Performance Is Not Everything: Audio Feedback Preferred Over Visual Feedback for Grasping Task in Virtual Reality

Ryan Canales*
Clemson University
rcanale@clemson.edu

Sophie Jörg
Clemson University
sjoerg@clemson.edu

ABSTRACT

In this work, we investigate the influence that audio and visual feedback have on a manipulation task in virtual reality (VR). Without the tactile feedback of a controller, grasping virtual objects using one's hands can result in slower interactions because it may be unclear to the user that a grasp has occurred. Providing alternative feedback, such as visual or audio cues, may lead to faster and more precise interactions, but might also affect user preference and perceived ownership of the virtual hands. In this study, we test four feedback conditions for virtual grasping. Three of the conditions provide feedback for when a grasp or release occurs, either visual, audio, or both, and one provides no feedback for these occurrences. We analyze the effect each feedback condition has on interaction performance, measure their effect on the perceived ownership of the virtual hands, and gauge user preference. In an experiment, users perform a pick-and-place task with each feedback condition. We found that audio feedback for grasping is preferred over visual feedback even though it seems to decrease grasping performance, and found that there were little to no differences in ownership between our conditions.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; **Visualization**.

KEYWORDS

virtual grasping, visual feedback, audio feedback, multimodal feedback, virtual hand interaction, virtual reality, avatar ownership, virtual characters

ACM Reference Format:
Ryan Canales and Sophie Jörg. 2020. Performance Is Not Everything: Audio Feedback Preferred Over Visual Feedback for Grasping Task in Virtual Reality. In *Motion, Interaction and Games (MIG '20), October 16–18, 2020, Virtual Event, SC, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3424636.3426897

*This is the corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MIG '20, October 16–18, 2020, Virtual Event, SC, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8171-0/20/10…$15.00
https://doi.org/10.1145/3424636.3426897

1 INTRODUCTION

Virtual interactions have been studied extensively, with various methods developed and tested for selecting and manipulating objects in virtual reality (VR). As hand tracking technology makes its way into commercial VR devices, research aimed at improving virtual interactions without controllers becomes increasingly relevant. Indeed, there have been several studies on the effects of visual feedback on virtual interactions [Canales et al. 2019; Prachyabrued and Borst 2014; Vosinakis and Koutsabasis 2018].

Studies incorporating multisensory feedback show that combining multiple feedback techniques can lead to better task performance [Cooper et al. 2018; Zhang et al. 2006]. However, the effects of multisensory feedback on virtual grasping when the real hand motions are tracked to control the virtual hands has not been investigated yet.

In this study, we focus on audio feedback in addition to the visual feedback that has been examined in previous studies. Audio feedback is widely used in video games to confirm that a requested action has been performed [Ekman 2013; Parker and Heerma 2008]. When designing sound for feedback, it is important to consider the characteristics of the audio, such as volume, pitch, and length. Inappropriately designed audio feedback might lead to misinterpretation, reducing its effectiveness [Sigrist et al. 2012]. Short and discrete audio clips ("earcons") have been shown to be effective at communicating that an interaction occurred in a user interface [Blattner et al. 1989]. In our study, we use earcons to assist the user in knowing when a grasp action or a release action has been detected. We investigate the effect that audio and visual feedback have on the users' performance, perceived ownership, and preferences.

2 RELATED WORK

As context for this study, we present previous work on visual and audio feedback for interaction in virtual environments and we provide a brief background on virtual hand ownership.

2.1 Visual Feedback for Virtual Grasping

When grasping an object in the real world, we receive a combination of cues that together indicate we have gripped the object. These cues include visual feedback, haptic feedback, and muscle tension. In VR, some of these cues are not present and the virtual hand might be able to penetrate the virtual object, making the interaction unrealistic. Additional feedback can be used to alleviate that impression. Visual feedback for grasping is a promising and simple technique that can be applied to many VR applications as no additional hardware is required.

Previous studies have shown that visual feedback can improve grasping performance with virtual objects. Prachybrued and Borst

[2014] found that visual feedback for the penetration depth of the tracked fingers improves targeted release speed and accuracy. Vosinakis and Koutsabasis [2018] found that visual feedback improved grasping and release performance in both desktop and immersive VR environments. Geiger et al. [2018] used visual feedback to guide users into gripping a virtual object in a specified way, finding that for complex grip types, such as whole hand grasping, visual feedback helped significantly. Cortes et al. [2018] found that rendering virtual shadows can affect 3D object placement.

2.2 Multimodal Feedback in VR

In addition to visual feedback, there has been extensive work on combining feedback techniques, such as audio, visual, and tactile, to improve virtual interactions. Multiple feedback modalities (multimodal feedback) for 2D user interfaces have been shown to improve target selection and acquisition [Akamatsu et al. 1995; Cockburn and Brewster 2005]. For 3D virtual environments, audio and visual feedback can improve presence [Biocca et al. 2001] and they show promise for use in rehabilitation in virtual environments [Wellner et al. 2008]. Combining audio and visual feedback tends to be subjectively preferred for 3D selection in virtual reality [Vanacken et al. 2009] and can improve selection performance by providing proximity based cues [Ariza et al. 2018]. Furthermore, providing multimodal feedback has been shown to improve task performance [Cooper et al. 2018] and reduce perceptual load [Santangelo and Spence 2007]. Cooper et al. [2018] studied the effect that audio, visual, and tactile (vibration) feedback have on task performance and subjective preference. They found that task performance is improved significantly when audio or tactile feedback is provided.

In our work, we test grasping and release performance when the user is provided with either no feedback, audio feedback, visual feedback, or both audio and visual feedback for grasping. In addition to examining the performance, we test for effects of each condition on virtual hand ownership and user preference.

2.3 Virtual Hand Ownership

The illusion that the virtual hands are actually one's real hands is referred to as the *virtual hand illusion*. When this occurs, the user can "feel" sensation on the virtual hands, as if it were their own hands contacting a virtual object. Longo et al. [2008] identify three key factors that affect the sense of embodiment (SoE): body ownership, sense of location, and sense of agency. Body ownership refers to the feeling that the virtual body is one's own body. Location is the feeling that one's body and the virtual body occupy the same space. Agency is the feeling that the virtual body is controlled by the user, i.e. its movements correspond to the user's intent.

Findings from previous studies indicate that the visual representation of the virtual hands and ownership are correlated. For example, Yuan and Steed [2010] find that ownership is significantly higher for a realistic virtual hand than for an arrow representing the hand. Lin and Jörg [2016] find that although ownership is strongest with a realistic hand, it can still be established, to a lesser degree, over a wooden block representation. Results from a study by Argelaguet et al. [2016], from which our experimental design was inspired, show that the virtual hand representation affects ownership and task performance. In addition to the representation of the virtual hands, the behavior of the virtual hands as they grasp a virtual object can also affect ownership. Canales et al. [2019] find that ownership tends to be higher when the virtual hand is kept from clipping through the object being grasped, replicating a realistic interaction.

In general, previous studies show that ownership can occur over any visual representation of the virtual hands, though it is stronger with anthropomorphic representations and strongest with realistic representations, and a realistic visualization when grasping might lead to a higher sensation of ownership in a similar way. Results from a study by Lugrin et al. suggest that audio feedback has no effect on body ownership [2016]. However, no previous work has examined if adding sound during a virtual interaction has an effect on ownership, which we investigate in this work.

3 METHOD

3.1 Design

This was a within-subjects experiment. Each participant performed a pick-and-place task with each of the four grasping feedback conditions in random order. The audio volume, avatar proportions, and virtual hand size were adjusted for each participant.

3.2 Participants

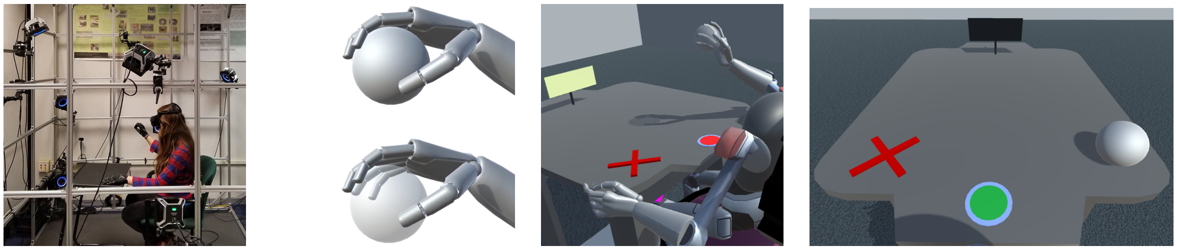
A total of 32 participants took part in our study. We excluded one participant's data from analysis due to tracking errors. The remaining data from 31 participants (17F, 14M), aged 18 to 42 years old (μ = 25), was used for our analysis. We obtained signed consent from all participants and pre-screened each participant for cybersickness. The study took participants between 25 and 35 minutes to complete. After completing the study, participants were debriefed and received a $5 gift card.

3.3 Apparatus

An Oculus Rift CV1 head mounted display (HMD) with the included headphones was used. Our HMD was connected to a computer with an Intel Core i7-7700 CPU and Nvidia GTX 1080 GPU. The experiment was developed in Unity. We used a high fidelity, marker-based hand tracking system with 16 cameras at 120fps (Figure 1(a)) following the method described by Han et al. [2018] to solve for the hand motion in real time.

3.4 Grasping Feedback Conditions

For our baseline condition without any feedback (NF), we visualize a hand that does not intersect with the ball geometry when grasping (see Figure 1(b)) as such a visualization is preferred in grasping tasks to simply displaying the motion of the real hands [Canales et al. 2019; Prachyabrued and Borst 2014]. A grasp is detected when at least two fingertips and the tip of the thumb collide with the ball. Then, the ball is attached to the hand. A release is confirmed when all fingertips have left the ball volume. Once a grasp has been detected, if a finger joint collides with the ball, then all the joints above the ones colliding with the ball in the hierarchy are locked and the joints below the colliding joints can rotate up to 60 degrees towards the ball relative to their parent or until they too make contact with the ball. With this method, which was developed



(a) (b) (c) (d)

Figure 1: (a) 16 camera hand motion capture system. (b) Top: NF and AF. Bottom: VF and AVF. (c) The seated robot avatar. (d) Virtual desk setup for pick and place task.

and adjusted in a series of tests, the fingers do not intersect with the ball's geometry and reach a natural grasp pose. To avoid that the fingers enter the ball geometry if the participants grasp very quickly, we slow down the joint's rotation when they get very close to the surface of the ball, which was not noticed by our test subjects.

Our four conditions are as follows:

(1) *NF:* No Feedback; there is no indication that the ball is being grasped or released.
(2) *AF:* Audio Feedback; sound is played when the ball is grasped and when it is released.
(3) *VF:* Visual Feedback; the ball becomes semitransparent upon being grasped and opaque when released (Figure 1(c)).
(4) *AVF:* Audio & Visual Feedback; the ball becomes semitransparent and a sound plays when it is grasped. Upon release, the ball becomes opaque and the release sound plays.

Note that three of the conditions, AF, VF, and AVF, provide feedback for when the ball is grasped and released. AF and VF provide only audio or visual feedback, allowing us to analyze the effects of each type of feedback independently. We choose object transparency for the visual feedback based on the preference results in Canales et al. [2019], in which object transparency was the most preferred visual feedback technique. The sound we use to indicate a grasp is a short knocking sound pitched at 525Hz and is about 30 milliseconds in length. The same sound but pitched 3 semitones ($\approx$ 100Hz) higher is used to indicate a release. There are no other sounds during the experiment.

3.5 Procedure

Calibration: All participants were represented in the virtual environment by the same sitting robot avatar (Figure 1(c)), which provides a neutral virtual representation and allows for detailed finger movement, but does not allow for flexing of the palms or skin. To change the height of the avatar, participants placed their hands on their lap, then the experimenter adjusted the height so that the robot hands rested on its lap. There are six sizes of the virtual hands, each of which is pre-computed based on the six glove sizes from which participants can choose from. The participants are asked to select the tightest fitting glove that permits comfortable hand movements, and the virtual hand size is matched to the selected glove size. Finally, to adjust the volume of the headset, the sound used when the ball is grasped is played three times and the participants are asked if they can hear the sound clearly.

Pre-Experiment Ranking: After calibration, videos of each visualization condition, with sound if present, are shown to the participants. Participants are asked to rank the conditions from 1 to 4, with 1 representing the most preferred and 4 the least, before using the conditions in practice. Then, a random feedback condition is activated and the participants practice grasping and releasing the ball until they feel reasonably skilled at doing so.

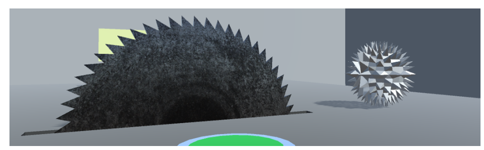


Figure 2: The spiky ball randomly appears twice during block two. The spinning saw appears once after block two.

Pick and Place Task: The main experimental task, called the *pick-and-place* task, is performed by the participant for each feedback condition. The task was implemented following elements from Argelaguet et al. [2016] and Canales et al. [2019] and is performed as follows: First, the participant "presses" a green button on the virtual desk. Then, they pick up the ball and move it onto a target, represented by a red X. The button, ball, and target are shown in Figure 1(d). Finally, they press the button again, which is now red, and perform the same action repeatedly. They were instructed to perform the task as smoothly and naturally as they could and also to attempt to center the ball over the target. Participants practiced this task eight times, twice with each condition in random order, before starting the main experiment task.

The *pick-and-place* task during the experiment is divided into three blocks. During the first block, the participants perform the task ten times with the current condition. The second block is similar to the first block; the participant performs the task ten times, but during two of the ten repetitions (selected randomly), the ball is replaced with a spiky ball after the green button is pressed. Finally, during the short third block, after the participant has pressed the button, a virtual spinning saw emerges from the table and the task is performed only once. There is no distinction between the blocks to the user. The task is performed with the user's dominant hand a total of 22 times for each feedback condition. The spiky ball and the spinning saw are used as threats to measure ownership based on Argelaguet et al. [2016] and Canales et al. [2019], see Figure 2.

Questionnaire: After the third block, a questionnaire inspired by Argelaguet et al. [2016] and Canales et al. [2019], which are based on the standard Botvinick and Cohen 9-question survey [1998], is displayed. For each statement, participants chose a rating on a seven-point Likert scale ranging from 1 for "strongly disagree" to 7 for "strongly agree."

We asked seven questions related to ownership: O1: "I felt as if the virtual hands were part of my body."; O2: "It sometimes seemed like my own hands came into contact with the virtual object."; O3: "I thought that the virtual hands could be harmed by a virtual danger."; O4: "I felt that my real body was endangered during the experiment."; O5: "I felt that my real hand was endangered during the experiment."; O6: "I anticipated feeling pain from the spinning saw on the screen."; and O7: "I tried to avoid the virtual saw while performing the task." We furthermore asked two questions related to agency for completeness: A1: "I felt as if I can control movements of the virtual hands." and A2: "I felt as if the virtual hands moved just like I wanted them to, as if they were obeying my will."

After answering all questions, participants were asked to rank the four conditions again in order of preference. At the end of the experiment, they were asked for anything they specifically liked or disliked in the conditions as well as for any further feedback.

3.6 Hypotheses

Based on related work, we formed the following hypotheses:

- **H1:** Feedback (AF, VF, AVF) of the grasp and release states will improve grasping and release performance.
- **H2:** Audio feedback (AF) will not affect ownership, however visual feedback (VF) could reduce ownership.
- **H3:** We expect that feedback (VF, AF, AVF) will be preferred over no feedback (NF), and AVF will be the most preferred.

We expect **H1** due to several previous studies [Cooper et al. 2018; Prachyabrued and Borst 2014; Vosinakis and Koutsabasis 2018] indicating that feedback improves task performance in VR. The first part of **H2** is based on results from Lugrin et al.'s study [2016], in which they found no effect of audio on body ownership. The second part of **H2** follows findings by Canales et al. [2019] indicating that less realistic interactions result in a reduced sense of ownership. Therefore, we think that if the level of ownership were to be affected by visual feedback (VF), it would be reduced. We expect **H3** because it follows previous evaluations of feedback on preference [Vanacken et al. 2009; Zhang et al. 2006], in which feedback is preferred and multimodal feedback is the most preferred.

4 RESULTS

4.1 Performance

Our analysis of performance uses the measurements from the first block of the pick-and-place trial for each condition and entails grasp and release performance. Before performing our analysis, we discarded measurements from trials in which errors occurred, such as grasps not being detected or markers briefly not being tracked. 98.4% (1220/1240) of the measurements from the first block are retained. One-way, repeated measures ANOVAs with Greenhouse-Geisser sphericity corrections are used to test for significant main effects of the feedback condition. If significant ($p < 0.05$) or close

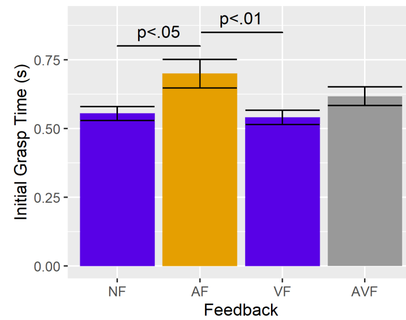


Figure 3: Mean initial grasp time. This time interval was significantly longer for AF than for NF and VF.

to significant ($p \approx 0.1$) effects were found, we performed a post-hoc pairwise t-test with Bonferroni adjusted p-values. Generalized eta-squared (η^2_G) is reported for effect sizes.

Grasp Performance: To analyze grasp performance, we use two measures. The first measure is the initial grasp time, which is the time taken from when the center of the virtual hand (close to the base of the middle finger) is less than 20cm from the center of the ball (r=5.48cm) to when the first grasp is detected. The second measure is the grasping time interval, which is measured from when the middle of the virtual hand is less than 20cm away from the center of the ball to the moment when the ball has been moved 20cm from its initial position. The grasping time interval includes the time it takes the user to realize that a grasp has occurred and also covers cases in which multiple grasp attempts occurred.

The ANOVA results evaluating an effect of feedback condition for the initial grasp time failed to reach significance ($F(3, 90) = 2.1, p \approx 0.10, \eta^2_G = 0.04$, Figure 3). However, our post-hoc test showed significant differences between AF and NF ($p \approx 0.03$) and AF and VF ($p < 0.01$), with participants taking longer to grasp when there was audio feedback. No significant effect of feedback condition was found for the grasping time interval ($F(3, 90) = 1.35, p = 0.26$).

Release Performance: Release performance is gauged by the target placement accuracy and a release time interval. The placement accuracy is the horizontal Euclidean distance from the center of the ball to the center of the target upon initial release. The release time interval is the time taken from when the center of the ball is less than 20cm from the target and when the center of the virtual hand has moved 20cm away from the target after placement.

We found no statistically significant main effect of condition on placement accuracy ($F(3, 90) = 1.67, p = 0.18$) or release time interval ($F(3, 90) = 0.54, p = 0.66$).

4.2 Ownership

To test for ownership over the virtual hands, we use measurements from the second and third blocks in which the virtual threats (the spiky ball and the spinning saw shown in Figure 2) appear. We furthermore analyze the responses to the questionnaire. Prior to analysis, we remove measurements from erroneous trials. 99.7% (1236/1240) of measurements from the second block are retained.

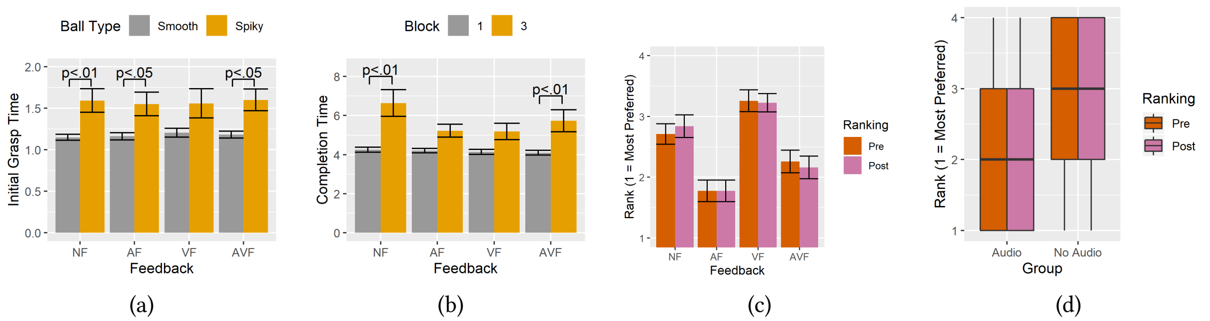


Figure 4: Measures related to ownership: (a) Block 2 and (b) Block 3. (c) Mean rank for each condition pre and post experiment. (d) Boxplot of averaged rank for feedback with audio (AF, AVF) and without audio (VF, NF) pre and post experiment.

No measurements are removed from the third block. Generalized eta-squared (η_G^2) is reported for effect sizes.

Virtual Threats: For the second block, the time between when the user presses the button and the time at which they have grasped the ball is analyzed. A two-way repeated measures ANOVA indicates that the spiky ball significantly increases the time taken to grasp the ball ($F(1, 30) = 28.17, p \ll 0.001, \eta_G^2 = 0.07$, Figure 4(a)). No significant interaction effect between the ball type and feedback condition was found ($F(3, 90) = 0.06, p = 0.98$).

A post-hoc least-squared means pairwise comparison showed that the spiky ball significantly increased the time to grasp for nearly all conditions ($p < 0.05$ for AF and AVF, $p < 0.01$ for NF, and $p \approx 0.07$ for VF).

For block three (spinning saw), we analyze the overall time to complete the task when compared to block 1 (see Figure 4(b)). A two-way repeated measures ANOVA found that the saw had a significant effect on completion time ($F(1, 30) = 26.24, p \ll 0.01, \eta_G^2 = 0.10$). A post-hoc least-square means analysis shows that the completion time is significantly higher for NF and AVF ($p < 0.01$). No significant interaction effect between feedback condition and the presence of the virtual saw was found ($F(3, 90) = 2.38, p = 0.075$).

Questionnaire: A Friedman rank test found no significant differences in responses between conditions for ownership and agency.

4.3 Preference

The mean rankings shown in Figure 4(c) indicate that, on average, AF was the most preferred, followed by AVF, NF, then VF. Two-way Chi-square tests for independence were used to further analyze the data in Table 1. In the pre-experiment ranking, VF was ranked significantly lower than AF ($\chi^2 = 22.9, p \ll 0.001$), NF ($\chi^2 = 9.69, p < 0.05$) and AVF ($\chi^2 = 14.11, p < 0.005$). NF was also ranked significantly lower than AF ($\chi^2 = 18.3, p < 0.005$). In the post-experiment ranking, VF was ranked significantly lower than both AF ($\chi^2 = 26.1, p \ll 0.001$) and AVF ($\chi^2 = 15.48, p < 0.005$). NF was again ranked significantly lower than AF ($\chi^2 = 14.9, p < 0.005$).

Table 1: Pre and post experiment ranking frequencies.

	Pre-experiment				Post-experiment			
Rank:	1st	2nd	3rd	4th	1st	2nd	3rd	4th
NF	2	13	8	8	4	7	10	10
AF	17	6	6	2	16	9	3	3
VF	3	3	8	17	1	5	11	14
AVF	9	9	9	4	10	10	7	4

5 DISCUSSION AND LIMITATIONS

Adding feedback for grasping did not improve grasping performance according to our measures. In fact, audio feedback (AF) seemed to slow down the speed at which users grasped the ball, and this same effect tended to occur when combining audio and visual feedback (AVF). This result is surprising and does not support hypothesis H1. One reason for this effect could be that users might have slowed down while waiting to hear the sound.

We found little to no indication that the level of perceived ownership was affected by any of the four conditions presented in this paper. The responses to our questionnaire showed that the type of grasping visualization did not affect ownership levels. Results from the second and third blocks point to similar levels of ownership for each condition. The spiky ball had the strongest effect in the NF condition and the spinning saw had a significant effect on completion time for NF and AVF. These results indicate that ownership might be higher when no feedback is present or when audio and visual feedback are both present. The first result could be based on the higher realism of the NF condition, though further experiments are necessary to provide this conclusion with confidence. One reason for our limited results regarding ownership might be the use of a robot avatar, as opposed to a realistic virtual human, which may have reduced the maximum level of ownership that can be established over the virtual hands in the first place. Though we did not find clear evidence that audio affects ownership, we cannot rule out the possibility. Audio has been linked to increased presence and immersion [Parker and Heerma 2008], which, under the correct circumstances, might elicit stronger levels of ownership. Maybe the ownership levels would be influenced by audio feedback if the experimental task were more engaging and if a realistic avatar were used. We only have weak evidence to confirm the second part of hypothesis H2, that visual feedback reduced ownership, based on the results from Block 3 compared to the NF condition.

Our rankings showed that audio feedback (AF) was preferred over both conditions without audio (NF, VF). While this is in line with our expectations and confirms parts of our third hypothesis, one still needs to mention that audio feedback did not improve performance (on the contrary, it reduced performance) and did not increase ownership. So why was it preferred? Comments from participants suggest that the audio feedback seemed to help, so maybe they were more confident or content when performing the task. The participants might have felt that audio made grasping easier for them.

Interestingly, AVF was not the most preferred condition and VF was the least preferred. These results suggest that the visual feedback in general was not liked, which is unexpected based on results from previous studies in which having visual or audio feedback was preferred over no feedback and multimodal feedback was the most preferred [Vanacken et al. 2009; Zhang et al. 2006]. Several users reported that they disliked the visual feedback because it was not realistic. It is possible that a different type of visual feedback such as a change of the color of the ball or hand as tested by Prachybrued and Borst [2014], would lead to different results where users would prefer conditions with further visual feedback. While a suddenly transparent object is not realistic, neither is an earcon when grasping or releasing an object. It could be that visual realism is perceived to be more important when it comes to preference or ownership than audio realism.

6 CONCLUSION AND FUTURE WORK

Our main findings are: (1) Our conditions with audio feedback were preferred over those without audio feedback: AF was the most preferred condition on average in both the pre and post experiment rankings, followed by AVF, NF, and VF. (2) Adding feedback for grasping and releasing did not improve performance. Interestingly, in our experiment audio feedback slows down grasp time.

We therefore recommend to add audio feedback to grasping interactions with tracked hands in virtual environments when user preference is the most important criteria. Although our results indicate that it might be better to avoid audio feedback if performance is crucial, further experimentation is needed to make this recommendation with confidence. We cannot make any strong conclusions about ownership based on the results in our study.

Future work could include investigating the interaction between other sensory modalities, especially haptic feedback, when grasping with tracked hands in VR. Additionally, this type of study could be reproduced using a more realistic avatar, different sound and visual designs for feedback, and more engaging tasks.

ACKNOWLEDGMENTS

The authors wish to thank Alexandra Adkins for helping to recruit participants. This work was supported in part by the National Science Foundation under Grant Number IIS-1652210.

REFERENCES

Motoyuki Akamatsu, I. Scott Mackenzie, and Thierry Hasbroucq. 1995. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics* 38, 4 (1995), 816–827. https://doi.org/10.1080/00140139508925152

Ferran Argelaguet, Ludovic Hoyet, Michaël Trico, and Anatole Lecuyer. 2016. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality*. 3–10. https://doi.org/10.1109/VR.2016.7504682

Oscar Ariza, Gerd Bruder, Nicholas Katzakis, and Frank Steinicke. 2018. Analysis of Proximity-Based Multimodal Feedback for 3D Selection in Immersive Virtual Environments. *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (March 2018), 327–334. https://doi.org/10.1109/VR.2018.8446317

Frank Biocca, Jin Kim, and Yung Kyun Choi. 2001. Visual Touch in Virtual Environments: An Exploratory Study of Presence, Multimodal Interfaces, and Cross-Modal Sensory Illusions. *Presence* 10 (06 2001), 247–265. https://doi.org/10.1162/105474601300343595

Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and Icons: Their Structure and Common Design Principles (Abstract Only). *SIGCHI Bull.* 21, 1 (Aug. 1989), 123–124. https://doi.org/10.1145/67880.1046599

Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands 'feel' touch that eyes see. *Nature* 391 (Feb. 1998), 756. https://doi.org/10.1038/35784

Ryan Canales, Aline Normoyle, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. Virtual Grasping Feedback and Virtual Hand Ownership. In *ACM Symposium on Applied Perception 2019 (SAP '19)*. ACM, Article 4, 9 pages. https://doi.org/10.1145/3343036.3343132

Andy Cockburn and Stephen Brewster. 2005. Multimodal feedback for the acquisition of small targets. *Ergonomics* 48, 9 (2005), 1129–1150. https://doi.org/10.1080/00140130500197260

Natalia Cooper, Ferdinando Milella, Carlo Pinto, Iain Cant, Mark White, and Georg Meyer. 2018. The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLOS ONE* 13, 2 (02 2018), 1–25. https://doi.org/10.1371/journal.pone.0191846

Guillaume Cortes, Ferran Argelaguet, Eric Marchand, and Anatole Lécuyer. 2018. Virtual Shadows for Real Humans in a CAVE: Influence on Virtual Embodiment and 3D Interaction. In *Proceedings of the 15th ACM Symposium on Applied Perception (SAP '18)*. ACM, Article 1, 8 pages. https://doi.org/10.1145/3225153.3225165

Inger Ekman. 2013. On the desire to not kill your players: Rethinking sound in pervasive and mixed reality games. In *Proceedings of the 8th International Conference on the Foundations of Digital Games 2013*. Society for the Advancement of the Science of Digital Games, 142–149.

Andreas Geiger, Imke Bewersdorf, Elisabeth Brandenburg, and Rainer Stark. 2018. Visual Feedback for Grasping in Virtual Reality Environments for an Interface to Instruct Digital Human Models. In *Advances in Usability and User Experience*. Springer International Publishing, 228–239. https://doi.org/10.1007/978-3-319-60492-3_22

Shangchen Han, Beibei Liu, Robert Wang, Yuting Ye, Christopher D. Twigg, and Kenrick Kin. 2018. Online Optical Marker-based Hand Tracking with Deep Labels. *ACM Trans. Graph.* 37, 4, Article 166 (July 2018), 10 pages. https://doi.org/10.1145/3197517.3201399

Lorraine Lin and Sophie Jörg. 2016. Need a Hand?: How Appearance Affects the Virtual Hand Illusion. In *Proceedings of the ACM Symposium on Applied Perception (SAP '16)*. ACM, 69–76. https://doi.org/10.1145/2931002.2931006

Matthew Longo, Friederike Schüür, Marjolein P M Kammers, Manos Tsakiris, and Patrick Haggard. 2008. What is embodiment? A psychometric approach. *Cognition* 107 (July 2008), 978–98. https://doi.org/10.1016/j.cognition.2007.12.004

Jean-Luc Lugrin, David Obremski, Daniel Roth, and Marc Erich Latoschik. 2016. Audio Feedback and Illusion of Virtual Body Ownership in Mixed Reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology (VRST '16)*. ACM, 309–310. https://doi.org/10.1145/2993369.2996319

Jim R. Parker and John Heerma. 2008. Audio Interaction in Computer Mediated Games. *International Journal of Computer Games Technology* (2008). https://doi.org/10.1155/2008/178923

Mores Prachyabrued and Christoph W. Borst. 2014. Visual Feedback for Virtual Grasping. *2014 IEEE Symposium on 3D User Interfaces (3DUI)* (March 2014), 19–26. https://doi.org/10.1109/3DUI.2014.6798835

Valerio Santangelo and Charles Spence. 2007. Multisensory Cues Capture Spatial Attention Regardless of Perceptual Load. *Journal of Experimental Psychology. Human Perception and Performance* 33 (Dec 2007), 1311–21. https://doi.org/10.1037/0096-1523.33.6.1311

Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2012. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review* 20 (Nov 2012), 21–53. https://doi.org/10.3758/s13423-012-0333-8

Lode Vanacken, Tovi Grossman, and Karin Coninx. 2009. Multimodal Selection Techniques for Dense and Occluded 3D Virtual Environments. *Int. J. Hum.-Comput. Stud.* 67, 3 (March 2009), 237–255. https://doi.org/10.1016/j.ijhcs.2008.09.001

Spyros Vosinakis and Panayiotis Koutsabasis. 2018. Evaluation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift. *Virtual Reality* 22, 1 (01 Mar 2018), 47–62. https://doi.org/10.1007/s10055-017-0313-4

Mathias Wellner, Audrey Schaufelberger, Joachim v. Zitzewitz, and Robert Riener. 2008. Evaluation of Visual and Auditory Feedback in Virtual Obstacle Walking. *Presence: Teleoperators and Virtual Environments* 17, 5 (2008), 512–524. https://doi.org/10.1162/pres.17.5.512

Ye Yuan and Anthony Steed. 2010. Is the Rubber Hand Illusion Induced by Immersive Virtual Reality?. In *Proceedings of the 2010 IEEE Virtual Reality Conference (VR '10)*. IEEE Computer Society, 95–102. https://doi.org/10.1109/VR.2010.5444807

Y. Zhang, T. Fernando, H. Xiao, and A. R. L. Travis. 2006. Evaluation of Auditory and Visual Feedback on Task Performance in a Virtual Assembly Environment. *Presence* 15, 6 (Dec 2006), 613–626. https://doi.org/10.1162/pres.15.6.613