Multidimensional processing of dynamic sounds: more than meets the ear

Estella H. Liu · Eduardo Mercado III · Barbara A. Church

Abstract Strong cross-modal interactions exist between visual and auditory processing. The relative contributions of perceptual versus decision-related processes to such interactions are only beginning to be understood. We used methodological and statistical approaches to control for potential decision-related contributions such as response interference, decisional criterion shift, and strategy selection. Participants were presented with rising-, falling-, and constant-amplitude sounds and were asked to detect change (increase or decrease) in sound amplitude while ignoring an irrelevant visual cue of a disk that grew, shrank, or stayed constant in size. Across two experiments, testing context was manipulated by varying the grouping of visual cues during testing, and cross-modal congruency showed independent perceptual and decision-related effects. Whereas a change in testing context greatly affected criterion shifts, cross-modal effects on perceptual sensitivity remained relatively consistent. In general, participants were more sensitive to increases in sound amplitude and less sensitive to sounds paired with dynamic visual cues. As compared with incongruent visual cues, congruent cues enhanced detection of amplitude decreases, but not increases. These findings suggest that the relative contributions of perceptual and decisional processing and the impacts of these processes on cross-modal interactions can vary significantly depending on asymmetries in within-modal processing, as well as consistencies in cross-modal dynamics.

Keywords Cross-modal interactions · Motion in depth · Looming · Audiovisual · Visual capture

E. H. Liu (✉)
Department of Psychology, University at Buffalo,
The State University of New York,
312 Park Hall,
Buffalo, NY 14260, USA
e-mail: hliu6@buffalo.edu

E. Mercado III
Department of Psychology, University at Buffalo,
The State University of New York,
350 Park Hall,
Buffalo, NY 14260, USA

B. A. Church
Department of Psychology, University at Buffalo,
The State University of New York,
373H Park Hall,
Buffalo, NY 14260, USA

Successful exploration of the environment often involves evaluating and comparing information from multiple modalities and along multiple dimensions. Cross-modal interactions in perception have been studied using behavioral and physiological measures (Alais, Newell, & Mamassian, 2010; Calvert, Spence, & Stein, 2004). Intersensory conflicts are often introduced in behavioral studies of cross-modal interaction (Soto-Faraco, Spence, Lloyd, & Kingstone, 2004b; Spence, 2011). Participants are asked to attend to signals from a target modality (e.g., audition) while ignoring potentially conflicting signals from an irrelevant modality (e.g., vision). In general, these studies demonstrate that the integration of multimodal stimuli depends on the temporal (Alais & Burr, 2003), spatial (Meyer, Wuerger, Rohrbein, & Zetzsche, 2005), semantic (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Meyer, Wuerger, & Greenlee, 2011), and synesthetic (Gallace & Spence, 2006) relationships between the target and irrelevant modalities. If the two signals provide conflicting information, the incongruent irrelevant cue typically causes the target signal to be perceived in a way that differs from when it is presented in isolation (Alais & Burr, 2004b; Ernst & Bülthoff, 2004; Helbig & Ernst, 2007; Sanabria, Soto-Faraco, & Spence, 2007a; Watanabe & Shimojo, 2001). In contrast, if the different modalities convey redundant information, the congruent irrelevant cue

typically facilitates perception of the target signal (Cappe, Thut, Romei, & Murray, 2009; Harrison, Wuerger, & Meyer, 2010; Laurienti et al., 2004; Soto-Faraco et al., 2004a, b; Zetzsche, Röhrbein, Hofbauer, & Schill, 2002). Behaviorally, such facilitation sometimes leads to faster responses, greater perceptual sensitivity, reduced perceptual ambiguity, or stronger orienting to cross-modal target stimuli.

Strong cross-modal interaction exists between the auditory and visual modalities (Alais et al., 2010; Brooks, van der Zwan, Billard, Petreska, Clarke, & Blanke, 2007; Soto-Faraco, & Kingstone, 2004; Wuerger, Hofbauer, & Meyer, 2003). For example, when a sound is played in the dark, perception of the origin of the sound can be biased toward the location of a concurrent flash—the spatial ventriloquist illusion (Alais & Burr, 2004b; Bertelson & Aschersleben, 1998; Bertelson & Radeau, 1976). A similar effect has been found when people perceive dynamic acoustic events. For example, in studies investigating cross-modal dynamic capture, listeners are asked to attend to a stream of auditory apparent motion while ignoring a stream of visual apparent motion. Listeners report perceiving motion differently depending on whether synchronous visual and auditory streams moved in the same or different directions (Soto-Faraco et al., 2004a; Soto-Faraco, Spence, & Kingstone, 2005). On about half of the trials on which sounds and lights moved in opposite directions, participants reported hearing the sounds moving in the direction of the visual motion. Audition is also shown to affect visual perception. For instance, when a single flash is presented with multiple beeps, participants falsely perceive that there is more than one flash (Shams, Kamitani, & Shimojo, 2000). Other recent work has shown that the availability of auditory cues impacts the detectability and discriminability of biological motion in point light displays (Brooks et al., 2007; Saygin, Driver, & de Sa, 2008; Schouten, Troje, Vroomen, & Verfaillie, 2011).

Perceptual and decision-related contributions to audiovisual processing

Previous studies have established clear interactions between auditory and visual processing. It is less clear, however, to what degree these past behavioral demonstrations of cross-modal modulation reflect decision-related/postperceptual biases versus perceptual processes. First, the observed interference from an incongruent distractor in a second modality may have arisen from listeners' confusion or response interference, rather than from the integration of audiovisual information at the perceptual level. When the target feature to be judged (e.g., whether a sound is moving to the right or left) shares the spatiotemporal properties with visual distractors, such stimulus–response compatibility can cause participants to inadvertently respond to the distractors

(Gallace & Spence, 2006; Sanabria, Spence, & Soto-Faraco, 2007b; Simon & Berbaum, 1990). In addition to response interference, it is also possible that performance may be affected either by a listener's conscious use of strategies (e.g., selectively attending to a particular spatial region) or by subconscious shifts in their response criterion, referred to as *response bias* (Alais & Burr, 2004a; Bertelson, 1998; Meyer & Wuerger, 2001; Sanabria et al., 2007a, b; Soto-Faraco, Kingstone & Spence, 2003; Talsma, Senkowski, Soto-Faraco & Woldorff, 2010; Welch & Warren, 1980). Cross-modal effects in these kinds of decision-related processes suggest that information from the irrelevant modality influences general response selection and execution, rather than (or in addition to) directly affecting perception (see also Spence, 2011).

Researchers have attempted to address the question of which cross-modal interactions are the result of perceptual or decision-related processes in several ways. First, studies exploring audiovisual motion capture have shown that performance interference from an incongruent visual stimulus takes place only when the auditory and visual cues are presented in synchrony (Soto-Faraco et al., 2004a). When one cue is presented after the other, interference disappears. Additionally, when listeners were asked to give confidence ratings after each response, even after high-confidence responses, they often reported hearing the sound moving in the direction of the visual motion (Soto-Faraco et al., 2003). These findings suggest that visual capture of auditory lateral motion may be a perceptual effect, rather than the result of confusion or response interference (Soto-Faraco & Kingstone, 2004).

Other studies have untangled perceptual and decisional contributions to audiovisual processing by combining methodologies from different experimental tasks to make the cross-modal conflict irrelevant to performance. For example, both spatial ventriloquism and audiovisual dynamic capture have been examined with a psychophysical staircase procedure (Bertelson & Aschersleben, 1998; Soto-Faraco et al., 2005). The results of these studies have shown that both phenomena occur even when response bias or strategy is controlled for by the experimental context. Other researchers have isolated perceptual from decision-related effects statistically. For example, Choe, Welch, Gilford & Juola (1975) investigated spatial ventriloquism by using signal detection measures. Listeners were presented with synchronous and asynchronous flashes and tones and were asked to judge whether the tone and flash came from the same location or not. Stimulus synchrony was found to affect the decision criterion, but not perceptual sensitivity. Consequently, the authors concluded that spatial ventriloquism was a result of changes in response bias, rather than of perceptual distortion caused by the concurrent flash (for counterclaims, see Bertelson & Radeau, 1976).

The role of coherence and correspondence in dynamic audiovisual processing

Whereas earlier studies often focused on determining whether cross-modal interactions were the result of perceptual integration or postperceptual selection and decisions, the current consensus is that both perceptual and decisional processes contribute to cross-modal processing (Alais et al., 2010; Sanabria et al., 2007b; Soto-Faraco, Spence & Kingstone, 2004a, b). Consequently, modern studies often focus less on the question of whether perceptual integration occurs during cross-modal processing of events and more on the issue of what conditions determine the degree to which information from multiple modalities is integrated. For example, Meyer and colleagues (Meyer et al., 2005) found that audiovisual integration of lateral motion reduced detection thresholds beyond what would be predicted statistically only when apparent visual and auditory motion were spatiotemporally coherent (see also Harrison et al., 2010). They concluded that effective perceptual integration of audiovisual motion requires all the dynamic spatial cues naturally available for localizing a moving target.

Recent studies of audiovisual perception of biological motion similarly have proposed that naturally relevant (and/or familiar) dynamics facilitate perceptual integration and detection (Brooks et al., 2007; Saygin et al., 2008; Schouten et al., 2011). These studies have shown that adding auditory information that is consistent (congruent) with a possible interpretation of biological motion within patterns of moving lights increases the likelihood that a person will detect biological motion. Again, both temporal coherence (e.g., phase locking of step sounds to leg movements in a point light display) and consistency with naturally occurring events (patterns corresponding to upright walkers vs. inverted walkers) proved to be crucial for enhancing detection of dynamic events (Saygin et al., 2008).

In the studies described above, increasing the spatiotemporal coherence of dynamic stimuli across modalities increased perceptual performance. Another factor that influences cross-modal processing relates to intrinsic or acquired links between modalities, sometimes referred to as *cross-modal correspondences* (Martino & Marks, 2001; Spence, 2011). For example, in synesthesia, higher pitch tones are often associated with more angular and smaller-sized shapes (Martino & Marks, 2001). Even in individuals with more typical perceptual experiences, certain nonredundant cross-modal features can facilitate or disrupt processing (e.g., perception of brightness interacts with perception of loudness, and perception of pitch interacts with perception of elevation; reviewed by Spence, 2011). The impacts of audiovisual correspondence (in either congruent or incongruent contrasts) have typically been assessed through measures of performance in speeded classification tasks in which participants must classify a stimulus varying along one dimension (e.g., pitch), while ignoring an irrelevant stimulus in a second modality that is varying along a different dimension (e.g., size). In these tasks, simultaneous presentation of features varying along corresponding dimensions across modalities leads to faster responses during congruent trials and slower responses during incongruent trials. In other words, the effects of cross-modal correspondence on performance of perceptual tasks are generally comparable to those associated with variations in the spatiotemporal coherence of dynamic events: Congruent variations enhance performance, and incongruent variations interfere with (or have no effect on) performance.

The present study

In the present study, we examined how visual cues affect the perception of sounds with rising and falling amplitude. Participants were asked to detect a change in amplitude in the presented sound, regardless of the direction of change, while ignoring an irrelevant concurrent visual cue (a disk growing, shrinking, or not changing in size). These cues—changes in sound amplitude and the size of visual images—are the most salient cues to auditory and visual motion in depth (Regan & Beverley, 1978; Rosenblum, Carello & Pastore, 1987). Because both sounds with rising amplitude and visual images that grow in size are associated with an approaching/looming motion, whereas both falling-amplitude sounds and visual images that shrink in size are associated with a departing/receding motion, henceforth, we refer to auditory and visual cues that converge in the direction of change as *congruent* (e.g., rising-amplitude sounds and growing disks) and cues that diverge in the direction of change as *incongruent* (e.g., rising-amplitude sounds and shrinking disks). On the basis of past work examining the role of sound amplitude changes in perception of motion in depth (Cappe et al., 2009; Kitagawa & Ichihara, 2002; Kitajima & Yamashita, 1999; Maier, Neuhoff, Logothetis & Ghazanfar, 2004; Zetzsche et al., 2002) and studies showing cross-modal integration in processing of lateral auditory motion (Harrison et al., 2010; Meyer et al., 2005; Sanabria et al., 2007b; Wuerger et al., 2003), we predicted that congruent changes in visual and auditory cues would enhance detection of changes in sound amplitude, whereas incongruent changes in these cues would interfere with detection of auditory changes. In addition, because processing asymmetries have been demonstrated in sounds with rising and falling amplitude (Hall & Moore, 2003; Maier & Ghazanfar, 2007; Neuhoff, 1998), congruency effects were examined separately for rising and falling sounds.

Building on previous studies of audiovisual perception of differences in lateral motion (Sanabria et al., 2007b), we designed our experiments so that the contributions of postperceptual processes to cross-modal interactions could be identified. To isolate the potential contributions of response bias, performance was analyzed using signal detection analysis, separating the detection data into the sensitivity index (d') and response bias (c) in the measures (Macmillan & Creelman, 1991). If visual cues directly impact the perception of amplitude-modulated sounds, congruency should affect d', but not c. In contrast, if a cross-modal congruency effect is exclusively the result of a shift in participants' decision criterion, congruency should affect c, but not d'.

Additionally, having participants attempt to detect a simple change in sound amplitude, independently of the direction of change, made it possible to account for the role of response interference in performance (Zetzsche et al., 2002). If cross-modal interactions between visual and auditory cues merely reflect listeners' response interference (or confusion), performance should not be disrupted when dynamic sounds are paired with an incongruently changing disk in this task, because both visual and auditory changes map onto the same response. In contrast, if pairing amplitude-modulated sounds with an incongruently growing or shrinking disk does interfere with performance, cross-modal interactions between visual and auditory cues must reflect some process other than response interference.

Finally, to explore the role of strategy or context in audiovisual processing of changes in sound amplitude, we conducted a second experiment that manipulated the distribution of trials containing different visual cues. In Experiment 1, trials involving growing and shrinking disks were blocked together, whereas trials involving static disks were blocked with the unimodal/no-disk baseline condition in which the sounds were classified without visual distractors. In Experiment 2, the trials representing the three audiovisual conditions were all blocked together, whereas trials for the unimodal/no-disk baseline condition were presented separately. If our sensitivity index d' measures the perceptual effect of cross-modal integration unconfounded by the use of strategy or differences in context, this change in the distribution of available visual cues between experiments should affect only the overall pattern of c, but not d'.

Experiment 1

Various perceptual and decision-related processes may all contribute to the interaction between stimuli from different modalities. To study how visual information affects the perception of auditory changes, one must account for potential decision-related confounds such as response biases, response interference, or even the use of strategy. In Experiment 1, participants were asked to detect whether or not the amplitude of the presented sounds changed, while ignoring a concurrent disk that grew, shrank, or remained constant in size. This change detection task helped us to rule out interference resulting from response conflict as a cause of the cross-modal interaction, by eliminating the direct mapping between an incongruent visual cue and conflicting response options. To further dissociate perceptual and decision-related contributions to cross-modal interactions, the data were analyzed in terms of d' and c within a signal detection framework. If concurrent visual cues influence only listeners' response bias, rather than making listeners more sensitive to changes in sound amplitude, congruency should affect c, but not d'. If concurrent visual cues are integrated to affect the perception of auditory changes, congruency should affect d'.

Method

Participants Fifty-one undergraduates with self-reported normal hearing and normal or corrected-to-normal vision were recruited for the experiment. Participants who missed more than 10 responses (3.1% of the trials) or performed at chance level or below (50% ± 11% for 80 trials) in the no-disk baseline condition were excluded. Thirty-five participants (21 males and 14 females, mean age = 19.23 years) were included in the analysis.[1] All received partial course credit in an introductory psychology course for their participation.

Apparatus and stimuli The stimuli were sine waves created in MATLAB 7.4 (Math Works, Natick, MA) and saved in wav format. The sounds were delivered binaurally via Audio-Technica ATH-M40fs stereo headphones. Visual stimuli were displayed on a 17-in. LCD monitor (HP Pavilion f70) with a resolution of 1,024 × 768 pixels at a refresh rate of 60 Hz. Stimulus presentation and response recording were controlled via DMDX (Forster & Forster, 2003) scripts. All auditory stimuli were 1,000 ms in duration, with a carrier frequency of 1 kHz. The carrier tones were sampled at 44.1 kHz and had 10-ms Hanning ramps at the onset and offset. Sound intensity was modulated by imposing an quadratic amplitude envelope on the waveforms (Maier & Ghazanfar, 2007). The falling-amplitude sounds were created by temporally reversing the rising-amplitude sounds. Two dynamic ranges were includ-

---

[1] Two and 14 participants were excluded by the missing trial and accuracy criteria, respectively (1 was excluded by both). This exclusion mainly reduced d' values. The significance patterns for d' and c were identical with or without exclusion. One bias became nonsignificant if all participants were included.

ed. From start to end, the sound intensity changed between 72 and 80 dB or between 66 and 80 dB sound pressure level (SPL, A-weighted). For each ramp size, a constant sound was created to match the global energy of the corresponding dynamic sounds. The visual stimuli were white disks on black backgrounds. The radius of each disk either grew or shrank at a constant rate for 1,000 ms from approximately 6.7° to 10° of visual angle.

Design and procedure The independent variables were sound type (rising, falling, and constant) and disk condition (growing, shrinking, constant, and no disk). Percent correct and response time (RT) were measured. We used a one-interval, two-alternative forced choice (2-AFC) task with 320 test trials. Half of the test trials contained a constant sound, and the other half contained a dynamic sound. Rising and falling sounds each made up half of the dynamic sounds. Within each type of sound, half had a greater and half had a smaller ramp size. The task had two blocks with an intervening break. One block contained sound presented without a disk or with a concurrent constant disk. The other block contained sound presented concurrently with either a growing or a shrinking disk. Each visual cue was presented with half of each type of sound in each block. Participants were instructed to determine whether the amplitude of the sound changed or not and were told to disregard variations in the visual cues. A schematic diagram of the experimental design is shown in Fig. 1.

At the beginning of each trial, a fixation cross appeared at the center of the screen and remained visible throughout the trial. Participants were told to fixate on the cross throughout the entire trial. A rising, falling, or constant target sound was presented 500 ms after the onset of the fixation cross. Participants responded by pressing keys labeled "change" or "no change" on a keyboard, depending on whether they detected an amplitude change (either louder or softer) in the target sound. At the beginning of each block, participants received eight practice trials. The practice trials encompassed all possible combinations of sound types and visual cues in the following block. The practice sounds (modulated between quiet and 80 dB) had a ramp size greater than that of the testing sounds (8 or 14 dB).

Participants were informed of the varying ramp sizes across trials and that they had 4 s to respond on each trial. They were instructed to guess if they were uncertain and that they would receive feedback after each response. We also emphasized the importance of fixating on the cross while performing the task, even if they thought that the visual cues interfered with their performance. Trial order was randomized within each block. Assignment of response keys and block order was counterbalanced between participants. Participants were debriefed upon completion of the task.

Results

The group-averaged proportions correct and RTs for correct responses are presented in Table 1 as a function of sound type and disk condition. The patterns of the two measures suggested that there was no speed–accuracy trade-off. To evaluate the effect of cross-modal integration, in further analyses, the dynamic visual cues were defined in terms of their congruency to the paired sounds (congruent, incongruent), rather than by their absolute change direction (growing, shrinking). The effects of visual cues on different sound types were tested using the sensitivity (d') and response bias (c) measures, as well as the RTs to the rising and falling sounds relative to the constant sound in the same disk condition. Specific predictions regarding d' and c were examined by planned contrasts. Unless otherwise specified, all statistical analyses noted henceforth were two-tailed and used an α level of .05. The joint α level of multiple post hoc comparisons was controlled by the Bonferroni correction, although all p values presented in the text are raw values without this correction. The

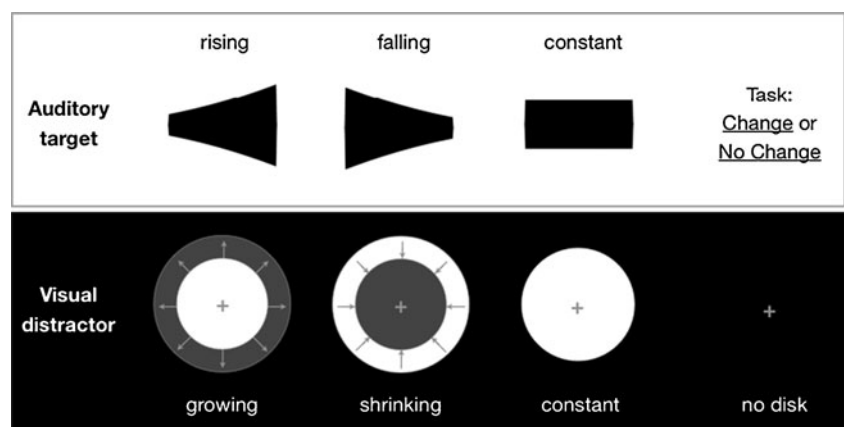**Fig. 1** Schematic diagram of the experimental design

Table 1 The mean proportions correct and response times as a function of sound type and disk condition in Experiment 1. Values in parenthesis represent standard errors of the means

Sound Type	Disk			
	Dynamic		Nondynamic	
	Growing	Shrinking	Constant	No Disk
Proportion Correct				
Rising	.82 (.03)	.71 (.03)	.81 (.02)	.82 (.03)
Falling	.71 (.03)	.71 (.03)	.69 (.02)	.69 (.03)
Constant	.66 (.02)	.77 (.02)	.80 (.02)	.77 (.02)
Response Time (ms)				
Rising	1,033.41 (42.79)	1,096.04 (50.17)	1,159.49 (42.80)	1,122.74 (47.25)
Falling	1,118.52 (48.00)	1,125.14 (42.78)	1,165.08 (43.04)	1,206.31 (43.30)
Constant	1,215.93 (50.02)	1,203.15 (47.81)	1,270.73 (41.75)	1,291.65 (50.93)

Greenhouse–Geisser correction for degrees of freedom was applied where appropriate. Effect sizes for ANOVAs and t-tests are reported in partial eta-squared (η_p^2) and Cohen's d, respectively.

Sensitivity (d') The mean d's for different sound types as a function of the visual cues are presented in Fig. 2a. A 2 (sound type: rising, falling)×4 (visual cue: congruent, incongruent, constant, no disk) repeated measures ANOVA revealed a significant main effect of sound type, $F(1, 34) = 13.15$, $p = .001$, $\eta_p^2 = .28$, showing that listeners were more sensitive to rising than to falling sounds regardless of the paired visual cue. The main effect of visual cue was also significant, $F(1.84, 62.46) = 6.10$, $p = .005$, $\eta_p^2 = .15$. Post hoc comparisons revealed that listeners were significantly less sensitive when sounds were paired with a dynamic disk (congruent or incongruent) than with a nondynamic disk (constant or no disk), $t(34) = 2.55$, $p = .016$, $d = 0.43$. In particular, relative to the no-disk condition, participants were less sensitive to dynamic sounds paired with incongruent disks, $t(34) = -2.81$, $p = .008$, $d = 0.47$. Differences in sensitivity between audiovisual and unimodal trials were not significant for sounds paired with congruent or constant disks, $|t s(34)| < 1.09$, $p s > .28$, $d s \leq 0.18$. Two planned contrasts were carried out comparing the congruency effect for rising and falling sounds, respectively, despite a lack of statistical significance for the sound type × visual cue interaction, $F(2.36, 80.36) = 2.04$, $p = .11$, $\eta_p^2 = .06$. We found that d' was greater when falling sounds were paired with a congruent rather than with an incongruent disk, $t(34) = 8.22$, $p < .001$, $d = 1.39$. Rising sounds, however, showed comparable d's regardless of whether the paired disk was congruent or incongruent, $t(34) = 1.07$, $p = .29$, $d = 0.18$. In addition, despite being enhanced by a congruent visual cue, d' was still numerically smaller for falling sounds than for rising sounds. These results suggest that sensitivity to amplitude change in falling sounds is affected more by the presence or absence of congruency across modalities.

Response bias (c) The mean cs for different sound types as a function of the visual cues are presented in Fig. 2b. The numbers were converted such that positive values represented tendencies to make the "change" response, and
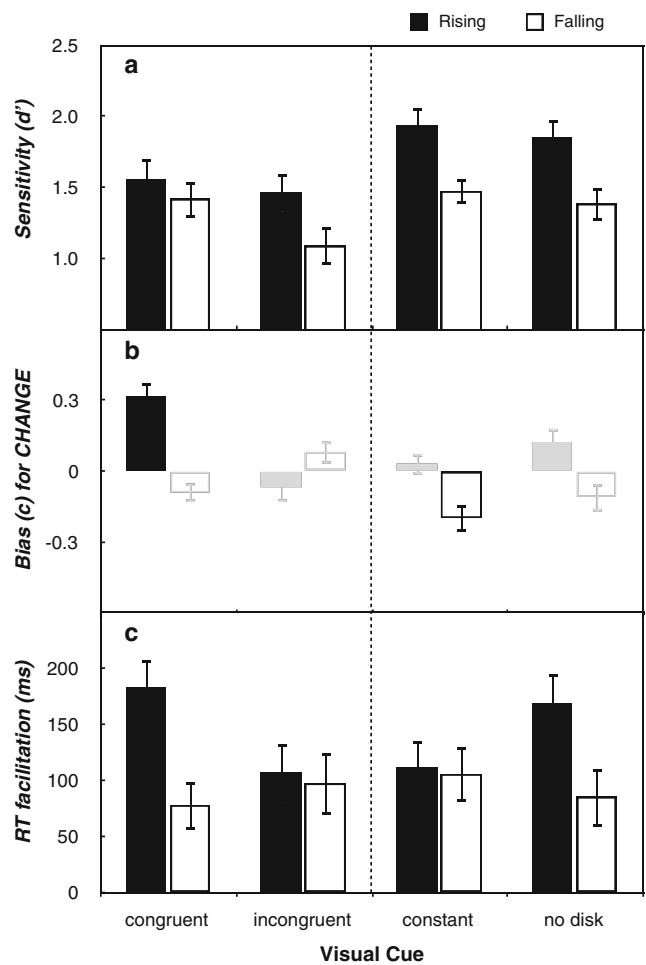

Fig. 2 Group-averaged **a** sensitivity (d′), **b** response bias (c) for *change* responses, and **c** response time (RT) facilitation (in milliseconds) measures as a function of sound type and visual cue in Experiment 1. Error bars represent the standard errors of the means

negative values represented tendencies to make the "no change" response. We conducted eight planned one-sample t-tests to examine whether a response bias was present (c significantly deviated from zero) in each condition. We found a positive bias when rising sounds were paired with a congruent disk, $t(34) = 5.48$, $p < .001$, $d = 0.93$, indicating that listeners were more liberal, making more false alarms when constant sounds were paired with a growing disk. We also found a negative bias when falling sounds were paired with a constant disk, $t(34) = 3.34$, $p = .002$, $d = 0.56$. This indicates that listeners were more conservative, making fewer hits when falling sounds were paired with a constant disk. No other test was significant.

RT facilitation The raw RTs for correct responses, as shown in Table 1, revealed that participants generally responded faster to sounds with changing amplitude than to constant sounds, even when they were not specifically instructed to make speeded responses. Although participants were not specifically instructed to make speeded responses, the present RT results demonstrate consistent patterns revealing how participants naturally respond to stimuli with different dynamics. To evaluate how natural response speed to different dynamic sounds was affected by the visual cues, we subtracted the raw RTs for the dynamic sounds from the raw RTs for the corresponding constant sounds in each disk condition (see Table 1). This measure is referred to as *RT facilitation* in the rest of the article. The resulting means as a function of sound type and visual cue are presented in Fig. 2c.

In general, listeners responded faster to dynamic sounds than to constant sounds. A 2 (sound type: rising, falling) × 4 (visual cue: congruent, incongruent, constant, no disk) repeated measures ANOVA revealed a significant effect of sound type, $F(1, 34) = 5.34$, $p = .027$, $\eta_p^2 = .14$, as well as of the sound type × visual cue interaction, $F(3, 102) = 3.77$, $p = .013$, $\eta_p^2 = .10$, but not of visual cue, $F(2.43, 82.68) < 1$. Overall, greater RT facilitation was found for rising sounds than for falling sounds, showing that relative to constant sounds paired with the same visual cue, listeners were faster to respond to rising than to falling sounds; RT facilitation was not significantly different for dynamic versus non-dynamic disks. Post hoc comparisons of the visual cue conditions showed no significant difference in RT facilitation for dynamic disk conditions, relative to the no-disk condition.

For direct comparison between RT facilitation and d', the significant interaction was further tested by two planned contrasts to assess the congruency effect separately for rising and falling sounds. Rising sounds showed greater RT facilitation when paired with a congruent versus an incongruent disk, $t(34) = 2.80$, $p = .008$, $d = 0.47$. In contrast, falling sounds demonstrated comparable RT

facilitation regardless of disk congruency, $|t(34)| < 1$. This pattern is consistent with the results reported by Cappe et al. (2009), who found that across multisensory conditions, speeded response was facilitated the most for audiovisual stimuli that increased in both amplitude and size. Additionally, it suggests that there is a dissociation between the effects of cross-modal congruency on RT facilitation (significant for rising but not for falling sounds) and on d' (significant for falling but not for rising sounds).

Discussion

In Experiment 1, participants were presented with rising, falling, or constant sounds and were asked to detect change in sound amplitude while ignoring any concurrent visual cues. To explore how cross-modal congruency affects perception of auditory changes, we examined participants' sensitivity, response bias, and RT for different audiovisual combinations. We found that independent perceptual and decision-related processes affected cross-modal processing of sound changes. This finding is consistent with previous research showing that multiple processes independently affect audiovisual interactions in the perception of lateral motion (Alais et al., 2010; Sanabria et al., 2007b; Soto-Faraco et al., 2004a, b) and static events (Bertelson & de Gelder, 2004; Choe et al., 1975). Sanabria et al. (2007b), for example, asked listeners to detect or classify whether an auditory apparent motion stream moved in a predefined target direction (left or right). In different testing blocks, the auditory streams were paired with concurrent visual streams that moved in either the same or the opposite direction. Signal detection analysis revealed that sensitivity to the auditory target direction was significantly reduced in cross-modal conditions, relative to unimodal conditions, and that this effect was independent of the relative direction of the visual distractor. However, shifts in response bias varied systematically as a function of the relative direction of the visual distractor. These results demonstrated a dissociation between the perceptual and decisional components involved in the processing of audiovisual motion. Their replication of this same pattern in two different tasks suggested that the perceptual and decision-related processes were largely independent.

Our results revealed different effects of congruency on the detectability of rising and falling sounds. Specifically, we found that congruency between auditory and visual cues affected sensitivity to falling but not to rising sounds. In contrast, congruency shifted the decision criterion and sped up responses for rising but not for falling sounds. These findings rule out the possibility that the cross-modal interaction found in the present experiment was a result of response interference, because if this were the sole source of audiovisual interference or facilitation, we should not

have observed any congruency effects. Furthermore, our findings support the notion that cross-modal processing of changes in sound intensity takes place, at least in part, at the perceptual level. This interpretation is suggested by the finding that congruency enhanced listeners' ability to detect falling amplitude in sounds.

The finding that detection of falling amplitude was facilitated by cross-modal congruency more than was detection of rising amplitude is consistent with the law of inverse effectiveness of multimodal integration. This law was proposed to encapsulate the neurophysiological observation that neural responses invoked by a unimodal stimulus are more strongly facilitated by congruent information from a second modality when the primary signal is weak (Meredith & Stein, 1983; Stein & Meredith, 1993). In audiovisual processing, this law suggests that redundant information from an irrelevant modality can benefit the perception of the target signal to the extent that the target signal is weak in the first place. Previous studies have shown that sounds with rising amplitude evoke greater neural responses than do falling sounds (Maier, Chandrasekaran & Ghazanfar, 2008; Maier & Ghazanfar, 2007), consistent with our finding of faster and more accurate detection of rising sounds across conditions. Given that processing of falling sounds was relatively inferior to processing of rising sounds, inverse effectiveness would predict a greater facilitation of detection for falling sounds when congruent cross-modal cues are available. This is consistent with our observation that the detection of falling sounds benefited more from the presence of congruent visual cues.

In addition to evidence suggestive of audiovisual integration, we also found evidence that decision-related processes contribute to cross-modal interactions in our task. In particular, audiovisual congruency shifted listeners' decision criterion for detecting intensity change in rising sounds (bias when paired with a congruent disk and no bias when paired with an incongruent disk), but not in falling sounds (no bias when paired with either a congruent or an incongruent disk). Furthermore, the finding that sensitivity was generally lower when moving disks were present could be interpreted as evidence of a general disruption of auditory processing associated with the presentation of changing visual stimuli. Given that participants were instructed to ignore visual stimuli, such disruption could reflect the fact that dynamically changing disks are more salient and, therefore, more difficult to ignore (Rapp & Hendel, 2003).

In this experiment, we controlled for the dynamic range of the auditory and visual stimuli, such that rising, falling, and constant sounds always started at different amplitudes. Likewise, growing, shrinking, and constant disks always started from different sizes. Potentially, participants could

have used this information, instead of dynamic stimulus properties, to distinguish the presented stimulus types. However, our data suggest that variations in the initial states of the stimuli did not contribute to our results. Given that the falling sounds were most consistently mapped by initial sound amplitude, if participants responded solely on the basis of this information, they should have been most accurate at responding to falling sounds. However, our data showed the opposite pattern. Among all sound types, participants were least accurate with the falling sounds (see Table 1). Therefore, the effect of initial amplitude would actually need to be overcome in order to produce the present results. Similarly, the dynamic disks affected sound amplitude perception differently than did the constant disks. It is not clear how this effect could be the result of differences in initial disk size, such that the largest and the smallest disks systematically affected amplitude change detection in a way that midsized disks did not. In an earlier study involving similar bimodal stimulus presentations, Cappe et al. (2009) presented sounds with a constant initial amplitude, as well as sounds ramping between comparable dynamic ranges (as in the present experiment). They found that both sets of sounds yielded cross-modal effects comparable to those observed in the present experiment, further suggesting that the specific features of initial sound amplitude are not driving these effects.

In summary, the present experiment showed that after controlling for possible effects of response interference by use of the nondirectional change detection task, the manipulation of cross-modal congruency revealed dissociable effects on sensitivity, decisional criterion, and RT facilitation for different types of sounds. To investigate the potential contributions of strategy use or context to cross-modal interactions during the perception of changes in sound amplitude, in Experiment 2, we modified the experimental context by changing the distribution of trials involving visual cues. We expected that this manipulation would affect only decision-related contributions to audiovisual processing (e.g., response bias) and would have no impact on perceptual contributions (e.g., cross-modal enhancement of sensitivity).

Experiment 2

Experiment 1 showed that when listeners tried to detect changes in sound amplitude while ignoring a concurrent visual cue, the congruency between the visual and auditory changes affected both their perceptual sensitivity to the sounds and their response bias. Because the growing disks were presented exclusively in blocks of trials that also included shrinking disks, and because constant disks were always blocked with no-disk trials, the varying visual cues

in any given block mapped onto only one of the two response options. To promote a strategy change, in Experiment 2, we modified the experimental context by blocking the three audiovisual conditions (growing, shrinking, and constant disk plus sound) together, while presenting the unimodal/no-disk baseline condition by itself. Participants performed the same change detection task as in Experiment 1. With all three disk conditions mixed together, the varying visual cues mapped onto both possible response options. As in Experiment 1, we dissociated perceptual and decision-related contributions to performance using d' and c. We predicted that any strategy shift resulting from this change in experimental context should affect only decision-related contributions to cross-modal interaction. Therefore, in Experiment 2, we expected to observe a pattern of cross-modal effects on d' comparable to those observed in Experiment 1, but a different pattern in response biases (c).

Method

Participants Fifty undergraduates with self-reported normal hearing and normal or corrected-to-normal vision were recruited in the experiment. Those who missed more than eight responses (3.1% of the trials) or performed at chance level or below (50% ± 12% for 64 trials) in the no-disk baseline condition were excluded, leaving 36 participants (19 males and 17 females, mean age = 19.28) in the analysis.[2] All received partial course credits in an introductory psychology course for their participation.

Apparatus and stimuli The apparatus and stimuli used in Experiment 2 were identical to those in Experiment 1.

Design and procedure The design and procedure used in Experiment 2 were identical to those in Experiment 1, except that all trials in the no-disk baseline condition were blocked together, whereas the other three visual cues were intermixed in the rest of the experiment. We counterbalanced the order in which participants received the baseline and the mixed-disk blocks. This blocking of trials involving visual cues was different from that in Experiment 1, where trials with no disk were blocked with trials with constant disks and growing disks were blocked with shrinking disks. Attempting to decrease participant fatigue, we also reduced the number of trials proportionally for all combinations of sound type and disk condition from 320 in Experiment 1 to 256 in Experiment 2.

---

[2] Four and 11 participants were excluded by the missing trial and accuracy criteria, respectively (1 was excluded by both). This exclusion mainly reduced d' values. The significance patterns for d' and c were identical with or without exclusion.

Results

The proportions correct and RTs for correct responses are presented in Table 2 as a function of sound type and disk condition. As in Experiment 1, the data suggest no speed–accuracy trade-off. The effect of cross-modal integration was examined by use of the sensitivity (d') and response bias (c) measures, as well as RT facilitation for rising and falling sounds. Specific predictions regarding d' and c were examined by planned contrasts.

Sensitivity (d') The mean d's as a function of sound type and visual cues are presented in Fig. 3a. A 2 (sound type: rising, falling) × 4 (visual cue: congruent, incongruent, constant, no disk) repeated measures ANOVA revealed a significant main effect of sound type, $F(1, 35) = 22.38$, $p < .001$, $\eta_p^2 = .39$, showing that listeners were more sensitive to rising sounds regardless of the paired visual cue. The main effect of visual cue was also significant, $F(2.38, 83.26) = 9.60$, $p < .001$, $\eta_p^2 = .22$. Post hoc comparisons revealed that, as compared with the no-disk baseline, listeners were less sensitive when sounds were paired with either incongruent or congruent changes in disk size, $|ts(35)| > 3.90$, $ps < .001$, $ds > 0.65$. When sounds were paired with a constant disk, this tendency was not significant, $t(35) = -1.94$, $p = .060$, $d = 0.32$. More generally, the presentation of dynamic disks reduced sensitivity to dynamic sounds, relative to nondynamic disks, $t(35) = 24.23$, $p < .001$. We also found a significant sound type × visual cue interaction, $F(3, 105) = 3.26$, $p = .025$, $\eta_p^2 = .09$. Planned contrasts revealed greater d's when falling sounds were paired with congruent rather than incongruent changes in disk size, $t(35) = 7.98$, $p < .001$, $d = 1.33$. However, rising sounds showed comparable d's regardless of the congruency of the paired disk, $|t(35)| < 1$. As in Experiment 1, we found a congruency effect for falling sounds, but not for rising sounds. Overall, despite variations in absolute magnitude, the pattern of cross-modal effects on d' replicated those observed in Experiment 1. This finding suggests that the context change between the two experiments—and potentially, a different response strategy adopted by the listeners—did not qualitatively alter how cross-modal congruency impacted a listener's perceptual sensitivity to detecting changes in sound amplitude.

Response bias (c) The mean cs as a function of sound type and visual cue are shown in Fig. 3b; the values were converted such that positive and negative values correspond to tendencies to make "change" and "no change" responses, respectively. We conducted eight planned one-sample t-tests to examine whether a response bias was present in each condition. We found that when sounds were presented in isolation, listeners showed a positive bias for rising sounds,

Table 2 The mean proportions correct and response times as a function of sound type and disk condition in Experiment 2. Values in parenthesis represent standard errors of the means

Sound Type	Disk			
	Dynamic		Nondynamic	
	Growing	Shrinking	Constant	No Disk
Proportion Correct				
Rising	.85 (.03)	.78 (.03)	.74 (.03)	.88 (.02)
Falling	.72 (.03)	.70 (.03)	.59 (.03)	.73 (.03)
Constant	.64 (.02)	.74 (.03)	.82 (.03)	.78 (.02)
Response Time (ms)				
Rising	1,043.82 (38.25)	1,103.91 (51.27)	1,110.72 (42.68)	1,127.70 (45.14)
Falling	1,093.40 (38.95)	1,093.08 (40.05)	1,116.29 (40.97)	1,230.46 (43.08)
Constant	1,221.33 (43.57)	1,207.18 (42.00)	1,139.47 (39.78)	1,293.64 (45.74)

$t(35) = 4.27$, $p < .001$, $d = 0.71$, but they showed no bias for falling sounds. This finding suggests that, as in Experiment 1, listeners were more liberal at reporting detection of changes in rising sounds.

Fig. 3 Group-averaged **a** sensitivity (d'), **b** response bias (c) for *change* responses, and **c** response time (RT) facilitation (in milliseconds) measures as a function of sound type and visual cue in Experiment 2. Error bars represent the standard errors of the means

We also found a positive bias when rising sounds were paired with a congruent growing disk, $t(35) = 8.39$, $p < .001$, $d = 1.40$, and when falling sounds were paired with an incongruent growing disk, $t(35) = 2.95$, $p = .006$, $d = 0.50$. These findings show that listeners were more liberal at reporting change detection in sounds whenever the sounds were paired with a growing disk. In contrast, listeners showed a negative bias when both rising and falling sounds were paired with a constant disk, $|ts(35)| > 3.81$, $ps < .001$, $ds > 0.64$, indicating that they were more conservative about saying that they detected an amplitude change when a constant disk was present. No other test was significant. Overall, listeners demonstrated a response bias in more conditions (five out of eight conditions) than in Experiment 1 (two out of eight conditions). This suggests that the contextual change in the blocking of visual cues led to different shifts in decisional criteria that may have resulted from differences in strategy.
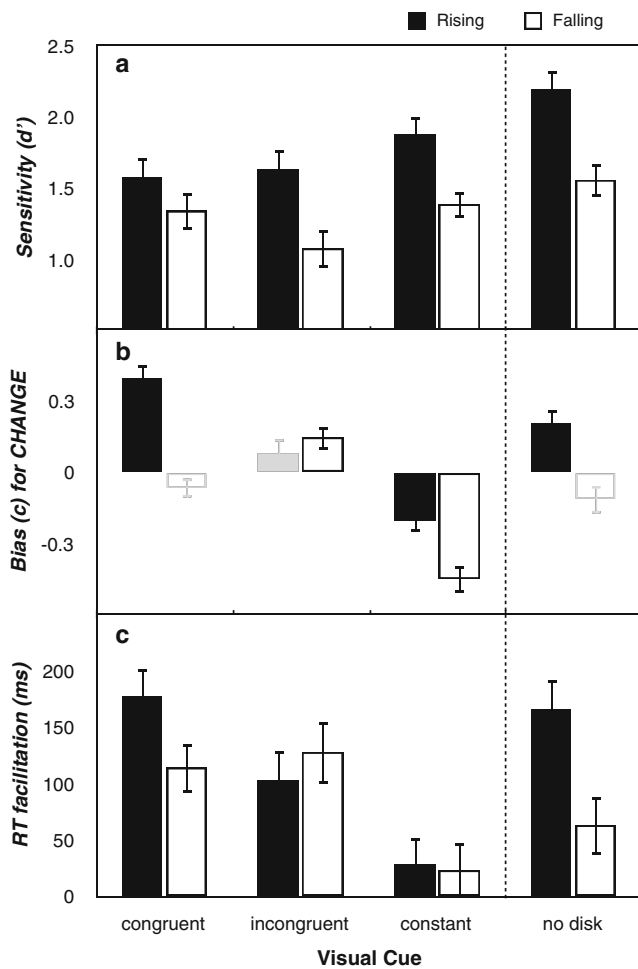
RT facilitation RTs to the dynamic sounds were subtracted from those to the corresponding constant sound in each disk condition to calculate RT facilitation (see Table 2). The means are presented in Fig. 3c as a function of sound type and visual cue. A 2 (sound type: rising, falling) × 4 (visual cue: congruent, incongruent, constant, no disk) repeated measures ANOVA revealed a significant effect of visual cue, $F(2.31, 80.81) = 9.06$, $p < .001$, $\eta_p^2 = .21$, as well as of the sound type × visual cue interaction, $F(2.44, 85.50) = 4.14$, $p = .013$, $\eta_p^2 = .11$. The main effect of sound type trended but did not reach significance, $F(1, 35) = 3.28$, $p = .079$, $\eta_p^2 = .09$. Post hoc comparisons showed that, relative to the no-disk baseline, participants responded to the dynamic sounds slower when sounds were paired with constant disks, $|t(35)| = 2.95$, $p = .006$, $d = 0.49$. However, when sounds were paired with dynamic disks, response to dynamic sounds was not significantly faster than the no-disk baseline, $|ts(35)| < 1.31$, $ps > .198$, $ds < 0.22$. Responses to sounds paired with dynamic disks were faster than

responses to sounds paired with nondynamic disks, $t(35) = 3.30$, $p = .002$.

For a direct comparison between the d' and RT facilitation measures, the significant interaction was further examined by two planned contrasts. Replicating the findings in Experiment 1 and Cappe et al. (2009), RT facilitation was greater when rising sounds were paired with a congruent versus an incongruent disk, $t(35) = 2.57$, $p = .014$, $d = 0.43$. In contrast, falling sounds showed comparable RT facilitation regardless of the direction of the paired visual changes, $|t(35)| < 1$. In other words, response speed to dynamic sounds was facilitated the most for audiovisual stimuli involving increases in amplitude and size. This finding again indicates a dissociation between the effects of cross-modal congruency on RT facilitation (significant for rising but not for falling sounds) and on d' (significant for falling but not for rising sounds).

To further explore how contextual change affected performance, we compared how values of dependent measures changed from Experiment 1 to Experiment 2. The differences as a function of sound type and visual cue are presented in Table 3, which shows that the greatest difference in RT facilitation across experiments for both rising and falling sounds was found in the constant disk condition (the condition that was blocked differently in Experiment 2). The greatest differences in c between experiments were also found in the constant disk condition. In contrast, for d', the greatest differences occurred in the no-disk baseline condition. These patterns of differences across experiments imply that the impact of testing context (and possibly strategy selection) on RT facilitation is more similar to its effect on response bias than to its effect on perceptual sensitivity.

Table 3 Magnitude change in d', c, and response time (RT) facilitation from Experiment 1 to Experiment 2 as a function of sound type and disk condition

Sound Type	Disk			
	Dynamic		Nondynamic	
	Growing	Shrinking	Constant	No Disk
Perceptual Sensitivity (d')				
Rising	0.021	0.173	-0.053	0.347
Falling	-0.074	-0.006	-0.082	0.177
Response Bias (c)				
Rising	0.079	0.149	-0.234	0.083
Falling	0.026	0.066	-0.249	-0.002
RT Facilitation (ms)				
Rising	-8.427	-5.890	50.760	-2.971
Falling	34.054	27.106	50.771	-22.163

Discussion

The grouping of visual conditions was changed from Experiment 1 to Experiment 2 in order to explore whether changes in context or strategy could influence measures of bimodal interactions in the perception of changes in sound amplitude. In Experiment 1, sounds paired with growing or shrinking disks were blocked together, and sounds paired with either a constant disk or no disk were blocked together. In Experiment 2, sounds paired with growing, shrinking, and constant disks were blocked together, whereas the diskless baseline condition was presented separately. Not surprisingly, isolating the no-disk trials and, thus, increasing the variability in visual cues in the bimodal block impacted the effect each cue had on perceptual sensitivity, relative to the sound-only baseline. Nevertheless, the effect of congruency on the sensitivity to dynamic sounds remained comparable to that in Experiment 1. Despite the change in trial distribution, cross-modal congruency again affected sensitivity to falling but not to rising sounds. On the other hand, listeners demonstrated more variations in response bias in Experiment 2 than in Experiment 1. This indicates a different pattern of shifts in decision criteria in Experiment 2, potentially induced by different response strategies adopted by the participants because of differences in the testing contexts.

The effects of testing context on RT facilitation were more mixed than its effects on either d' or c. On the one hand, the overall results of RT facilitation revealed similar patterns across the two experiments, and the asymmetry in the congruency effect was also replicated. This suggests that, like the d's, RT facilitation might be relatively unaffected by the change in testing context. On the other hand, the change in testing context appeared to slow down the faster responses to rising sounds. Furthermore, RT facilitation changed in ways that were more similar to measures of response bias than to measures of perceptual sensitivity. In summary, testing context affected the decision-related measure (c), but not the perceptual measure (d') of listeners' change detection performance, and the effect of testing context on a listener's response speed was complex.

General discussion

To study cross-modal interaction, it is important to tease apart processing at the decisional and perceptual levels (Meyer & Wuerger, 2001; Sanabria et al., 2007b; Soto-Faraco & Kingstone, 2004; Soto-Faraco et al., 2003). Here, we combined methodological and statistical approaches to dissociate perceptual and decision-related processes in change detection of sound amplitude. In two experiments,

listeners were asked to attend to amplitude-modulated sounds and to ignore a concurrent visual cue. Response interference was controlled by mapping dynamic auditory and visual cues onto the same response key. Perceptual and remaining decision-related effects were further dissociated using signal detection analyses and by varying the distribution of trial types. We found strong evidence for coexisting perceptual and decisional processes independently affecting cross-modal interaction. Even after controlling for response interference, participants' responses were not bias free. On the contrary, they responded more liberally or conservatively depending on the testing context and the type of visual cue paired with the sounds. In contrast, visual effects on auditory sensitivity remained consistent regardless of testing context. Our findings complement previous research demonstrating independent perceptual and decisional influences coexisting in the audiovisual perception of lateral motion (Bertelson & de Gelder, 2004; Choe et al., 1975; Sanabria et al., 2007b; Schirillo, 2011) and static audiovisual events (Bertelson & de Gelder, 2004; Choe et al., 1975; Sanabria et al., 2007b; Schirillo, 2011). In addition, they raise new questions about how changes in auditory and visual cues interact. Two interpretations are considered below to account for these results: (1) Participants process acoustic and visual cues as independent indicators of motion in depth, and (2) visual cues differentially disrupt participants' ability to attend to auditory cues depending on the cross-modal correspondence of the audiovisual dimensions.

Audiovisual motion in depth

Change rates in sound amplitude and visual image size are the most reliable auditory and visual cues to motion in depth (Regan & Beverley, 1978; Rosenblum et al., 1987) that covary when a signal source moves closer or away from an observer (Zetzsche et al., 2002). Looming sounds (typically, an increase in amplitude) can serve as a warning of impending contact (Bach, Schachinger, Neuhoff, Esposito, Di Salle, Lehmann & Seifritz, 2008; Ghazanfar, Neuhoff & Logothetis, 2002) and are more likely to speed up responses, bias attention and change distance estimates (Ghazanfar et al., 2002; Maier et al., 2004; Neuhoff, 1998, 2001), and induce stronger neural activation than are receding or constant sounds (Bach et al., 2008; Maier et al., 2008; Maier & Ghazanfar, 2007; Romei, Murray, Cappe & Thut, 2009; Seifritz et al., 2002). Although frequency spectrum, reverberant energy, and interaural differences also vary with sound motion in depth (Hall & Moore, 2003; Zahorik, 2002), modulating sound amplitude is sufficient to produce the response biases associated with looming sounds (Bach, Neuhoff, Perrig & Seifritz, 2009) and activate cortical areas known to respond to auditory motion

(Seifritz, Neuhoff, Bilecen, Scheffler, Mustovic, Schachinger & Di Salle, 2002).

Although our participants were not specifically instructed to report on whether they perceived sounds or images to be moving, it is likely that some stimuli induced perception of motion in depth. In a similar study, Cappe et al. (2009) found that participants rated amplitude-modulated sounds as lower and dynamic audiovisual stimuli as higher on subjective motion. It is important to note, however, that Cappe et al. instructed participants to attend to both visual and auditory changes, whereas our participants were instructed to ignore visual changes. Consistent with their findings, in postexperiment debriefings, several participants in our study reported having a stronger sensation of motion when congruently changing disks were presented with sounds. A few prior studies have also reported interactive auditory and visual cues associated with motion in depth (Cappe et al., 2009; Liu, Church, & Mercado, 2008; Maier et al., 2004; Schouten et al., 2011; Zetzsche et al., 2002).

Our data on response criteria and speed parallel past findings on perceptual bias for looming sounds (Neuhoff, 1998, 2001; Rosenblum et al., 1987; Rosenblum, Wuestefeld & Saldana, 1993). Audiovisual looming stimuli have been shown to attract more attention (Maier et al., 2004), to be responded to faster, and to be rated as having stronger subjective motion (Cappe et al., 2009). Consistent with these findings, we showed that congruent visual dynamics facilitated natural response speed only to rising sounds and that participants were more liberal at detecting rising sounds. In addition, we found that shifts in criterion varied with the strategies adopted in different testing contexts, possibly reflecting the greater attention to sounds with rising amplitude demonstrated in previous research (Ghazanfar et al., 2002; Neuhoff, 1998).

Across experiments, we also found consistent cross-modal effects on perceptual sensitivity showing asymmetric processing of looming versus receding stimuli. Participants were more sensitive to rising than to falling sounds, except for when sounds were paired with congruent visual cues. Overall, they were less sensitive to auditory changes when sounds were paired with dynamic than with static or no visual stimuli. Other studies of audiovisual integration have noted that audiovisual stimuli differentially enhance or reduce sensitivity depending on whether both modalities show congruent changes (Sanabria et al., 2007b; Zetzsche et al., 2002). Also varying sound amplitude and visual image size, Zetzsche et al. found that congruent changes in auditory and visual cues reduced the threshold for change detection in either target modality beyond what was predicted by probability summation. Therefore, the present findings raise new questions about how these asymmetries relate to the processing of redundant or contradictory

inputs. As was noted earlier, one possible explanation is suggested by the law of inverse effectiveness in multimodal integration (Meredith & Stein, 1983; Schirillo, 2011; Stein & Meredith, 1993). Although the original studies demonstrated this in animal neurophysiology, a recent study directly showed this asymmetry in perception: The cross-modal benefits to the detectability of a target sound were inversely related to its unimodal signal strength, as defined by signal-to-noise ratios (Schirillo, 2011). Given that in the present study participants were generally less able to detect decreases in sound amplitude, the effects of visual cues on detection may have been more pronounced. Cappe et al. (2009) found that participants rated unimodal receding sounds as having the least movement among auditory, visual, and audiovisual stimuli moving in both directions in depth. The reduced "salience" of sounds with falling amplitude thus could increase the potential benefit gained from the integration of congruent visual cues.

Cross-modal correspondence

Regardless of whether participants perceived audiovisual stimuli as looming, receding, or stationary, it is clear that their attempts at responding only to auditory changes were unsuccessful. Visual cues impacted participants' responses differently depending on the dynamics of those cues. In particular, images of growing disks impacted performance differently from shrinking disks, and dynamic disks impacted performance differently from constant disks. These effects may be related to the naturalness of the audiovisual combinations (Zetzsche et al., 2002), to differences in the processing levels that are engaged by properties of the stimuli (Rapp & Hendel, 2003), or to differences in how distinct properties are mapped across modalities (Melara, Marks & Potts, 1993; Spence, 2011). In principle, changes in image size may map more readily onto changes in certain acoustic features than in others. If cross-modal interference is viewed as a failure of selective attention, incongruent changes across modalities are expected to disrupt performance, relative to congruent changes. RT measures for rising sounds were consistent with this prediction (slower responses when they were paired with shrinking disks), but this effect was not seen for falling sounds. In terms of perceptual sensitivity, introduction of dynamic visual cues appeared to decrease sensitivity to rising sounds, regardless of how disks changed in size. Thus, there was no simple pattern of cross-modal interactions related to correspondences between changes in disks and changes in sounds.

The dissociable effects of congruency on d' and c in the present study highlight the complexity of cross-modal interactions. We found evidence of a congruency facilitation in d' only for the falling sounds, and in c only for the rising sounds. The effect of cross-modal congruency on partic-

ipants' readiness to respond, as measured by RT facilitation, revealed a more mixed pattern than that observed in d' and c. This complex pattern suggests that audiovisual interactions reflect more than one factor involved in the cross-modal processing of dynamic cues. As was noted above, previous cross-modal studies have established a variety of benefits of congruency across modalities, such as facilitation of response speed, enhancement of perceptual sensitivity, reduction of perceptual ambiguity, and stronger orienting to cross-modal target stimuli. The present results indicate that these effects are probably not driven by a single "binding" process. Instead, different measures may tap into different processes involved in a cross-modal task and, hence, might be appropriate for investigating different aspects of cross-modal processing.

A long line of work has shown that stimuli presented in one modality can enhance, distort, or interfere with processing in other modalities. Although there is growing consensus that both subcortical integration and cortically mediated, postperceptual selection processes probably contribute to audiovisual processing, little progress has been made toward predicting the circumstances under which two properties will (or will not) interact in a particular perceptual task. Ross and colleagues (Ross, Saint-Amour, Leavitt, Javitt & Foxe, 2007) proposed that audiovisual integration in speech processing is most likely to occur when neither modality alone provides adequate cues for recognition, whereas Kitagawa & Ichihara (2002) and others (Meyer et al., 2005; Zetzsche et al., 2002) suggested that, for audiovisual motion cues, maximal cross-modal interactions occur when the cues are maximally redundant/consistent. The complex interactions between perceptual and decisional (and likely, attentional) processes revealed in the present study suggest that audiovisual processing is not adequately captured by dichotomous models in which processing is either unimodally segregated (followed by selection of the most informative modality) or perceptually integrated (merged at early stages of processing) depending on the modal information content (see also Talsma et al., 2010). Our findings indicate that participants' representations of changes within a particular modality are, instead, modulated by a host of perceptual and nonperceptual factors and that even "irrelevant" cues from other modalities or contexts can potentially shape an individual's processing of dynamic audiovisual events.

References

Alais, D., & Burr, D. (2003). The "flash-lag" effect occurs in audition and cross-modally. *Current Biology, 13,* 59–63.

Alais, D., & Burr, D. (2004a). No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Research, 19,* 185–194. doi:10.1016/j.cogbrainres.2003.11.011

Alais, D., & Burr, D. (2004b). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262. doi:10.1016/j.cub.2004.01.029

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving, 23,* 3–38. doi:10.1163/187847510x488603

Bach, D. R., Neuhoff, J. G., Perrig, W., & Seifritz, E. (2009). Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology, 74,* 28–33. doi:10.1016/j.ijpsycho.2009.06.004

Bach, D. R., Schachinger, H., Neuhoff, J. G., Esposito, F., Di Salle, F., Lehmann, C., et al. (2008). Rising sound intensity: An intrinsic warning cue activating the amygdala. *Cerebral Cortex, 18,* 145–150. doi:10.1093/cercor/bhm040

Bertelson, P. (1998). Starting from the ventriloquist: The perception of multimodal events. In M. Sabourin, C. Fergus, & M. Robert (Eds.), *Advances in psychological science: Vol. 2. Biological and cognitive aspects* (pp. 419–439). Hove: Psychology Press.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review, 5,* 482–489.

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 151–177). Oxford: Oxford University Press.

Bertelson, P., & Radeau, M. (1976). Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola. *Perception & Psychophysics, 19,* 531–535.

Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., & Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia, 45,* 523–530. doi:10.1016/j.neuropsychologia.2005.12.012

Calvert, G., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. Cambridge: MIT Press.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory–visual looming cues by humans. *Neuropsychologia, 47,* 1045–1052. doi:10.1016/j.neuropsychologia.2008.11.003

Choe, C. S., Welch, R. B., Gilford, R. M., & Juola, J. F. (1975). Ventriloquist effect: Visual dominance or response bias? *Perception & Psychophysics, 18,* 55–60.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8,* 162–169. doi:10.1016/j.tics.2004.02.002

Forster, K. I., & Forster, J. C. (2003). Dmdx: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68,* 1191–1203.

Ghazanfar, A. A., Neuhoff, J. G., & Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. *Proceedings of the National Academy of Sciences, 99,* 15755–15757. doi:10.1073/pnas.242469699

Hall, D. A., & Moore, D. R. (2003). Auditory neuroscience: The salience of looming sounds. *Current Biology, 13,* R91–R93.

Harrison, N. R., Wuerger, S., & Meyer, G. F. (2010). Reaction time facilitation for horizontally moving auditory–visual stimuli. *Journal of Vision, 10,* 1–21.

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179,* 595–606. doi:10.1007/s00221-006-0814-y

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416,* 172–174.

Kitajima, N., & Yamashita, Y. (1999). Dynamic capture of sound motion by light stimuli moving in three-dimensional space. *Perceptual and Motor Skills, 89,* 1139–1158.

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, 158,* 405–414. doi:10.1007/s00221-004-1913-2

Liu, E. H., Church, B. A., & Mercado, E., III. (2008, November). *Congruent audiovisual dynamics facilitate perception of looming stimuli.* Poster presented at the the 7th Annual Auditory Perception, Cognition, and Action Meeting, Chicago, IL.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge: Cambridge University Press.

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology, 18,* 963–968. doi:10.1016/j.cub.2008.05.043

Maier, J. X., & Ghazanfar, A. A. (2007). Looming biases in monkey auditory cortex. *Journal of Neuroscience, 27,* 4093–4100. doi:10.1523/jneurosci.0330-07.2007

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus monkeys. *Neuron, 43*(2), 177–181.

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Currrent Directions in Psychological Science, 10,* 61–65.

Melara, R. D., Marks, L. E., & Potts, B. C. (1993). Early-holistic processing or dimensional similarity. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1114–1120.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221,* 389–391.

Meyer, G. F., & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *NeuroReport, 12,* 2557–2560.

Meyer, G. F., Wuerger, S., & Greenlee, M. (2011). Interactions between auditory and visual semantic stimulus classes: Evidence for common processing networks for speech and body actions. *Journal of Cognitive Neuroscience, 23,* 2291–2308. doi:10.1162/jocn.2010.21593

Meyer, G. F., Wuerger, S. M., Rohrbein, F., & Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research, 166,* 538–547. doi:10.1007/s00221-005-2394-7

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature, 395,* 123–124.

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13,* 87–110.

Rapp, B., & Hendel, S. K. (2003). Principles of cross-modal competition: Evidence from deficits of attention. *Psychonomic Bulletin & Review, 10,* 210–219.

Regan, D., & Beverley, K. (1978). Looming detectors in the human visual pathway. *Vision Research, 18,* 415–421.

Romei, V., Murray, M. M., Cappe, C., & Thut, G. (2009). Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Current Biology, 19,* 1799–1805. doi:10.1016/j.cub.2009.09.027

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception, 16,* 175–186.

Rosenblum, L. D., Wuestefeld, A. P., & Saldana, H. M. (1993). Auditory looming perception: Influences on anticipatory judgments. *Perception, 22,* 1467–1482.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17,* 1147–1153. doi:10.1093/cercor/bhl024

Sanabria, D., Soto-Faraco, S., & Spence, C. (2007a). Spatial attention and audiovisual interactions in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 927–937. doi:10.1037/0096-1523.33.4.927

Sanabria, D., Spence, C., & Soto-Faraco, S. (2007b). Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: A signal detection study. *Cognition, 102,* 299–310. doi:10.1016/j.cognition.2006.01.003

Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: Judgments of audiovisual temporal relations are enhanced for upright walkers. *Psychological Science, 19,* 469–475. doi:10.1111/j.1467-9280.2008.02111.x

Schirillo, J. A. (2011). Cross-modal detection using various temporal and spatial configurations. *Attention, Perception, & Psychophysics, 73,* 237–246. doi:10.3758/s13414-010-0012-7

Schouten, B., Troje, N. F., Vroomen, J., & Verfaillie, K. (2011). The effect of looming and receding sounds on the perceived in-depth orientation of depth-ambiguous biological motion figures. *PLoS One, 6,* e14725. doi:10.1371/journal.pone.0014725

Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schachinger, H., et al. (2002). Neural processing of auditory looming in the human brain. *Current Biology, 12,* 2147–2151.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408,* 788.

Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: The 'Stroop effect' vs. The 'Simon effect'. *Acta Psychologica, 73,* 159–170.

Soto-Faraco, S., & Kingstone, A. (2004). Multisensory integration of dynamic information. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 49–68). Cambridge: MIT Press.

Soto-Faraco, S., Kingstone, A., & Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia, 41,* 1847–1862. doi:10.1016/s0028-3932(03)00185-4

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004a). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 330–345. doi:10.1037/0096-1523.30.2.330

Soto-Faraco, S., Spence, C., & Kingstone, A. (2005). Assessing automaticity in the audiovisual integration of motion. *Acta Psychologica, 118,* 71–92. doi:10.1016/j.actpsy.2004.10.008

Soto-Faraco, S., Spence, C., Lloyd, D., & Kingstone, A. (2004b). Moving multisensory research along: Motion perception across sensory modalities. *Current Directions in Psychological Science, 13,* 29–32.

Spence, C. (2011). Crossmodal correspondences: A tutorial. *Attention, Perception, & Psychophysics, 73,* 971–995. doi:10.3758/s13414-010-0073-7

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge: MIT Press.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14,* 400–410. doi:10.1016/j.tics.2010.06.008

Watanabe, K., & Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science, 12,* 109–116.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88,* 638–667.

Wuerger, S. M., Hofbauer, M., & Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics, 65,* 1188–1196.

Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustical Society of America, 111,* 1832–1846. doi:10.1121/1.1458027

Zetzsche, C., Röhrbein, F., Hofbauer, M., & Schill, K. (2002). Audio-visual sensory interactions and the statistical covariance of the natural environment. In A. Calvo-Manzano, A. Perez-Lopez, & J. S. Santiago (Eds.), *Proceedings of the Forum Acusticum. Sevilla.*