Available online at www.sciencedirect.com
SCIENCE DIRECT®
Brain and Cognition xxx (2004) xxx–xxx
Brain and Cognition
www.elsevier.com/locate/b&c

Flexible and abstract resolutions to crossmodal conflicts

Christa-Lynn Donovan,[a,*] D. Stephen Lindsay,[b] and Alan Kingstone[a]

[a] *University of British Columbia, Canada*
[b] *University of Victoria, Canada*

Accepted 12 February 2004

Abstract

Participants judged whether two sequential visual events were presented for the *same* length of time or for *different* lengths of time, while ignoring two irrelevant sequential sounds. Sounds could be either the same or different in terms of their duration or their pitch. When the visual stimuli were in conflict with the sound stimuli (e.g., visual events were the same, but the sounds were different) performance declined. This was true whether sounds varied in duration or in pitch. The influence of sounds was eliminated when visual duration discriminations were made easier. Together these results demonstrate that resolutions to crossmodal conflicts are flexible across the neural and cognitive architectures. More importantly, they suggest that interactions between modalities can span to abstract levels of same/different representation.

1. Introduction

A rapidly expanding body of research is tackling the question of how information presented concurrently to different senses is combined to form a coherent unitary percept. One particularly important class of multisensory interactions concerns crossmodal conflict. In the classic ventriloquist illusion the position of a sound is mislocalized toward a visual event presented at the same time but at a different position (see Soto-Faraco & Kingstone, in press, for a review). Instances in which auditory information affects the perception of visual events are far less common, but they do occur. For example, Morein-Zamir, Soto-Faraco, and Kingstone (in press) have demonstrated that the perception of the time at which a visual event occurs is biased toward a subsequent auditory event.

The presence of such crossmodal biases is typically interpreted as reflecting the fact that when multisensory conflict occurs, the perceptual system weights the information in a manner that favours the modality most appropriate for the domain that is being processed (Welch, DuttonHurt, & Warren, 1986). Thus, vision biases the perceived spatial position of a sound because vision provides more accurate spatial information than audition. Conversely, audition biases the perceived temporal position of a light because audition provides more accurate temporal information than does vision.

Note that two assumptions are implicit to this modality appropriateness account. First, that during intersensory conflict, dominance of one sense over the other is flexible. That is, in one stimulus domain vision can bias audition, but in another domain audition can bias vision. The dominant modality is the one that provides the most accurate information for that domain. Second, the point of intersensory conflict occurs over a common domain of sensory representation, e.g., space or time. Both assumptions were tested in the present study.

2. Method

2.1. Subjects

Forty-eight undergraduates were divided equally into three groups. Testing lasted approximately 1 h.

* Corresponding author.
E-mail address: cjdonova@interchange.ubc.ca (C.-L. Donovan).

0278-2626/$ - see front matter © 2004 Published by Elsevier Inc.
doi:10.1016/j.bandc.2004.02.019

2.2. Design

Each subject judged whether two sequential visual events were presented for the *same* or *different* length of time. Two irrelevant sequential sounds also occurred. Sounds could be either the same or different in terms of their duration or their pitch. Same/different visual events and same/different auditory events were manipulated orthogonally with random selection.

Group 1. Visual Task (Hard Duration)–Auditory Task (Easy Duration). We expected that the visual judgments would be affected by the irrelevant auditory durations because the accuracy of the temporal information will be superior for audition (Morein-Zamir et al., in press).

Group 2. Visual Task (Easy Duration)–Auditory Task (Easy Duration). If the weighting placed on intersensory conflict is flexible, then for Group 2 the influence of auditory stimuli on visual judgments should decline relative to Group 1. This is because the advantage gained by the superior temporal processing of audition should decline.

Group 3. Visual Task (Hard Duration)–Auditory Task (Easy Pitch). Here the visual duration judgment was difficult (as it was in Group 1), but sounds varied in pitch rather than duration. It is unclear whether there will be any influence of auditory stimuli under these conditions. On the one hand, there is strictly speaking no intersensory conflict in duration judgments, for the auditory stimuli varied in terms of pitch. Yet, it is possible that intersensory conflict may span to a more abstract level of stimulus representation—i.e., a representation that is not tied to duration representations per se, but to "same" and "different" stimulus perceptions. If this is the case, then auditory stimuli may affect visual duration judgments.

For all groups, auditory stimuli could occur at the same time as the visual events (synchronous) or at different times (asynchronous). The asynchronous condition was used to assess for an influence of post-perceptual processes such as response bias. It is well established that crossmodal integration declines rapidly when information is delivered to two senses asynchronously by 200–300 ms. Therefore there should be no true crossmodal perceptual effects in the asynchronous conditions. In contrast, if post-perceptual processes underlie the effect obtained in the synchronous condition, then those same effects should also appear in the asynchronous conditions (Soto-Faraco, Spence, & Kingstone, in press).

2.3. Apparatus and procedure

Visual stimuli were presented on an 18-in. monitor and two speakers, placed on either side of the monitor, 59 cm apart, delivered the auditory stimuli. Accuracy measures were based on keyboard responses.

Group 1. The stimuli were two 1° black asterisks that were flashed in succession at fixation. Each flash was either 50 or 150 ms; the interstimulus interval was 900 ms. Either synchronous or asynchronous with flash onset a pure 440 Hz 64 dB tone was presented for 50 or 150 ms. Flash and sound durations were combined orthogonally to produce 16 conditions. In the asynchronous condition, the tones preceded the asterisks by 450 ms. Again audiovisual pairings produced 16 conditions. Subjects were presented randomly both the synchronous and asynchronous conditions in each block.

Group 2. Same as Group 1 except the longer flash was 250 ms to make the visual task easier.

Group 3. Visual stimuli were as in Group 1. Auditory stimuli were replaced with two 64 dB, 8 ms pure tones. These tones were either 440 or 880 Hz.

Subjects completed 8 blocks of 64 trials. Subjects pressed the 'z' key for same, and 'x' for different. Subjects were told to ignore the tones.

3. Results

Fig. 1 presents mean response accuracy for each group as a function of congruency and synchrony.

Group 1 performed as predicted. Visual performance declined when incongruent auditory information was presented synchronously $F(1,60) = 4.030$, $p < .05$. This shows that auditory stimuli can distort visual perceptions when the visual task involves temporal discriminations. This congruency effect is perceptual because it was absent in the asynchronous condition. Group 2 performance also agreed with our predictions. When the discrimination of visual duration was easier, the interfering effect of synchronous incongruent auditory information disappeared $F(1,52) = 3.741$, $p > .05$. Consistent with these observations an ANOVA comparing Groups 1 and 2 yielded a significant three-way Group × Congruency × Synchrony interaction, $F(1,120) = 9.43$, $p < .005$.

Group 3 performance, with irrelevant sound pitches, matched Group 1 performance, i.e., incongruent < congruent accuracy for synchronous trials $F(1,60) = 6.423$, $p < .05$. Again this congruency effect is perceptual because it was absent for asynchronous trials. These observations were confirmed by an ANOVA comparing Groups 1 and 3, which yielded a Congruency × Synchrony interaction, $F(1,120) = 4.98$, $p < .03$, but no interaction involving group all $F_s < 1.0$.

To summarize, there was a significant effect of congruency in the synchronous condition for Group 1 and Group 3 but not for Group 2. An ANOVA conducted with all three groups yielded a highly significant Group × Congruency × Synchrony, $F(2,172) = 4.37$, $p < .02$.

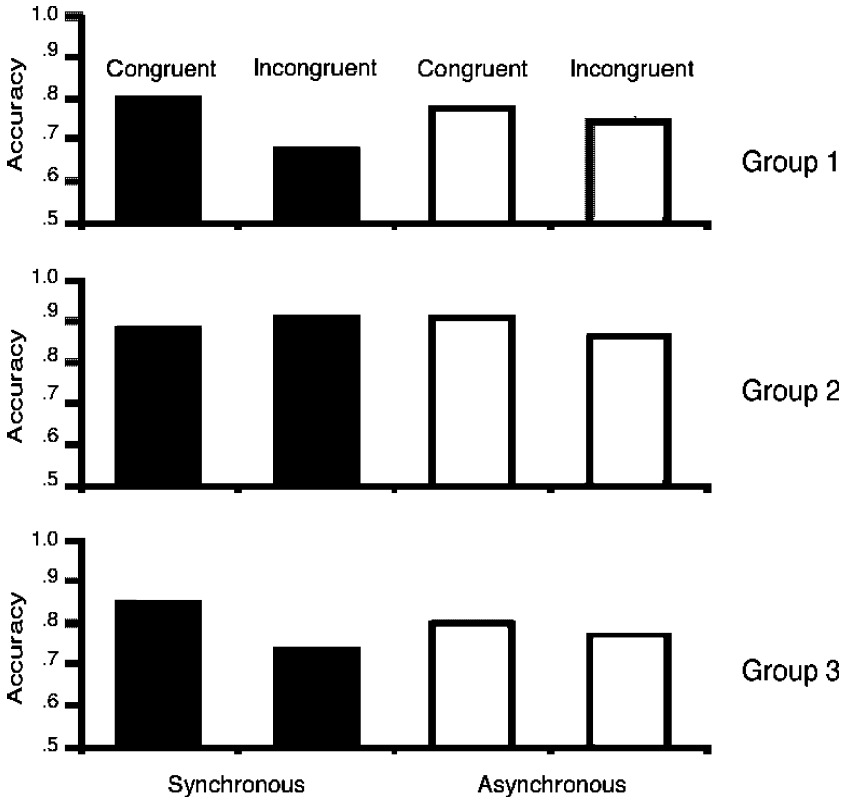

Fig. 1. Performance accuracy for each of the three groups as a function of visual–sound congruency (congruent or incongruent) and synchrony (synchronous or asynchronous). The key point to note is that there was a significant effect of congruency (congruent > incongruent) in the synchronous condition for Group 1 and Group 3 but not for Group 2.

4. Discussion

The present study demonstrated that crossmodal conflict is not specific to vision dominating sound perception, as visual duration perceptions were affected by irrelevant sound information. This study adds to a small, but growing list of investigations reporting that a visual–auditory interaction does not necessitate visual dominance (Morein-Zamir et al., in press; Welch et al., 1986). Hence, our current investigation shows that the relationship between modalities is a flexible one.

In terms of the modality appropriateness hypothesis, flexibility between modalities is determined largely by which modality provides the most accurate information within a given stimulus domain. Sound provides superior temporal information when compared to vision. Thus, we expected that performance in Group 1—in which subjects made a difficult same/different visual duration judgment while irrelevant sounds of the same or different duration were presented—would be affected by whether the sounds were congruent or incongruent with the visual stimuli. This is precisely what we found.

Importantly, and consistent with the modality appropriateness hypothesis, we discovered that when visual durations were made easier for Group 2, the congruency effect of the sounds was eliminated. This demonstrates that the relationship between modalities is flexible and highlights the fact that it is the relative quality of information between modalities that largely determines the nature of crossmodal dominance.

Finally, Group 3 demonstrated that visual–auditory interactions can occur at a remarkably high level of stimulus representation. Here, same/different visual duration perceptions were affected by the congruency of same/different sound pitches in a manner that was indistinguishable from the congruency effect of same/different sound durations observed in Group 1. In terms of the modality appropriateness hypothesis, this suggests that crossmodal perceptual interactions are not restricted to common domains shared across low-level perceptual representations, such as where or when a stimulus has occurred. Rather, they can span higher-order categorical levels of intersection, such as whether they both map onto a "same" or "different" stimulus representation.

In sum, our study has provided fundamental empirical support for the assumption that crossmodal interactions do not reflect a fixed relation between the neural and cognitive architecture, but are instead flexible in nature. In addition, it suggests a significant expansion to the level on which crossmodal interactions may occur, spanning both low-level and higher-levels of stimulus

representations. The role that post-perceptual factors may play in these interactions, if any, is an important and exciting avenue for future investigation.

Acknowledgments

This research was supported by an Undergraduate Research Award to Christa-Lynn Donovan from the Natural Sciences and Engineering Research Council; and by grants to Alan Kingstone from the Human Frontier Science Program, NSERC, and the Michael Smith Foundation for Health Research.

References

Morein-Zamir, S., Soto-Faraco, S., Kingstone, A. (in press). The capture of vision by audition: Deconstructing temporal ventriloquism. *Cognitive Brain Research*.

Soto-Faraco, S., Kingstone, A. (in press). Multisensory integration of dynamic information. In G. Calvert, C. Spence, B. Stein (Eds.), *Handbook of multisensory processes*. Cambridge, MA: MIT Press.

Soto-Faraco, S., Spence, C., Kingstone, A. (in press). Crossmodal dynamic capture. *Journal of Experimental Psychology*: *Human Perception and Performance*.

Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics, 39*, 294–300.