2= 0.832	F(5,85)= 18.850, p<0.001, η^2= 0.526

The results raise the following question: since subjects made fewer errors with C8 feedback, one could expect to see better accuracy. Yet, the throughput results show that subjects' performance decreased significantly. Thus, we also analyzed the standard deviation (SD_x) [40] for auditory feedback, selection technique, and ID. SD_x was not normally distributed, S = 10.93, K = 153.10 and we found significant interactions for auditory feedback F(2,34) = 42.039, $p < 0.001$, $\eta^2 = 0.712$ and ID (F(5,85) = 4.695, $p < 0.01$, $\eta^2 = 0.216$). Selection technique was not significant for SD_x (F(1,17)= 1.968, n.s., η^2= 0.104). The results for auditory feedback in Fig. 2(h) show that the precision of the subjects significantly decreased when there was no auditory feedback.

3.5.1 Subjective results

Most subjects, 14 out of 18, preferred a C4 tone, 2 preferred no auditory feedback, and 2 preferred the C8. We asked participants about their thoughts on the C8 tone, and they commented that it was "annoying", "very uncomfortable", "it [C8] made me feel very rushed and panicked", "very irritating", "quite alarming", "I tried to avoid that [C8] as much as possible", "it was disturbing. I believe it affected my speed, slowed down a lot", "distracting", and that it "scared me". In contrast, one subject commented that "The irritating sound helped me [to] focus better".

3.5.2 Fitts' Law Analysis

Linear regressions according to Fitts' Law, Equation 1, show that movement time can be modelled as MT = 0.24 + 0.22*ID, $R^2 = 0.19$ without sound feedback, MT = 0.17 + 0.22*ID, $R^2 = 0.83$ for the C4 pitch and MT = 0.93 + 0.27*ID, $R^2 = 0.93$ for the C8. These results are shown in Fig. 3(a). The results for selection technique are similarly shown in Fig. 3(b), virtual hand MT = 0.19 + 0.2*ID, $R^2 = 0.93$ and ray casting MT = 0.14 + 0.26*ID, $R^2 = 0.87$.

3.6 Study 1 Discussion

Our first study investigated how different forms of auditory feedback affect user performance in terms of time, error rate, and throughput. We extended our results further by analyzing the standard deviation of the selection points to better understand the effects of different auditory error feedback.

Looking at the data shown in Table 2, we can see that changing the feedback's pitch can alter user performance. From previous work, we know that auditory cues can increase the spatial accuracy for virtual object interaction [74]. Also, the interaction between auditory feedback and selection technique in our results here identifies a performance difference between ray casting and virtual hand selection

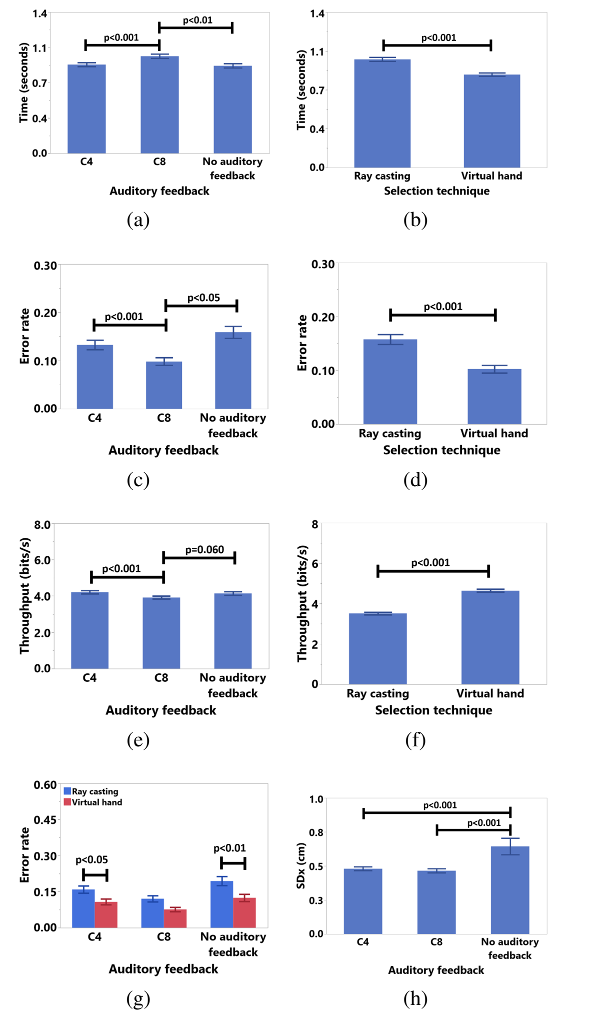


Figure 2: Time results for (a) auditory feedback and (b) selection technique; error rate results for (c) auditory feedback and (d) selection technique; effective throughput results for (e) auditory feedback and (f) selection technique. (g) Error rate interaction results for auditory feedback and selection technique and (h) SD_x interaction results for the first study.

techniques. However, we did not observe a significant difference for the C8 tone between ray casting and virtual hand, since subjects seem to have focused mostly on not making any errors with this form of auditory feedback. These results support our hypothesis **H1.1**, where different forms of auditory feedback affect user performance differently, regardless of the VR selection technique.

In **H1.2**, we hypothesized that high-pitch feedback decreases the error rate and throughput and increases execution time. According to our results, high-frequency auditory feedback, i.e., a C8, increased the execution time compared to middle-frequency auditory feedback, i.e., a C4, and no auditory feedback. On the other hand, the error rate significantly decreased for the C8 condition. Since throughput combines time, precision, and accuracy, the increase in the execution time had detrimental effects on the overall throughput. Thus, our results on time, precision, and throughput support **H1.2**.

We speculate that this result is related to participants' perception of the different sound frequencies and the associated cognitive load. In the real world, high-pitch sounds are often correlated to danger or potential damage and used to attract the attention of a person [24,53].

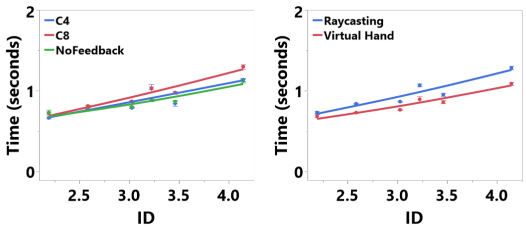


Figure 3: Fitts' law model for (a) sound feedback conditions and (b) task execution strategies.

Thus we believe that when the subjects heard the high-frequency sounds in this task, they changed their task execution strategy to select the targets more precisely. Since human cognitive capacity is limited, this meant that they had to de-prioritize execution time. This shift in priority also decreased their throughput.

The other explanation is that the used pitch did irritate or annoy participants [4, 7]. Thus, participants might have actively tried to avoid hearing the error sound, which then decreased their throughput and increased execution time. The participants' comments on the high-pitch sound support this speculation. Since subjects were not notably more precise with the C8 tone compared to the C4, our results for standard deviation also support this hypothesis. Their execution time increased while their error rate decreased, meaning that subjects only focused on selecting the targets correctly. However, if this kind of high-pitch auditory feedback would be used as positive feedback, e.g., [75], it could increase task execution time and decrease throughput. Thus, we see it as important to choose appropriate feedback for a given task. For pointing it is usually much more likely that participants hit the target, so giving unpleasant feedback for non-errors seems counter-intuitive.

Overall, the results of our first study identify that user performance in terms of time, precision, and error rate is dependent on the pitch used as error feedback. Thus, one might be able to control trainee performance by using varying sound feedback, which could be used to increase training efficiency.

4 User Study 2

4.1 Motivation & Hypotheses

Our motivation for the second study was to investigate how we can use different forms of auditory feedback to elicit different task execution strategies, which could then be applied in training systems to improve user performance. To investigate this effect, we introduced adaptive error feedback by increasing the error feedback sound frequency with each error. To evaluate this approach, we compared the effects of adaptive pitch for error feedback on user performance against a constant pitch.

H2.1. **Adaptive auditory error feedback increases user performance:** Based on the results of study 1, we speculate that increasing the sound frequency with each error in a round of trials could increase the participants' awareness of their performance, which could improve overall user performance. As participants might get annoyed if they hear the same tone repeatedly [61], we increased the pitch step-wise, which enabled participants to adapt their task execution strategy to reduce their error rate.

H2.2. **Adaptive auditory error feedback decreases user performance when selecting targets with the ray casting selection technique:** Adaptive feedback is designed to increase the participants' awareness of an increasing error rate. Also, based on the results of study 1, participants made more errors with the ray casting selection technique. Thus, when subjects select targets with a ray, we expect to observe a speed-precision trade-off where they get slower and their throughput decreases, to avoid increasing their error rate and hearing higher-frequency sounds.

4.2 Subjects

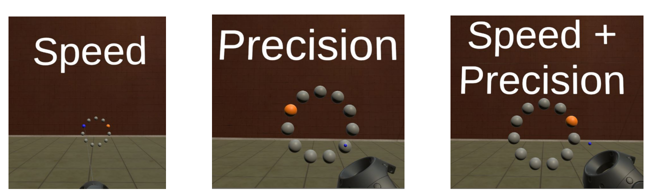
The same subjects as in the first study participated in this second study, as part of one experimental "session". We did not counterbalance the user studies since we wanted subjects to be already familiar with different pitches for feedback for this second study.

4.3 Procedure

This study was conducted remotely using the same participant pool as in study 1 within the same experimental session. After filling out a pre-questionnaire for the second study, subjects returned to the same VE as in the first study. Like in the first study, participants performed the second study with two different pointing conditions and six different IDs. In the second study, we changed the **auditory feedback** and introduced adaptive sound feedback, where we increased the pitch of the error sound with each error within each round of 11 targets. In other words, when the subjects made their first error, they heard a C4 (262Hz) as auditory feedback, for the second error we used a C5 (523Hz), for the third a C6 (1046Hz), for the fourth a C7 (2093Hz), and for five or more errors a C8 (4186Hz). For the constant-frequency baseline, we used a C4 as auditory error feedback. As in user study 1, we played each sound for 0.25 s.

In this study, subjects performed the experiment with three different **task execution strategies**. In the first, we asked participants to perform the tasks "as fast as possible," focusing only on their task execution time (Fig. 4(a)). In the second condition, we asked subjects to perform the tasks "as precisely as possible" (Fig. 4(b)), i.e., we asked subjects to focus only on selecting the targets precisely while ignoring speed. As the third condition, we asked subjects to perform the tasks "as fast and as precise as possible" (Fig. 4(c)), where they focus simultaneously on their speed and precision.

We showed the current task execution strategy as floating text behind the targets during the experiment to help participants keep track of the current strategy. The experimenter also monitored each participants' performance remotely via teleconference during the study to ensure that participants followed the current task execution strategy, and used verbal feedback if participants were deviating clearly from the strategy. For instance, if a subject was making errors in the "as precise as possible" condition, the experimenter encouraged the participant to slow down and to focus on the precision.



(a) (b) (c)

Figure 4: Illustration of study 2 task execution strategy conditions: (a) speed task with distal pointing (b) precision task with virtual hand and (c) speed and precision task with virtual hand.

At the end of the study, subjects filled a questionnaire about the auditory feedback. Overall, the second study took around 20 minutes and we counterbalanced the **sound feedback** and **task execution strategies** with a Latin square to eliminate potential learning effects.

4.4 Experimental Design

In this study, we used a three-factor within-subjects design with two **auditory feedback** (2_{AF} = adaptive sound feedback, constant sound feedback) conditions with three different **task execution strategies** (3_{TES} = as fast as possible, as precise as possible, and as fast and as precise as possible) and two **selection techniques** (2_{ST} = virtual hand and ray casting) conditions, comprising a $2_{AF} \times 3_{TES} \times 2_{ST}$ design. We measured task execution time (seconds), error rate (%), and effective throughput (bits/s) of the subjects. We varied the index of difficulty *ID*, by using three target sizes (3_{TS} =1.5, 2.5 and 3.5 cm) and two target distances (2_{TD} = 12.5 and 25 cm), which created 6 unique IDs between 2.19 and 4.14. Each subject performed $2_{AF} \times 3_{TES} \times 2_{ST} \times 6_{ID} \times$ 11 repetitions = 792 trials.

Table 3: Study 2 Experimental Conditions, AF = Auditory feedback, ASF = Adaptive sound feedback, CSF = Constant sound feedback, ST = Selection technique, VH = Virtual hand, RC = Ray casting, TES = Task execution strategy, S = Speed, P = Precision, and S+P = Speed and Precision.

$2_{AF} \times 3_{TES}$ conditions counterbalanced with Latin Square

AF	ASF	CSF	ASF	CSF	ASF	CSF
TES	S	S	P	P	S+P	S+P

2_{ST} independently counterbalanced conditions

ST	VH	RC

Randomized task variables for each condition

A	12.5	12.5	12.5	25	25	25
W	1.5	2.5	3.5	1.5	2.5	3.5

4.5 Results

As for the data from the first study, we used three-way RM ANOVA with SPSS 24 and only report significant results. We considered data as normal when the S and K were within ± 1.5 [30,42].

The one-way RM ANOVA results for selection technique are similar to the outcomes of user study 1. Thus, we just show them in Table 4 and do not discuss these results further here.

Table 4: Study 2 results

	Auditory Feedback	Task Execution Strategy	Selection Technique	ID
Time	F(1, 17) = 20.517, p<0.001, η^2=0.547	F(2, 34) = 213.094, p<0.001, η^2=0.926	F(1, 17) = 72.59, p<0.001, η^2=0.001	F(5, 85) = 189.763, p<0.001, $\eta^2 = 0.918$
Error rate	F(1, 17) = 1.859, n.s., η^2=0.099	F(1.45, 24.73) = 99.488, p<0.001, η^2=0.854	F(1, 17) = 13.36, p<0.001, η^2=0.44	F(3.581,60.871) =73.951, p<0.001, $\eta^2 = 0.813$
Throughput	F(1,17) = 2.945, n.s., η^2=0.148	F(2, 34) = 155.449, p<0.001, η^2=0.901	F(1, 17) =202.904, p<0.001, η^2=0.923	F(3.298,56.067) =31.967, p<0.001, $\eta^2 = 0.653$

Time: The time dependent variable was not normally distributed, S = 1.33 and K = 2.35, thus we used ART [71] for analysis. According to the one-way RM ANOVA results in Table 4, Fig. 5(e) and Fig. 5(f), subjects were faster with constant sound feedback (Mean (M) = 0.779, Standard Error of Mean (SEM) = 0.012) compared to adaptive sound feedback (M = 0.805, SEM = 0.011) and the "as fast as possible" task execution strategy. We also found a significant two-way interaction between task execution strategy and auditory feedback for time (F(1,17) = 80.286, $p < 0.001$, $\eta^2 = 0.825$ in Fig. 5(e)). Results showed that subjects were slower with the "as precise as possible" task execution strategy. Furthermore, we found a significant interaction between selection technique and auditory feedback for time (F(1,17) = 80.286, $p < 0.001$, $\eta^2 = 0.825$ Fig. 5(f). According to these results, the task execution time of the participants increased for ray casting with adaptive auditory feedback. Also, three-way RM ANOVA results showed that there is a significant interaction between auditory feedback, task execution strategy, and ID (F(10,170) = 6.686, $p < 0.01$, $\eta^2 = 0.282$). Subjects were slower with the virtual hand condition compared to ray casting, except when they experienced constant feedback with ID = 2.19. Four-way ANOVA results showed that there is a significant interaction between auditory feedback, task execution strategy, selection technique, and ID (F(10,170) = 10.357, $p < 0.01$, $\eta^2 = 0.379$). Subjects were particularly slower with the 1.5 cm target size with ray casting

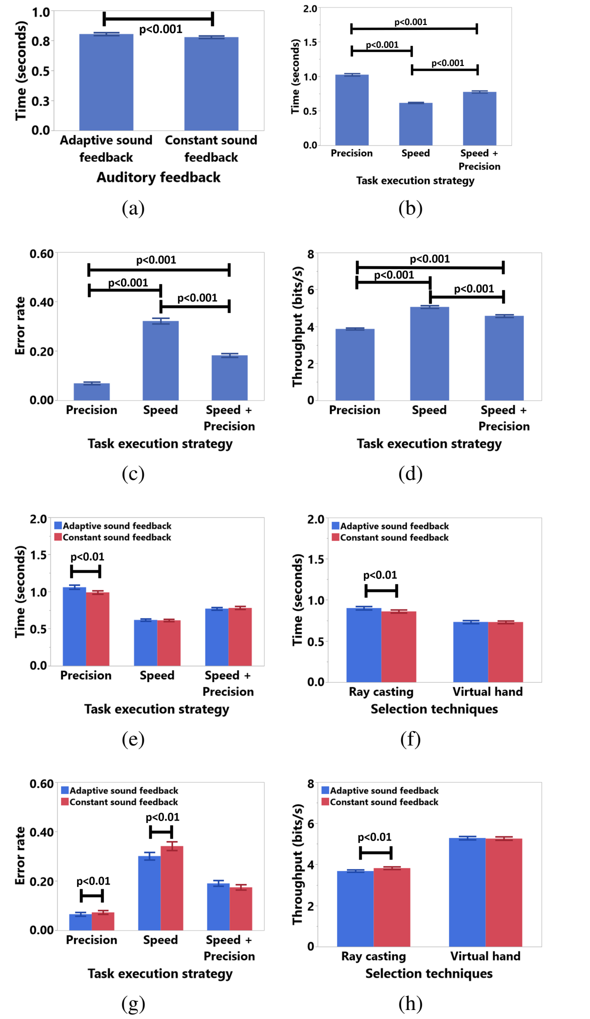


Figure 5: Time results for (a) auditory feedback and (b) task execution strategy; error rate results for (c) task execution strategy and effective throughput results for (d) task execution strategy. Time results for (e) interaction of auditory feedback and task execution strategy and (f) interaction of auditory feedback and selection technique; error rate results for (g) interaction of auditory feedback and task execution strategy; throughput results for (h) interaction of auditory feedback and selection technique.

when they executed the task "as precise as possible" compared to the virtual hand technique.

Error rate: The error rate was normally distributed with S = 1.19 and K = 1.07. According to the one-way RM ANOVA results shown in Fig. 5(c) and Table 4, subjects made more errors with the "as fast as possible" task execution strategy and fewer with "as precise as possible". We also found a significant two-way interaction between task execution strategy and task auditory feedback for error rate (F(2,34) = 3.123, $p < 0.05$, $\eta^2 = 0.155$ in Fig. 5(g)), where subjects made fewer errors with the "as fast as possible" and "as precise as possible" task execution strategies with adaptive feedback. Also, three-way RM ANOVA results showed that there is a significant interaction between auditory feedback, task execution strategy, and ID (F(6.432,109.35) = 5.037, $p < 0.01$, $\eta^2 = 0.229$). The results showed that the error rate of the subjects was lower with constant feedback compared to the adaptive feedback condition with the virtual hand interaction technique when A=25. Yet, four-way

ANOVA results identified no significant interaction between auditory feedback, task execution strategy, selection technique, and ID (F(7.341,124.798) = 1.246, p = 0.265, $\eta^2 = 0.068$).

Throughput: Throughput was normally distributed, S = 0.34, K = 1.07. The one-way RM ANOVA results in Fig. 5(d) and Table 4 show that the throughput performance of subjects was higher with the "as fast as possible" task execution strategy. Furthermore, we found a significant interaction between selection technique and auditory feedback for throughput (F(1,17) = 4.743, $p < 0.05$, $\eta^2 = 0.218$ in Fig. 5(h)), where the throughput performance of the participants decreased for ray casting selection with adaptive auditory feedback. Three-way RM ANOVA results showed that there is significant interaction between auditory feedback, task execution strategy, and ID (F(10, 170) = 2.068, $p < 0.05$, $\eta^2 = 0.108$). According to these results, participants achieved a higher throughput while they were instructed to follow "as fast as possible" strategy compared to "as precise/as fast and precise as possible" with adaptive and constant sound feedback for ID = 4.14, i.e., the most complex task. Four-way RM ANOVA results showed that there is a significant interaction between auditory feedback, task execution strategy, selection technique, and ID (F(10,170) = 3.777, $p < 0.001$, $\eta^2 = 0.182$). According to results, participants' throughput performance was higher with virtual hand interaction when they were asked to be "as precise as possible" with ID = 2.19.

4.5.1 Subjective results

As in the first study, we asked subjects about their insights and thoughts for the adaptive error feedback. Most, 13 out of 18, participants found the adaptive sound feedback useful and commented: "It was functional to focus [me] back on the task when I kept on making mistakes, but it also felt like I had a bit of room to make mistakes (the interval between medium to high pitch). Overall, it allowed me to know how bad I was doing and therefore try to focus more, but it also made the task more stressful as it increased the perceived pressure", "it may help increase the precision, but it gives me pressure and stress at the same time", "the inclining pitch when making mistakes made me want to slow down my speed and focus more on my precision", "Interesting, gives a sense of progression", "I find it helpful because it makes me more careful", "I liked it, warning without too annoying", "I believe it increased my awareness of errors", "Effective in terms of adaptive level of your errors", and "When the sound got high pitch I was trying to be more precise in order not to hear the sound anymore, although, I don't think I was driven by it as much as the fact that I was missing targets". The other 5 participants thought the adaptive feedback was "irritating", "aggressive" and "was good up to a point. I didn't like the highest pitch but before that it was good".

We also included a multiple-choice question to query for which task execution strategy participants thought that adaptive sound feedback increased their performance. Two participants chose "as fast as possible", 10 chose "as precise as possible", 14 chose "as fast and as precise as possible", and one chose "none" of the task execution strategies.

4.5.2 Fitts' Law Analysis

Linear regressions with Equation 1 show that user movement time can be modelled as MT = 0.02 + 0.25*ID, $R^2 = 0.96$ for the adaptive feedback and MT = 0.91 + 0.23*ID, $R^2 = 0.92$ for the constant frequency sound feedback. These results are shown in Fig. 6. Similarly, the results for "as fast as possible", "as precise as possible" and "as fast and as precise as possible" are shown in Fig. 6, with MT = 0.09 + 0.16*ID, $R^2 = 0.98$, MT = -0.04 + 0.34*ID, $R^2 = 0.93$ and MT = 0.05 + 0.22*ID, $R^2 = 0.0.94$, respectively.

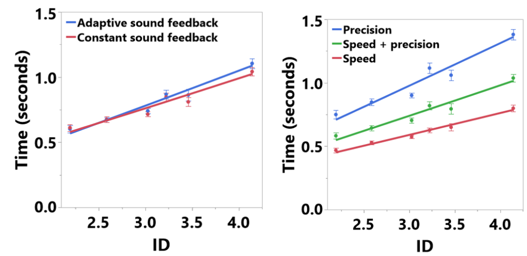


Figure 6: Fitts' law model for (a) sound feedback conditions and (b) task execution strategies.

4.6 Study 2 Discussion

In this second study, we investigated adaptive auditory error feedback, i.e., an error feedback mechanism that increases the sound frequency with each error that occurred. We also tested this method with three possible task execution strategies: "as fast/as precise/as fast and as precise as possible".

According to the results, adaptive auditory error feedback decreased task execution time overall. This result is in line with previous findings, where varying-frequency sounds were used to increase the motor performance [60]. However, we were not able to identify significant differences in overall error rate and throughput. For the "as precise as possible" task execution strategy, adaptive sound feedback decreased the error rate while increasing the task execution time, and throughput was also negatively affected. On the other hand, the error rate significantly decreased with the adaptive feedback for the "as fast as possible" task execution strategy. The majority (13 out of 18) participants also found adaptive sound feedback useful. Thus, we can say that our hypothesis **H2.1**, that *adaptive auditory error feedback increases user performance*, is partially supported. When we compare our results with those from previous work [4,39], one potential explanation for the different outcomes is that discreet (as in our work) and continuous sound feedback (as in [60]) affect the user performance in different ways. In this study, we increased the pitch of the discrete sound feedback when the participant made more errors in a round of trials, while previous work mapped the quality of the user movement to sound frequency [39,60]. These differences could explain the variation in results between our work and previous studies.

When we look at the interaction between selection technique and auditory feedback in terms of time and throughput, we can see that subjects were slower and their throughput decreased with the adaptive sound feedback. However, we did not observe any significant interaction in the error rate results between selection technique and auditory feedback. Thus, our results support hypothesis **H2.2**, that *Adaptive auditory error feedback decreases user performance when selecting targets with the ray casting selection technique*, which identifies a speed-precision trade-off. Previous work has shown that limitations of tracking systems, such as jitter [10–12, 67], can increase the error rate for the ray casting selection technique. Thus, we hypothesize that the subjects de-prioritized execution time, which decreased their throughput performance, to enable them to be more careful with the adaptive auditory error feedback so that they could avoid hearing the higher pitch.

Participants' positive comments on the adaptive sound feedback also validate the findings for the "as precise as possible" task execution strategy. The majority of the participants (13 out of 18) observed and commented on the decrease of their error rate with adaptive sound feedback, and 10 participants indicated that adaptive auditory feedback improved their performance for the "as precise as possible" task execution strategy. Yet, 14 out of 18 subjects thought that adaptive auditory feedback also improved their overall performance with the "as fast and as precise as possible" task execution strategy. Yet, we were unable to observe any significant performance improvement in terms of time, error rate, and throughput for this condition. Thus, we highlight that practitioners, developers, and researchers need to investigate this phenomenon further, since the trainee's motivation could decrease if their perception and objective results do not match.

5 General discussion

In this paper, we presented two user studies on the effects that auditory error feedback has on user performance in VR training systems and simulators that use Fitts' law tasks and throughput for performance assessment. In the first study, we investigated high-frequency auditory feedback and found a generally negative effect. Based on these results, we introduced adaptive auditory feedback in our second study to enable participants to change their user performance based on the task execution strategy.

Since VR HMDs and controllers exhibit great potential as training systems and simulators, we hope our findings will inform the creation of better training systems. We also believe that the results presented here can be used for 3D user interface design and interaction studies. Overall, we suggest the following:

Test sound feedback before starting trials: During the design of VR training systems or simulators, developers and practitioners should carefully choose an appropriate error sound, since its pitch can significantly change user performance [7]. If possible, the used frequency should be tested before a performance evaluation, even with different interaction techniques. As can be seen from the results of study 2, subjects' motor performance can vary with different task execution strategies with different auditory error feedback. This suggestion also extends to Fitts' law research, and we thus globally recommend reporting the sound frequency used in each user study, e.g., [1, 31].

Use a higher pitch to focus trainees' attention on their error rate: For situations where the trainee needs to decrease their error rate, high-pitch auditory feedback might help them shift their focus from execution time to error rate. However, such an approach can create a trade-off between the participant's time/throughput and their error rate. Moreover, since users did not prefer high-pitch auditory feedback, we do not recommend using this approach frequently.

Design the sound feedback to match the desired task execution strategy: To improve training outcomes, previous work suggested prioritizing participants' precision by asking them to execute the task "as precise as possible" [5]. In this work, we showed that adaptive error sound feedback can decrease participants' error rates during task execution, while it did not impact the throughput for the "as precise as possible" task execution strategy. Thus we believe that this strategy should be further investigated as a means to improve trainee motor performance.

In this work, we tested different auditory feedback and task execution strategies within an ISO 9241-411:2012 task [34]. Thus, all studies that use the Fitts' law task and throughput as a performance assessment criterion can use our results as a baseline. Previous work on VR simulators and training systems for motor performance training [8, 13] did not investigate auditory feedback. Also, we took care to choose our participants from the general public, not a specific group. Since novice trainees typically come from different backgrounds, the findings here could thus be applied to other VR training systems and simulators. We also believe that before generally advocating the use of VR training systems and simulators, we need to further investigate the overall effects of a VR system and its technical limitations on user performance before focusing on task-specific groups. This (more cautious) approach eliminates the unnecessary cost, time, and effort spent on creating VR training systems that do not perform well due to design issues.

We still acknowledge that the previous literature on skill transfer from VR systems to the real-world is inconclusive [29, 45, 72]. We see this as further evidence that we first need to understand the fundamental effects of VR-based training systems before designing them for specific tasks.

In this work, we only investigated the pitch of auditory error feedback. Based on the results from and suggestions included in previous work [4], we also only analyzed frequencies equal to or higher than a C4, i.e., 262 Hz, in our experiments. However, there are several other dimensions of sound feedback, including loudness, quality, timing, and localization [32, 43]. While our study aims to extend previous work in terms of the pitch for error feedback, the other dimensions should also be analyzed and investigated in terms of their usefulness to facilitate motor learning [61].

We focused on the effects of auditory error feedback and task execution strategies. Since ray casting and virtual hand selection techniques are widely used, we also investigated the interaction between these selection techniques, task execution strategies, and auditory error feedback. Still, the results presented here may vary for non-standard ray casting [47] and virtual hand techniques [46]. Also, our results could apply for positive auditory feedback [75], however experimenters should consider that it might take up to 8 minutes for subjects to familiarize themselves with a given form of sound feedback before they start improving their motor performance [60].

Due to COVID-19 related restrictions and following Steed et al.'s suggestions [63], we distributed our experiment and ran participants remotely. Participants used the headsets available in their homes with their own controllers, which increases the external validity of our work. Thus, data collection might have been affected by different computer specifications, resulting in a larger variation of latency relative to a lab study [67]. Still, we found no evidence that user system characteristics would have influenced our outcomes in a major way. For VR headsets without built-in headphones, participants used their own earphones or headsets. We only asked subjects to fix the volume at 40% on their computer. Distributing the experiment enabled us to report more externally valid results since the experiment involved several different headsetsThus, we can say that our results are not only limited to one particular VR system, but likely generalize to most common commercial VR headsets.

When we look at the 7-point Likert scale results, we see that participants did not report substantial mental or physical fatigue in both studies, with an average of 3.833 for both mental and physical fatigue in user study 1 and 4.5 and 4.11 for physical and mental fatigue in user study 2 (1-I feel extremely relaxed, 4-I feel normal, 7-I feel extremely tired).

6 Conclusion & Future Work

In this paper, we extended previous work on the effect of different forms of auditory error feedback on user performance and showed how different pitches impact user performance with different task execution strategies. According to the results, high-pitch auditory error feedback can be used to focus the attention of a user on their error rate, which, as a trade-off, decreases their throughput performance. Moreover, adaptive auditory feedback can be used to further decrease the error rate without affecting throughput for the "as precise as possible" task execution strategy, but (still) does not increase overall motor performance. We believe that our results inform better VR-based training systems and simulators to improve human eye-hand coordination motor performance.

In the future, we plan to apply the results of our findings to VR and AR simulators and training systems that are commercially available on the market. Furthermore, we want to investigate how other forms of auditory feedback, such as varying pitch with the distance of the "hit" to the target center or the length of the tone, affect individual learning in different tasks. We also want to apply our results to rehabilitation and medical research [3].

REFERENCES

[1] M. Akamatsu, I. MacKenzie, and T. Hasbroucq. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, 38(4):816—827, April 1995. doi: 10.1080/00140139508925152

[2] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke. Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 327–334, 2018. doi: 10.1109/VR.2018.8446317

[3] B. M. Asken, J. P. Mihalik, J. D. Schmidt, A. C. Littleton, K. M. Guskiewicz, and J. B. Hopfinger. Visual performance measures and functional implications in healthy participants: A sports concussion perspective. *Athletic Training and Sports Health Care*, 8(4):145–153, 2016.

[4] A. U. Batmaz. *Speed, precision and grip force analysis of human manual operations with and without direct visual input*. PhD thesis, Strasbourg, 2018.

[5] A. U. Batmaz, M. de Mathelin, and B. Dresp-Langley. Getting nowhere fast: trade-off between speed and precision in training to execute image-guided hand-tool movements. *BMC psychology*, 4(1):55, 2016.

[6] A. U. Batmaz, M. de Mathelin, and B. Dresp-Langley. Seeing virtual while acting real: Visual display and strategy effects on the time and precision of eye-hand coordination. *PloS one*, 12(8):e0183789, 2017.

[7] A. U. Batmaz, M. de Mathelin, and B. Dresp-Langley. Different sound pitch effects on motor performance of individuals in head-mounted virtual reality. In *Perception*, vol. 48, pp. 21–21, 2019.

[8] A. U. Batmaz, A. K. Mutasim, M. Malekmakan, E. Sadr, and W. Stuerzlinger. Touch the wall: Comparison of virtual and augmented reality with conventional 2d screen eye-hand coordination training systems. In *Conference on Virtual Reality and 3D User Interfaces*, VR '20, pp. 184–193. IEEE, Mar 2020. doi: 10.1109/VR46266.2020.1581205539914

[9] A. U. Batmaz, A. K. Mutasim, and W. Stuerzlinger. Precision vs. power grip: A comparison of pen grip styles for selection in virtual reality, Mar 2020. doi: 10.1109/VRW50115.2020.00012

[10] A. U. Batmaz, M. R. Seraji, J. Kneifel, and W. Stuerzlinger. No jitter please: Effects of rotational and positional jitter on 3D mid-air interaction. In *Future Technologies Conference*, vol. AISC 1289 of *FTC '20*. Springer, Nov 2020. doi: 10.1007/978-3-030-63089-8_52

[11] A. U. Batmaz and W. Stuerzlinger. The effect of rotational jitter on 3D pointing tasks. Extended Abstract, May 2019. doi: 10.1145/3290607.3312752

[12] A. U. Batmaz and W. Stuerzlinger. Effects of 3d rotational jitter and selection methods on 3d pointing tasks, Mar 2019. doi: 10.1109/VR.2019.8798038

[13] A. U. Batmaz, X. Sun, D. Taskiran, and W. Stuerzlinger. Hitting the Wall: Mid-air interaction for eye-hand coordination. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, pp. 30:1–30:5. ACM, New York, NY, USA, 2019. doi: 10.1145/3359996.3364249

[14] A. U. Batmaz, X. Sun, D. Taskiran, and W. Stuerzlinger. Eye-hand coordination training for sports with mid-air vr. In *26th Symposium on Virtual Reality Software and Technology*, VRST '20. ACM, Nov 2020. doi: 10.1145/3385956.3418971

[15] J. D. Bell and K. L. Macuga. Goal-directed aiming under restricted viewing conditions with confirmatory sensory feedback. *Human Movement Science*, 67:102515, 2019. doi: 10.1016/j.humov.2019.102515

[16] A. Bernardo. Virtual reality and simulation in neurosurgical training. *World neurosurgery*, 106:1015–1029, 2017.

[17] B. Bideau, R. Kulpa, N. Vignais, S. Brault, F. Multon, and C. Craig. Using virtual reality to analyze sports performance. *IEEE Computer Graphics and Applications*, 30(2):14–21, 2009.

[18] D. Bowman, C. Wingrave, J. Campbell, and V. Q. Ly. Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments, 2001.

[19] E. O. Boyer, B. M. Babayan, F. Bevilacqua, M. Noisternig, O. Warusfel, A. Roby-Brami, S. Hanneton, and I. Viaud-Delmon. From ear to hand: the role of the auditory-motor loop in pointing to an auditory source. *Frontiers in computational neuroscience*, 7:26, 2013.

[20] D. S. Brungart, B. D. Simpson, R. L. McKinley, A. J. Kordik, R. C. Dallman, and D. A. Ovenshire. The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources. Georgia Institute of Technology, 2004.

[21] K. J. Ciuffreda and B. Wang. *Vision Training and Sports*, pp. 407–433. Springer US, Boston, MA, 2004. doi: 10.1007/978-1-4419-8887-4_16

[22] R. M. Clifford, H. Khan, S. Hoermann, M. Billinghurst, and R. W. Lindeman. Development of a multi-sensory virtual reality training simulator for airborne firefighters supervising aerial wildfire suppression. In *2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good)*, pp. 1–5. IEEE, 2018.

[23] A. Covaci, C. Postelnicu, A. N. Panfir, and D. Talaba. A virtual reality simulator for basketball free-throw skills development. In *Doctoral Conference on Computing, Electrical and Industrial Systems*, pp. 105–112. Springer, 2012.

[24] J. Edworthy and N. Stanton. A user-centred approach to the design and evaluation of auditory warning signals: 1. methodology. *Ergonomics*, 38(11):2262–2280, 1995. PMID: 7498187. doi: 10.1080/00140139508925267

[25] H. Engelbrecht, R. Lindeman, and S. Hoermann. A swot analysis of the field of virtual reality for firefighter training. *Frontiers in Robotics and AI*, 6:101, 2019.

[26] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

[27] S. G. Fussell and D. Truong. Preliminary results of a study investigating aviation student's intentions to use virtual reality for flight training. *International Journal of Aviation, Aeronautics, and Aerospace*, 7(3):2, 2020.

[28] A. Godbout and J. E. Boyd. Rhythmic sonic feedback for speed skating by real-time movement synchronization. *International Journal of Computer Science in Sport (International Association of Computer Science in Sport)*, 11(3), 2012.

[29] R. Gray. Behavior of college baseball players in a virtual batting task. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5):1131, 2002.

[30] J. F. Hair Jr, W. C. Black, B. J. Babin, and R. E. Anderson. Multivariate data analysis, 2014.

[31] B. C. Hatfield, W. R. Wyatt, and J. B. Shea. Effects of auditory feedback on movement time in a fitts task. *Journal of Motor Behavior*, 42(5):289–293, 2010. PMID: 20826421. doi: 10.1080/00222895.2010.504759

[32] T. Hermann and A. Hunt. Guest editors' introduction: An introduction to interactive sonification. *IEEE multimedia*, 12(2):20–24, 2005.

[33] Y. Huang, L. Churches, and B. Reilly. A case study on virtual reality american football training. In *Proceedings of the 2015 Virtual Reality International Conference*, p. 6. ACM, 2015.

[34] ISO 9241-411:2012. Ergonomics of human-system interaction – part 411: Evaluation methods for the design of physical input devices. ISO, May 2012.

[35] G. Jakus, K. Stojmenova, S. Tomažič, and J. Sodnik. A system for efficient motor learning using multimodal augmented feedback. *Multimedia Tools and Applications*, 76(20):20409–20421, 2017.

[36] W. Kil, M.-J. Son, and J.-Y. Lee. Development of vr ship environment for the educational training of ship survey. *Journal of the Society of Naval Architects of Korea*, 55(4):361–369, 2018.

[37] N. H. Kim, M. Wininger, and W. Craelius. Training grip control with a fitts' paradigm: A pilot study in chronic stroke. *Journal of Hand Therapy*, 23(1):63–72, 2010.

[38] M. Kleiman-Weiner and J. Berger. The sound of one arm swinging: a model for multidimensional auditory display of physical motion. In *International Conference on Auditory Display*. Georgia Institute of Technology, 2006.

[39] N. Konttinen, K. Mononen, J. Viitasalo, and T. Mets. The effects of augmented auditory feedback on psychomotor skill learning in precision shooting. *Journal of Sport and Exercise Psychology*, 26(2):306–316, 2004.

[40] I. S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 1633–1636. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/

1357054.1357308
[41] I. S. MacKenzie and A. Oniszczak. A comparison of three selection techniques for touchpads. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 336–343. ACM Press/Addison-Wesley Publishing Co., 1998.
[42] P. Mallery and D. George. SPSS for windows step by step: a simple guide and reference. *Allyn, Bacon, Boston,*, 2003.
[43] G. Marentakis and S. A. Brewster. A comparison of feedback cues for enhancing pointing efficiency in interaction with spatial audio displays. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, MobileHCI '05, p. 55–62. Association for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1085777.1085787
[44] R. R. McKnight, C. A. Pean, J. S. Buck, J. S. Hwang, J. R. Hsu, and S. N. Pierrie. Virtual reality and augmented reality—translating surgical training into surgical technique. *Current Reviews in Musculoskeletal Medicine*, pp. 1–12, 2020.
[45] H. C. Miles, S. R. Pop, S. J. Watt, G. P. Lawrence, and N. W. John. A review of virtual environments for training in ball sports. *Computers & Graphics*, 36(6):714 – 726, 2012. doi: 10.1016/j.cag.2012.04.007
[46] A. Moore, M. Kodeih, A. Singhania, A. Wu, T. Bashir, and R. McMahan. The importance of intersection disambiguation for virtual hand techniques. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 310–317. IEEE, 2019.
[47] A. G. Moore, J. G. Hatch, S. Kuehl, and R. P. McMahan. Vote: A raycasting study of vote-oriented technique enhancements. *International Journal of Human-Computer Studies*, 120:36–48, 2018.
[48] A. K. Mutasim, W. Stuerzlinger, and A. U. Batmaz. Gaze tracking for eye-hand coordination training systems in virtual reality. Extended Abstract, Apr 2020. doi: 10.1145/3334480.3382924
[49] D. Narciso, M. Melo, J. V. Raposo, J. Cunha, and M. Bessa. Virtual reality in training: an experimental study with firefighters. *Multimedia Tools and Applications*, 79(9):6227–6245, 2020.
[50] W. T. Nelson, L. J. Hettinger, J. A. Cunningham, B. J. Brickman, M. W. Haas, and R. L. McKinley. Effects of localized auditory information on visual target detection performance using a helmet-mounted display. *Human factors*, 40(3):452–460, 1998.
[51] J. M. Obrenović, V. D. Nešić, and M. Nešić. The reaction time in relation to the modality of stimulation. *Facta universitatis-series: Physical Education*, 1(3):85–90, 1996.
[52] G. Papagiannakis, P. Trahanias, E. Kenanidis, and E. Tsiridis. Psychomotor surgical training in virtual reality. In *The Adult Hip-Master Case Series and Techniques*, pp. 827–830. Springer, 2018.
[53] R. D. Patterson. Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 327(1241):485–492, 1990.
[54] D. R. Perrott, J. Cisneros, R. L. McKinley, and W. R. D'Angelo. Aurally aided visual search under virtual and free-field listening conditions. *Human factors*, 38(4):702–715, 1996.
[55] D. Pham and W. Stuerzlinger. Is the pen mightier than the controller? a comparison of input devices for selection in virtual and augmented reality. In *25th Symposium on Virtual Reality Software and Technology*, VRST '19. ACM, Nov 2019. doi: 10.1145/3359996.3364264
[56] A. C. Pierno, A. Caria, and U. Castiello. Comparing effects of 2-d and 3-d visual cues during aurally aided target acquisition. *Human factors*, 46(4):728–737, 2004.
[57] L. Rosenberg and S. Brave. Using force feedback to enhance human performance in graphical user interfaces. In *Conference Companion on Human Factors in Computing Systems*, CHI '96, pp. 291–292. ACM, New York, NY, USA, 1996. doi: 10.1145/257089.257327
[58] E. J. Scheme and K. B. Englehart. Validation of a selective ensemble-based classification scheme for myoelectric control using a three-dimensional fitts' law test. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(4):616–623, 2013. doi: 10.1109/TNSRE.2012.2226189
[59] J. Shepherd, L. Carter, G. Pepping, and L. Potter. Towards an operational framework for designing training based sports virtual reality performance simulators. In *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 2, p. 214, 2018.
[60] R. Sigrist. Visual and auditory augmented concurrent feedback in a complex motor task. *Presence*, 20(1):15–32, 2011. doi: 10.1162/pres_a_00032
[61] R. Sigrist, G. Rauter, R. Riener, and P. Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review*, 20(1):21–53, 2013.
[62] E. M. Staurset and E. Prasolova-Førland. Creating a smart virtual reality simulator for sports training and education. In *Smart Education and e-Learning 2016*, pp. 423–433. Springer, 2016.
[63] A. Steed, F. R. Ortega, A. S. Williams, E. Kruijff, W. Stuerzlinger, A. U. Batmaz, A. S. Won, E. S. Rosenberg, A. L. Simeone, and A. Hayes. Evaluating immersive experiences during covid-19 and beyond. *Interactions*, 27(4):62–67, Jul 2020. doi: 10.1145/3406098
[64] J. Sterkenburg, S. Landry, and M. Jeon. Influences of visual and auditory displays on aimed movements using air gesture controls. In *International Conference on Auditory Display*. Georgia Institute of Technology, 2017.
[65] J. Sterkenburg, S. Landry, and M. Jeon. Design and evaluation of auditory-supported air gesture controls in vehicles. *Journal on Multimodal User Interfaces*, 13(2):55–70, 2019.
[66] Y. Tan, C. Niu, and J. Zhang. Head-mounted, display-based immersive virtual reality marine-engine training system: A fully immersive and interactive virtual reality environment. *IEEE Systems, Man, and Cybernetics Magazine*, 6(1):46–51, 2020.
[67] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie. Effects of tracking technology, latency, and spatial jitter on object movement. In *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on*, pp. 43–50. IEEE, 2009.
[68] R. J. Teather and W. Stuerzlinger. Pointing at 3D targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 87–94, March 2011. doi: 10.1109/3DUI.2011.5759222
[69] R. J. Teather and W. Stuerzlinger. Visual aids in 3D point selection experiments. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pp. 127–136. ACM, 2014.
[70] K. Valentino, K. Christian, and E. Joelianto. Virtual reality flight simulator. *Internetworking Indonesia Journal*, 9(1):21–25, 2017.
[71] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 143–146. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963
[72] F. T. J. M. Zaal and R. J. Bootsma. Virtual reality as a tool for the study of perception-action: The case of running to catch fly balls. *Presence: Teleoperators and Virtual Environments*, 20(1):93–103, 2011.
[73] M. A. Zahariev and C. L. MacKenzie. Auditory, graphical and haptic contact cues for a reach, grasp, and place task in an augmented environment. In *Proceedings of the 5th International Conference on Multimodal Interfaces*, ICMI '03, p. 273–276. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/958432.958481
[74] M. A. Zahariev and C. L. Mackenzie. Auditory contact cues improve performance when grasping augmented and virtual objects with a tool. *Experimental brain research*, 186(4):619–627, 2008.
[75] D. Zielasko, N. Neha, B. Weyers, and T. W. Kuhlen. A reliable nonverbal vocal input metaphor for clicking. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 40–49. IEEE, 2017.