Auditory, Graphical and Haptic Contact Cues
for a Reach, Grasp, and Place Task
in an Augmented Environment

Mihaela A. Zahariev, Christine L. MacKenzie
Simon Fraser University
8888 University Drive, Burnaby, B.C., Canada
1-604-291-5794
mihaelaanca_zahariev@sfu.ca, christine_mackenzie@sfu.ca

ABSTRACT

An experiment was conducted to investigate how performance of a reach, grasp and place task was influenced by added auditory and graphical cues. The cues were presented at points in the task, specifically when making contact for grasping or placing the object, and were presented in single or in combined modalities. Haptic feedback was present always during physical interaction with the object. The auditory and graphical cues provided enhanced feedback about making contact between hand and object and between object and table. Also, the task was performed with or without vision of hand. Movements were slower without vision of hand. Providing auditory cues clearly facilitated performance, while graphical contact cues had no additional effect. Implications are discussed for various uses of auditory displays in virtual environments.

Categories and Subject Descriptors

H.5.2. [**Information Interfaces and Presentation**]: User Interfaces - *evaluation/methodology, auditory (non-speech) feedback.*

General Terms

Human Factors, Performance, Measurement.

Keywords

Auditory displays, visual information, multimodal displays, prehension, proprioception, virtual reality, object manipulation, human performance, Fitts' law.

1. INTRODUCTION

In augmented and virtual environments, information is usually displayed graphically. With the increasing availability of auditory and haptic displays, several modalities can be combined to provide a rich multisensory experience, closer to the way we interact with objects in the physical world.

For certain tasks presented in virtual worlds, multimodal user feedback can lead to enhanced human performance. Acoustic signals can be used as substitutes for feedback that would normally be available through other modalities (e.g., sound for force), or to provide redundant feedback [9]. The benefits of sensory substitution for force feedback through auditory and vibrotactile feedback were demonstrated in teleoperator peg-in-hole tasks [7].

In this experiment we investigated the effects of auditory and graphical contact cues on a direct manipulation task performed in a tabletop augmented environment. The task was to reach and grasp a physical object and place it on a graphical target. The cues provided enhanced feedback about making contact between hand and object and between object and table. We expected that the additional feedback would aid subjects to perform the task more effectively.

In order to characterize subjects' movements, we used three-dimensional kinematic measures such as movement time, peak velocity and deceleration time toward the target [3]. Movement time has long been used to assess human performance and to characterize the difficulty of a task. According to Fitts [1], movement time increases with increases in index of difficulty. Fitts' law was extended to two-dimensional target acquisition tasks [5], and also to reaching to acquire computer-generated target objects [6]. Availability of haptic contact information and of visual feedback of the moving limb has been shown to have profound effects on human performance. In this paper, the groundwork is being laid to extend Fitts' law to three-dimensional movement augmented with auditory cues.

2. METHOD

2.1 Subjects

Sixteen university students (11 males and 5 females, aged from 19 to 33 years) were paid to participate in the study. All subjects were right-handed with normal or corrected-to-normal vision, and reported to have normal hearing. This research was approved by the Simon Fraser University Ethics Review Committee.

2.2 Apparatus and Set-up

The experimental set-up is illustrated in Figure 1: an upside down SGI monitor displays an image on a half-silvered mirror that lies between the tabletop and computer screen. Thus, the image is reflected by the mirror so that it appears as a graphical image in a workspace on the table surface. Subjects wore CrystalEYES

goggles so that the reflected image appeared 3-D, and headphones (Sony) to hear the auditory cue.

Infrared emitting diodes (IREDs) were used in conjunction with a dual-sensor OPTOTRAK 3-D motion analysis system (Northern Digital Inc., Waterloo, Canada) to sample (200 Hz) and record the 3-D movement pattern of each subject and transport of the cube. IREDs were placed on the goggles, on the subject's right hand (wrist, thumb and index finger), and on the cube. Markers on the goggles and cube drove the stereoscopic, head-coupled graphics display.
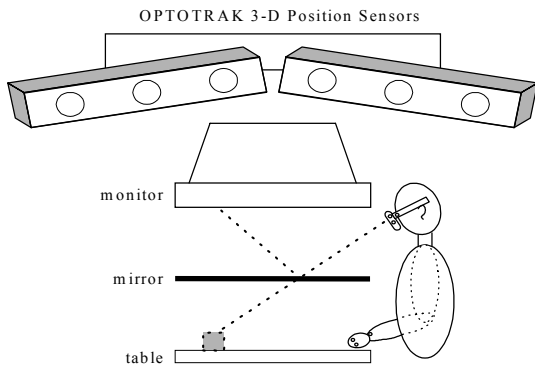

Figure 1. Experimental set-up of the Enhanced Virtual Hand Laboratory

The object used was a black polyvinyl cube (1.5 cm, 46.71 g), containing an embedded load cell (Entran Devices Inc., Fairfield, NJ) to quantify grasping forces. The signals from the load cell were amplified and transferred to a computer by a 12 bit A/D converter called the OPTOTRAK Data Acquisition Unit (ODAU, Northern Digital Inc., Waterloo, Canada).

The physical cube was augmented with a veridical, solid brown graphic overlay. The cube was presented at a 45-degree angle so as to make it easier for the subject to grasp the object. The graphical target consisted of a graphical white square (1.5 cm), also presented at a 45-degree angle, that appeared to be on the table surface.

2.3 Procedure
Seated subjects started with the thumb placed over a start mark. The appearance of the graphical cube and target was the indicator for the subjects to begin their movement. The distance between the start mark and graphical target was 28 cm; the cube was placed at 14 cm. Subjects were instructed to move as fast and as accurately as possible.

2.3.1 Experimental conditions
There were three experimental conditions: vision of hand, cue type, and cue location. In half of the trials, the subjects could see their hand, the physical cube on the workspace tabletop, and the 3-D graphical cube (vision condition). In the other half, vision of the hand and physical cube was prevented (no vision condition) by placing a piece of black cardboard under the mirror. Vision and no vision were counterbalanced.

For each vision condition, blocks of trials were performed for each cue type: auditory (A), graphical (G), or both auditory and graphical cues presented simultaneously (AG). The auditory cue was a 2000-Hz pure tone, 200ms duration. The graphical cue was a change in brightness of the graphical cube from dark brown to light brown.

Haptic feedback was present always during physical interaction with the object. For each of the three cue types (A, G, AG), the subject would receive the cue either upon object contact (touch), triggered by grasping force, or upon object placement on the table (place), triggered by the 3-D cube position. The order of presentation of cue type and cue location was counterbalanced. The subjects were informed about the cue type and cue location prior to each block of trials. Each block consisted of 12 trials.

In addition to the experimental conditions outlined above, four blocks of trials with no additional auditory or graphical cues were included; in these trials subjects had only haptic feedback (H). The total number of trials was 202, including 10 practice trials at the beginning of the experimental session.

2.3.2 Data Analysis
The task was divided into *three phases*: reach, acquire object, and place object. The start of the reach phase was defined as wrist velocity greater than 3mm/s. The first exertion of grasping force defined the end of the reach phase and start of the acquire object phase. Lifting the object off the table defined the end of the acquire object phase and start of the place phase. The place phase ended with object placement on the table, defined according to the 3-D position of the object.

3. RESULTS
Kinematic features were analyzed as dependent measures. They included: movement time, peak velocities and their time of occurrence, and aperture profiles. Repeated measures ANOVAs were performed on the 2 (vision) X 3 (cue type) X 2 (cue location) design for each phase of the movement. An a priori alpha level of .05 was used for all statistical tests.

3.1 Vision Effects
As in other studies [2], vision of the hand had a significant impact on performance. Whether or not subjects saw their hands affected all phases of movement, as expected. Movement time was slower, peak velocity was lower, and peak aperture was larger for 'no vision', compared to 'vision' conditions.

3.2 Cue Effects
For the *reach phase* of the movement, significant main effects of cue type and cue location are outlined in Table 1 and Table 2, respectively. Included are the means (numbers in parentheses represent standard error) for movement time (MT), peak velocity (PV), time to peak velocity (TPV), time after peak velocity (TAPV) and percent time after peak velocity (PTAPV). All results presented for cue effects have been computed over both vision and no vision conditions.

Cue type affected the reach phase of movement, with faster movement time and higher peak velocity for AG and Auditory, compared to Graphical only cues. Similarly, cue location affected the reach phase of movement, with faster movement time and higher peak velocity when the cue occurred at 'touch' compared to 'place'.

Table 1. Reach phase: Cue type main effects. Kinematic measures

	AG	Auditory	Graphical
MT (ms)	455 (29)	455 (29)	488 (32)
PV (mm/s)	471 (22)	463 (22)	444 (21)
TPV (ms)	275 (10)	278 (10)	289 (10)
TAPV (ms)	180 (21)	176 (21)	198 (24)

Table 2. Reach phase: Cue location main effects. Kinematic measures

	Touch	Place
MT (ms)	456 (27)	476 (32)
PV (mm/s)	464 (21)	454 (21)
TAPV (ms)	176 (20)	194 (23)
PTAPV (%)	36 (2)	38 (2)

For the reach phase of the movement, cue type effects on movement time and peak velocity are outlined in Figure 2 and Figure 3, respectively. For comparison purposes, the means for the Haptic only condition (H) are included.
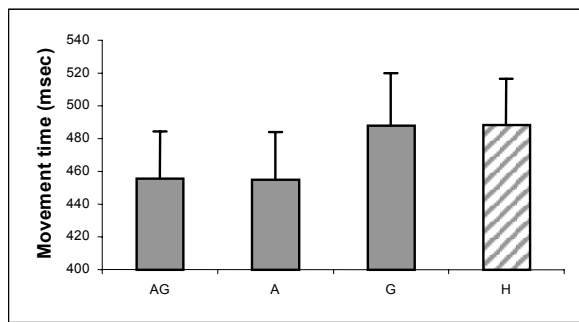
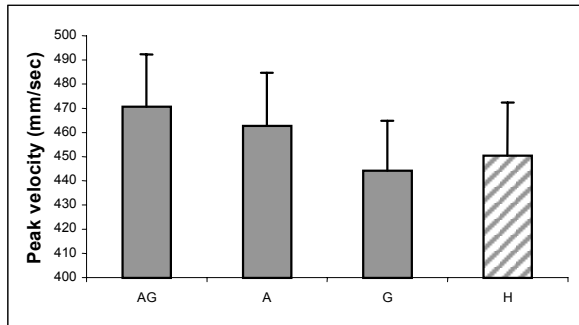In Figure 4, as in Figure 2 and 3, the Graphic only results are similar to Haptic only results, and AG results are similar to Auditory only results. No significant cue main effects were found for the acquire object phase.
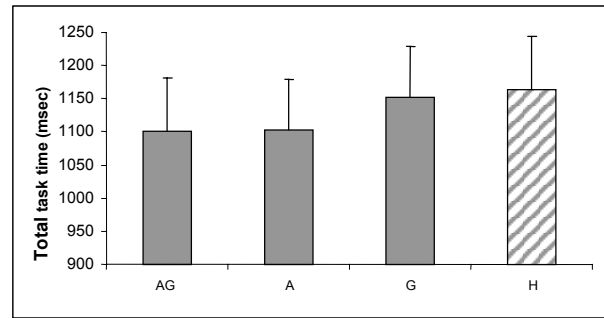

Figure 2. Reach phase movement time

Figure 3. Reach phase peak velocity

Figure 4. Total task time

4. DISCUSSION

In this prehension task, subjects always had natural haptic contact information. Providing redundant auditory, but not graphical, contact cues significantly enhanced performance. Providing graphical only cues provided no benefit, with performance being similar to haptic only condition. Providing both graphical and auditory cues provided no added benefit, compared to auditory only cues.

It is possible that the intensity of the graphical cue (a change in brightness) was too weak to have an effect on performance. The facilitatory effect of the auditory stimuli might be due to their more pronounced alerting or 'attention-grabbing' properties, compared to graphical stimuli [8]. Regardless of the underlying reason for the facilitatory effect, auditory cues likely have great potential for uses other than alerts and alarms in HCI. While the 50 ms advantage for auditory cues during object manipulation may seem small, this could quickly add up to significant time savings over hundreds of object manipulations in enhanced environments.

At the outset, we expected that the additional auditory and graphical cues would act as feedback, that is, movements would be affected *after* the occurrence of these cues. However, cue effects did occur during the reach phase, *before* the cues were provided, suggesting that subjects anticipated their presence. Prior to each block of trials, subjects were informed about the type and location of the cues, so they could predict them with certainty. It is possible that the expectation of enhanced contact feedback may have influenced the planning stages of the movement. Moreover,

For the *place phase* of the movement, the same trend was observed, with faster movement time for AG and Auditory (538 ms and 537 ms, respectively), compared to Graphical cues (554 ms). Although the cue type effect for the place phase did not reach significance (p>.05), preplanned contrasts revealed a significant difference between Auditory only and Graphical only cues (p<.05).

Cue type also affected the *total task time* (p<.05), which was about 50 ms faster for AG and Auditory only (1100 ms and 1102 ms, respectively), compared to Graphical only cues (1152 ms), as seen in Figure 4. The mean for Haptic only condition is again included for comparison. Total task time was from the start of reach to the time of object placement. The cue type effect in the reach phase mostly contributed to the differences in total task time.

reaching movements were faster, with a higher peak velocity and a lower proportion of time spent after peak velocity with enhanced feedback at 'touch' (hand contacts object), compared to 'place' (object contacts table) conditions. The percent of time spent in the deceleration phase has been shown to increase as the precision required for the movement increases, which has been termed as the 'precision effect' [4]. Although this effect has been shown with targets of varying sizes, and our experiment used a single target size, it can be argued that providing the cues to enhance the haptic feedback of making contact between hand and object acts to decrease the 'precision' requirement of the reaching phase of the task.

The availability, expectation and modalities of feedback can all influence motor performance. For designing effective multimodal displays, it is essential to consider the complex interactions among different sensory modalities, and also to establish the most appropriate sensory combinations for specific tasks. Future work might explore the potential benefits of auditory cues for direct manipulation tasks with varying difficulty, and for manipulating virtual objects, when haptic feedback is not available.

In conclusion, our results suggest the benefits of auditory cues for enhancing manipulation performance in the design of multimodal displays for interactive augmented and virtual environments. Auditory displays may also provide cost benefits by reducing the requirement for expensive high quality graphics or haptics displays.

5. ACKNOWLEDGMENTS

6. REFERENCES

1. Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381-391.

2. Jeannerod, M. (1981). Intersegmental coordination during reaching at natural visual objects. In J. Long & A. Baddeley (Eds.) *Attention and Performance* IX, Hillsdale, NJ: Erlbaum, 153-169.

3. MacKenzie, C. L. & Iberall, T. (1994). *The grasping hand*. Advances in Psychology, Volume 104, Amsterdam: North-Holland (Elsevier Science B.V.), 109-201.

4. MacKenzie, C.L., Marteniuk, R.G., Dugas, C., Liske, D., Eickmeier, B. (1987). Three-dimensional movement trajectories in Fitts' task: implications for control. *The Quarterly Journal of Experimental Psychology,* 39A: 629–647.

5. MacKenzie, I.S., and Buxton, W. (1992). Extending Fitts' law to two-dimensional tasks. *Proceedings of the CHI '92 Conference on Human Factors in Computing Systems.* New York: ACM, 219-226.

6. Mason, A.H., Walji, M.A., Lee, E.J., & MacKenzie, C.L. (2001). Reaching movements to augmented and graphic objects in virtual environments, CHI Notes – *Proceedings of the ACM Conference on Computer-Human Interaction*, 426-433.

7. Massimino, M.J., & Sheridan, T.B. (1993). Sensory substitution for force feedback in teleoperation. *Presence: Teleoperators and Virtual Environments*, 2(4), 344-352.

8. Posner, M. I., Nissen, M. J., and Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance, *Psychological Review*, Vol. 83, No. 2, 157-171.

9. Stanney, K.M., Mourant, R.R., and Kennedy, R.S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments*, 7(4), 327-351.