Effects of Auditory Cues on Grasping a Virtual Object with a Bare Hand

Mie Sato
Utsunomiya University
7-1-2 Yoto, Utsunomiya
Tochigi, Japan 321-8585
mie@is.utsunomiya-u.ac.jp

Haruna Kimura
Utsunomiya University
7-1-2 Yoto, Utsunomiya
Tochigi, Japan 321-8585

ABSTRACT

We have been developing an augmented reality (AR) system that allows a user to grasp a virtual object with a bare hand. To enhance the user's perception of grasping the virtual object, we employ multisensory integration in our AR system. Our experimental results show that presenting a virtual object with an auditory cue is statistically more effective than presenting one without an auditory cue as regards grasping, holding, translating, rotating and releasing the virtual object.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Mixed / augmented reality**;

KEYWORDS

auditory cues, interaction, augmented reality

ACM Reference format:

Mie Sato and Haruna Kimura. 2017. Effects of Auditory Cues on Grasping a Virtual Object with a Bare Hand. In *Proceedings of SIGGRAPH '17 Posters, Los Angeles, CA, USA, July 30 - August 03, 2017,* 2 pages. https://doi.org/10.1145/3102163.3102181

Figure 1: Our AR system

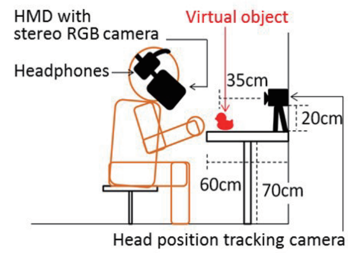


Figure 2: Experimental environment

1 INTRODUCTION

Bare hand interaction with a virtual object has the great advantage of reducing the discomfort caused by devices mounted on a user's hand. There have been some studies on such bare hand interaction [Benko et al. 2012]. We have developed an augmented reality (AR) system that allows a user to grasp a virtual object with a bare hand. However, the bare hand interaction means there is no real physical feedback to the user. Our previous study indicated that a bare hand interaction is needed to provide the user with a greater perception of grasping a virtual object [Sato et al. 2016].

Studies on the multisensory integration of haptic, auditory and visual information have been actively conducted. The results showed that presenting stimuli to haptic, auditory and visual sensory receptors simultaneously improved perceptions of human behavior as a result of this multisensory integration.

SIGGRAPH '17 Posters, July 30 - August 03, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5015-0/17/07…$15.00
https://doi.org/10.1145/3102163.3102181

Our challenge is to improve the perception of grasping a virtual object with a bare hand. We apply multisensory integration to our AR system and present visual information with an auditory cue. We then examine whether the auditory cue is effective in enhancing the user's perception of grasping a virtual object with a bare hand.

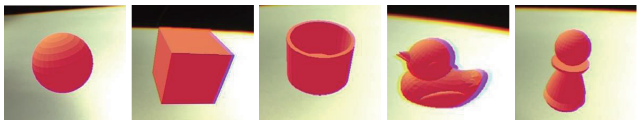
2 EXPERIMENT

Using our AR system, we conducted an experiment to examine the effects of presenting an auditory cue when a participant grasps and releases a virtual object. Twenty participants evaluated their feelings of grasping, holding, translating, rotating and releasing a virtual object with and without an auditory cue.

2.1 Our AR System

Our AR system is composed of a head-mounted display (HMD) (Oculus Rift DK2, Oculus VR) with a stereo RGB camera (Ovrvision 1 for DK2, Wizapply), headphones (ATH-M20x, Audio-Technica), a head position tracking camera (HMD accessory) and a PC. A participant wears the HMD with the stereo camera and the headphones. Figure 1 shows an overview of our AR system.

A stereoscopic view of a virtual object and the participant's hand captured with the stereo camera is displayed on the HMD in real time. By checking the positions of the virtual object and the participant's thumb and index finger, the AR system determines whether the thumb and index finger collide with the virtual object. While the collision occurs, the shape of the participant's hand is modified so that the thumb and index finger touch the surface of the virtual object. With a hidden-surface process, the participant can grasp the virtual object with his/her bare hand on the HMD. The AR system is described in detail in our previous study [Sato et al. 2016].



(a) sphere (b) cube (c) cup (d) duck (e) chess pawn

Figure 3: Virtual objects

Table 1: Signaling tones

pitch interval	grasp (length)	release (length)
major 3rd	C#5, 554Hz (150ms)	A4, 440Hz (150ms)
major 3rd	C5, 523Hz (150ms)	G#4, 415Hz (150ms)
perfect 4th	C5, 523Hz (150ms)	G4, 329Hz (150ms)

Table 2: Sounds associated with each virtual object

virtual object	associated sound (length)	
sphere	grasp	hitting table tennis ball (120ms)
	release	table tennis ball bouncing (140ms)
cube	grasp	bell tinkling 1 (980ms)
	release	bell tinkling 2 (900ms)
cup	grasp	water dripping 1 (130ms)
	release	water dripping 2 (630ms)
duck	grasp	crushing rubber duck toy (170ms)
	release	releasing rubber duck toy (570ms)
chess pawn	grasp	grasping game piece (90ms)
	release	releasing game piece (190ms)

2.2 Virtual Objects and Auditory Cues

We prepared five virtual objects with various shapes as shown in Figure 3. The virtual objects were in the shape of a sphere, a cube, a cup, a duck and a chess pawn.

Five auditory cues were prepared for each virtual object. We presented one of the auditory cues each time a participant grasped and released a virtual object. There were three types of auditory cues: signaling tones, voices, and sounds associated with the virtual object. Table 1 shows the signaling tones whose impressions are "decided," "agree," "okay," "yes" and so on. The voices say the words "grasping" and "releasing" (both 690 ms in length, in Japanese). The sounds associated with each virtual object are shown in Table 2. In our preliminary experiment, we confirmed that these associated sounds matched the impressions of the virtual objects when the virtual objects were grasped and released.

2.3 Procedure

Figure 2 shows a side view of our experimental environment. A participant is asked to grasp, hold, translate, rotate and release each virtual object in Figure 3(a)-(e). We randomly present one of the five auditory cues when the participant grasps and releases the virtual object. After releasing the virtual object, the participant is asked to evaluate the following feelings on a six-point scale: (i) grasp, (ii) hold, (iii) translate, (iv) rotate and (v) release the virtual object. An evaluation without any auditory cue is also carried out. After presenting the virtual object with and without the five auditory cues, we ask the participant whether it is better with or without the auditory cue. Note that to eliminate the influence of unfamiliarity with the AR system, the participant first practices the above with a different virtual object from those shown in Figure 3.

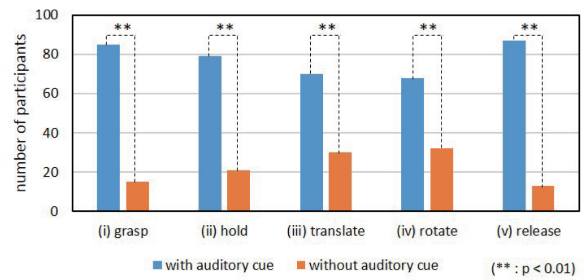


Figure 4: Results of data comparison

3 RESULTS AND CONCLUSION

We had 600 data (six auditory stimuli (five auditory cues and no auditory cue) × five virtual objects × twenty participants) for each of the feelings (i)-(v). We had a total of 100 data (five virtual objects × twenty participants) for comparing the results with and without an auditory cue.

Focusing on the feelings of (i) grasp and (v) release the virtual object, we found that the signaling tones tended to receive a higher evaluation than the voices and the associated sounds. In particular, "perfect 4th (C5, 523Hz and G4, 329Hz)" worked well with the virtual objects "sphere," "cup" and "chess pawn," and "major 3rd (C#5, 554Hz and A4, 440Hz)" worked well with the virtual objects "cube" and "duck." A Steel-Dwass test for multiple comparison ($p < 0.05$) revealed that there was no significant difference between the five auditory cues, although there were significant differences between most of the five auditory cues and when there was no auditory cue.

Figure 4 shows data comparing the results with and without the auditory cues for all of the virtual objects. A binomial test ($p < 0.01$, two-tailed) revealed significant differences between the results. From Figure 4, we can see that presenting a virtual object with an auditory cue was more effective than when there was no auditory cue for all the feelings (i)-(v). In addition, the complex virtual objects (chess pawn, duck, cup) showed this tendency more than the simple virtual objects (sphere, cube).

In conclusion, the presentation of an auditory cue improves the perception of grasping a virtual object with a bare hand. To realize a natural interaction with a virtual object, we study presenting a wider variety of auditory cues with different impressions.

This work was supported by JSPS KAKENHI Grant Number 15K00373.

REFERENCES

Hrvoje Benko, Ricardo Jota, and Andrew Wilson. 2012. MirageTable: Freehand Interaction on a Projected Augmented Reality Tabletop. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 199–208.

Mie Sato, Daiki Ebihara, Sota Suzuki, and Sho Kato. 2016. An AR System on Manipulating a Virtual Object with a Bare Hand. In *SIGGRAPH ASIA 2016 Posters*. 19:1.