Do Multimodal Signals Need to Come from the Same Place? Crossmodal Attentional Links Between Proximal and Distal Surfaces

Rob Gray
Department of Applied Psychology,
Arizona State University East
7001 E Williams Field Road, Building 20,
Mesa, AZ 85212
Email: robgray@asu.edu; Phone: (480)
727-1340; Fax: (480) 727-1777

Hong Z. Tan & J. Jay Young
Haptic Interface Research Laboratory,
Purdue University
1285 Electrical Engineering Building, West
Lafayette, IN 47907
Email: hongtan@ecn.purdue.edu; Phone:
(765) 494-6416; Fax: (765) 494-6951

Abstract

Previous research has shown that the use of multimodal signals can lead to faster and more accurate responses compared to purely unimodal displays. However, in most cases response facilitation only occurs when the signals are presented in roughly the same spatial location. This would suggest a severe restriction on interface designers: to use multimodal displays effectively all signals must be presented from the same location on the display. We previously reported evidence that the use of haptic cues may provide a solution to this problem as haptic cues presented to a user's back can be used to redirect visual attention to locations on a screen in front of the user [1]. In the present experiment we used a visual change detection task to investigate whether (i) this type of visual-haptic interaction is robust at low cue validity rates and (ii) similar effects occur for auditory cues. Valid haptic cues resulted in significantly faster change detection times even when they accurately indicated the location of the change on only 20% of the trials. Auditory cues had a much smaller effect on detection times at the high validity rate (80%) than haptic cues and did not significantly improve performance at the 20% validity rate. These results suggest that the use haptic attentional cues may be particularly effective in environments in which information cannot be presented in the same spatial location.

1. Introduction

With the ever-increasing complexity of visual displays being presented to interface operators, many system designers are looking towards multimodal interfaces that use auditory and/or tactile channels as a supplemental or alternative means of information transfer. One of the most encouraging research findings for the developers of multi-modal interfaces is that responses to multimodal stimulation can often be stronger at the cellular level [2], and faster and more accurate (e.g., [3] [4, 5]) at the behavioral level relative to responses to unimodal stimuli.

However, these basic research findings also seem to impose a very serious design limitation, namely that for multimodal displays to be effective the signals must come from roughly the same locations relative to the operator. Previous research has shown that response facilitation only occurs when the stimuli from different modalities are in roughly the same spatial location [6, 7]. Conversely, responses are inhibited (e.g., reaction times are slowed) when the stimuli from the different modalities are in very different spatial locations. So, for example, in order to reorient an operator's attention to an important visual event (e.g., an impending collision in a car or airplane) an auditory signal must be presented in roughly the same location as the visual stimulus. If the stimuli from the different modalities are presented in very different locations the use of a multimodal display may hinder performance by diverting attention away from the critical event [8], thus increasing the time to detect the visual stimulus [5]. This situation can be even further exacerbated when an event occurs in one modality when it is expected from another [8]. Given that presenting multimodal signals in the same spatial location is difficult if not impossible in many operating environments this would seem to be a grave problem indeed.

But is it truly the case that facilitatory interactions only occur when signals from the different modalities are presented in the same spatial location? Because our sense of touch informs us

about events on our skin surface, in many instances it would not be advantageous to reorient our visual or auditory attention to the exact spatial location that received tactile stimulation. For example, when you feel a proximal tap on your shoulder your visual attention is reoriented to distal space instead of the location on your shoulder that was tapped. Thus, the spatial relationships between multimodal signals that produce response facilitation may be very different for the sense of touch. Whereas, visual and auditory stimuli need to be in close spatial proximity to produce facilitation [2], our somatosensory system may make transformations between proximal locations on the skin surface and distal areas in external space. In a previous study we provided evidence supporting this proposal. We used haptic cues (taps on the user's back) to reorient visual attention in a visual change-detection task that was displayed on a monitor in front of the user [9] (see Fig 1A&B below). When the haptic cue accurately indicated the location of the change, visual detection times were reduced by 1630 ms relative to baseline.

In the current study we expanded on our previous work in two ways. When studying crossmodal interactions it is important to rule out the possibility that the cueing effects are produced by a learned strategic shift in attention (e.g., that could be produced by an irrelevant verbal cue), rather than low-level sensory interactions [10]. To address this issue we varied the validity of the haptic cue in the present study. If attentional reorienting occurs at low cue validity rates (when it is actually a hindrance to performance on the majority of trials) it suggests that the effect is the result of a natural sensory integration process rather then a higher-level learning effect [10]. In the current study we also investigated whether auditory cues (presented behind the user in the same spatial arrangement as the haptic cues) can be used to improve performance in the same visual detection task. If visual-haptic interactions are inherently different than visual-auditory interactions (in terms of their spatial properties) than we might expect that user's are better able to make front-to-back transformations for haptic cues than auditory cues.

2. Methods

2.1 Apparatus

The visual stimuli used were similar to that described in our previous study [9]. Briefly, the presentation of the visual stimulus was based on the flicker paradigm [11] used to study "change blindness". In this paradigm, two scenes, differing only slightly (e.g., in the position or size of one of the objects in the scene), are presented in an alternating order with a blank screen inserted between each alternation. Previous research using this paradigm has shown that participants can take upwards of 20-30 sec to detect even large changes in the scene (reviewed in [12]) and that the time required for change detection can be dramatically reduced by manipulations that draw attention to the changing element [13]. In the present study, the scenes were comprised of rectangular elements which had either a horizontal or vertical orientation (Figure 1A) and the orientation of one of the rectangles was different for the two scenes (e.g., the top-right rectangle in Fig. 1A). The duration of the two patterned scenes was 80 ms and the duration of the blanks was 120 ms. The visual stimuli were displayed on a 17''computer monitor.
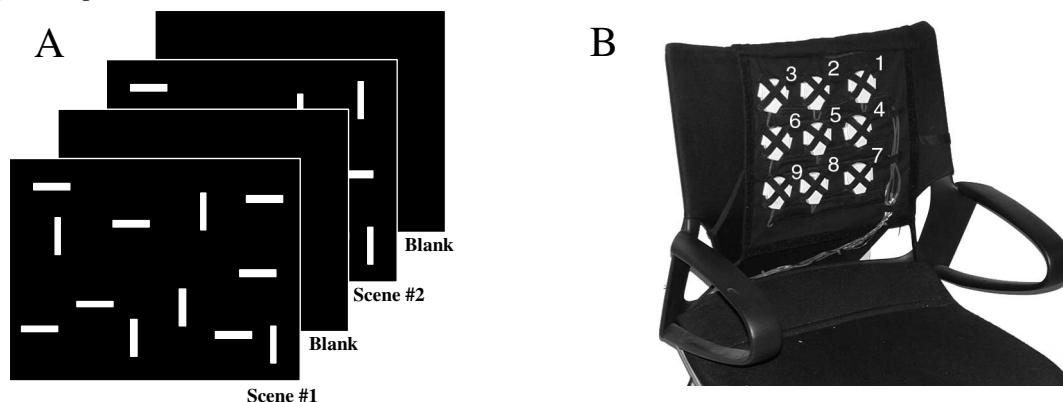
Figure 1. A: The stimulus display for a visual change detection task presented on a monitor in front of the user. B: The haptic directional display used for tactile cueing. See text for details.

The haptic stimuli were delivered via a 3-by-3 vibrotactile display, developed at the Purdue Haptic Interface Research Laboratory, draped over the back of an office chair (Figure 1B). In the present study, only the four corner tactors (i.e., No. 1, 3, 7 and 9 in Fig 1B) were used. Each tactor could be independently driven by a 60 ms sinusoidal pulse with a frequency between 209-306 Hz (corresponding to the resonant frequency of the four tactors). Under unloaded conditions, the intensity of the vibration was between 26.1-27.9 dB SL. In conditions in which haptic stimuli were used, white noise was played continuously through headphones so that participants could not hear the tactors.

The auditory stimuli were 60 msec bursts of white noise. These stimuli were presented via a 2 x 2 array of loudspeakers positioned 20 cm behind the back of the participant's seat. This arrangement was chosen so as to make the spatial positions of auditory and tactile cues as similar as possible. The separation between the speakers was identical to the separation between tactors 1, 3, 7 and 9 in Fig 1B.

2.2. Procedure

To ensure that the participants could clearly feel the vibrations and could correctly identify the four tactor locations, an absolute identification experiment was conducted before each new session. The participant's task was to click on one of the four quadrants of the monitor (represented by four large rectangles) in response to a vibration on the back (e.g., the correct response to a vibration near the right shoulder would be to click the upper-right quadrant of the monitor). Each participant had to complete one run of 60 trials with 100% correct performance before proceeding to the main experiment. An analogous experiment was used to ensure that participant's could identify the locations of the auditory stimuli.

The main experiments involved a visual change detection task. On each trial, participants were presented with alternating scenes as shown in Fig 1A and were instructed to click the left mouse button as soon as the changing element was detected (without moving the cursor over the element). Following this initial mouse click the screen froze and the color of the bars changed from white to pink. Participants were then required to make a second mouse click with the cursor on the element that was perceived to change. The time of the first mouse click was recorded as the detection time. The x-y positions of

the second mouse click were used to discard trials where the wrong element was identified.

The independent variables used in this study were the Cue Type and Cue Validity. Two different cue types were used: (1) haptic cueing and (2) auditory cueing. For the haptic cueing conditions, each trial began with a 60 msec vibration sent through one of the four corner tactors. The screen was blank during the tactile stimulation. The alternating visual scenes began 200 msec after the onset of the tactile vibration. Data for auditory cueing was collected on separate runs. Auditory cues were presented for the same duration and had the same SOA as haptic cues. The intent of these manipulations was to draw the participant's visual attention to one of the four quadrants of the monitor (although they were not explicitly told to do this). For each participant data was also collected for a no cue baseline condition in which everything was identical expect that the tactors or speakers were turned off. Two levels of cue validity were used: 20% and 80% where these values refer to the % of trials in which the haptic or auditory cue was valid (e.g., the location of the cue coincided with the location of the changing element in the visual display).

Both independent variables were between-subjects manipulations, thus participants were randomly assigned to one of four possible groups: (i) 20%_haptic, (ii) 80%_haptic, (iii) 20%_auditory or (iv) 80% _auditory. Data from 5 participants was collected for each group and each participant completed 480 trials. In separate runs, each participant also completed 180 trials of the no cue baseline condition.

2.3. Data Analysis

The main dependent measure was the time for visual change detection. To measure the effect of cueing on mean detection times (DT) we first split data from cueing conditions into valid and invalid conditions. We then calculated the difference between the mean cued DT and the mean DT for the no-cue baseline for each of these conditions. These differences were then analyzed using 2X2 between-subjects ANOVAs.

3. Results

Figure 2A plots the mean difference in DTs between each of the cueing conditions and the no-cue baseline. These data are for conditions in which the

A

B
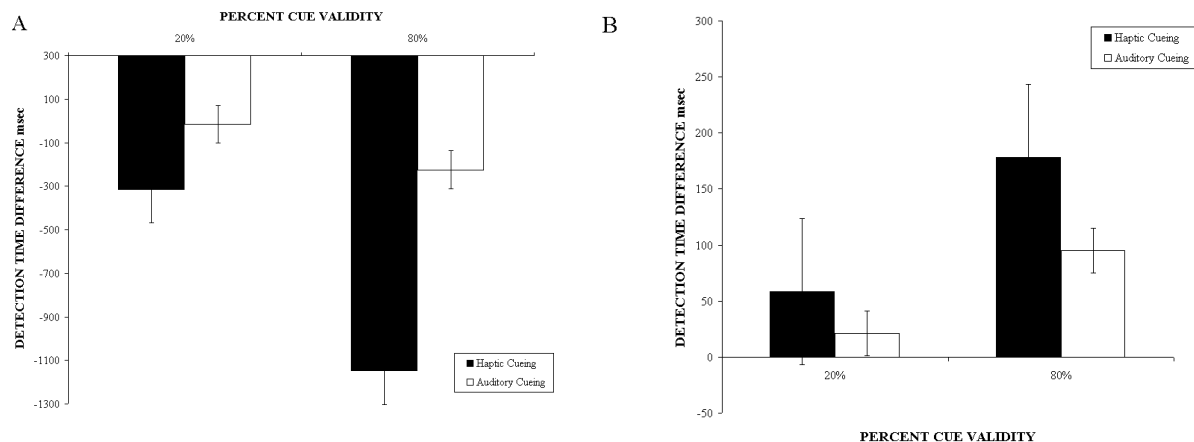
Figure 2. A: Mean differences in detection times (cued vs. baseline) for valid cues. B: Mean differences in detection times for invalid cues. Error bars are standard errors.

cue was valid. Haptic cueing (solid bars) lead to a substantial decrease in detection times (i.e., negative difference) for both the 20% and 80% validity conditions. Conversely, for auditory cues there was no appreciable cueing effect for the 20% validity condition. The ANOVA performed on the data from valid cueing conditions revealed significant main effects of Cue Type and Cue Validity: the DT difference was significantly larger for the haptic cues [$F(1,16)=21.6$, $p<0.001$] and was significantly larger for the 80% validity condition [$F(1,16)=16.3$, $p<0.001$]. There was also a significant type X validity interaction. From Fig 2A it is evident that this interaction was due to the fact that % cue validity had a much larger affect on haptic cues.

Figure 2B shows mean differences in DTs for invalid conditions. Invalid haptic cues increased detection times for both validity conditions and the effects of invalid auditory cues were again smaller than observed for invalid haptic cues. However, the cueing effects were smaller and the variability was considerably larger than for valid conditions. The ANOVA performed on the invalid cueing data revealed no significant effects. This asymmetry between cueing effects for valid and invalid conditions is quite unusual for attentional cueing experiments as the speeded reaction times produced by valid cues are thought to reflect the same mechanism as the slowed reaction times produced by invalid cues. Namely, both effects indicate that the cue is drawing attention to a specific location in space. One possible explanation for the present findings is that the haptic cues in our paradigm lead to a spatially diffuse distribution of attention (see also [14]) such that responses are not slowed as much for invalid cues i.e., attention still covers the location of

the change even when it is drawn away by the invalid cue. The large variability evident in Fig 2B indicates that this effect may differ across individuals. We are currently exploring these issues in more detail.

4. Conclusions

In this study we compared the reorienting of visual attention produced by haptic and auditory cues when the cue and visual target are presented in very different locations in space. We found that (1) haptic cues that are valid 80% of the time reduce reaction times for detection of a visual change significantly more than comparable auditory cues, and (2) detection times are substantially shortened (average 313 msec) for haptic cues that are valid on only 20% of the trials while no significant effects occur for auditory cues in this condition. These findings provide further support for our proposal that visual-haptic interactions are qualitatively different than auditory-visual interactions.   Unlike auditory and visual signals, proximal haptic signals can be used to effectively reorient attention to areas in distal space.

Our results have important implications for the designers of multimodal interfaces. In many environments, multimodal warning signals and the critical information that must be attended cannot be positioned in the same location in the display for obvious practical reasons. Our findings suggest that haptic cues presented a large distance from the main visual display can effectively reorient attention in these situations. The use of such signals would also help to reduce the visual clutter that can be a hindrance to performance in many human-machine interfaces [15].

IEEE
COMPUTER
SOCIETY

5. Acknowledgments

6. References

[1] Tan, H.Z., et al., Haptic cueing of a visual change-detection task: Implications for multimodal interfaces, in Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality (Vol. 1 of the Proceedings of the 9th International Conference on Human-Computer Interaction), R.J. Koubek, Editor. 2001, Lawrence Erlbaum Associates: Mahwah, NJ. p. 678-682.

[2] Stein, B.M. and M.A. Meredith, *The Merging of the Senses*. 1993, Cambridge, MA: MIT Press.

[3] Allport, D.A., B. Antonis, and P. Reynolds, *On the division of attention: a disproof of the single channel hypothesis.* Q J Exp Psychol, 1972. **24**(2): p. 225-35.

[4] Treisman, A.M. and A. Davies, *Divided attention to ear and eye*, in *Attention and performance*, S. Kornblum, Editor. 1973, Academic: New York. p. 101-117.

[5] Spence, C., et al., Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. Perception & Psychophysics, 1998. **60**(4): p. 544-557.

[6] Meredith, M.A., J.W. Nemitz, and B.E. Stein, Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci, 1987. **7**(10): p. 3215-29.

[7] Driver, J. and C. Spence, *Cross-modal links in spatial attention.* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 1998. **353**(1373): p. 1319-1331.

[8] Spence, C., M.E.R. Nicholls, and J. Driver, *The cost of expecting events in the wrong sensory modality.* Perception & Psychophysics, 2001. **63**(2): p. 330-336.

[9] Tan, H.Z., et al., *A haptic back display for attentional and directional cueing.* IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans., 2002. **under review**.

[10] Spence, C. and J. Driver, *Audiovisual links in endogenous covert spatial attention.* Journal of Experimental Psychology-Human Perception and Performance, 1996. **22**(4): p. 1005-1030.

[11] Rensink, R.A., J.K. O'Regan, and J.J. Clark, *To See or Not to See: The Need for Attention to Perceive Changes in Scenes.* Psychological Science, 1997. **8**(368-373).

[12] Rensink, R.A., *Change Detection.* Annual Review of Psychology, 2002. **53**: p. 245-277.

[13] Scholl, B.J., Attenuated change blindness for exogenously attended items in a flicker paradigm. Visual Cognition, 2000. **7**: p. 377 - 396.

[14] Chong, T. and J.B. Mattingley, Preserved cross-modal attentional links in the absence of conscious vision: Evidence from patients with primary visual cortex lesions. Journal of Cognitive Neuroscience, 2000: p. 103A.

[15] Wogalter, M.S., M.J. Kalsher, and B.M. Racicot, Behavioral compliance with warnings: Effects of voice, context and location. Safety Science, 1993. 16: p. 637-654.

IEEE
COMPUTER
SOCIETY