Effects of temporal asynchrony and stimulus magnitude on competitive audio–visual binding

Jonathan M. P. Wilbiks & Benjamin J. Dyson

Attention, Perception, & Psychophysics

ISSN 1943-3921
Volume 75
Number 8

Atten Percept Psychophys (2013)
75:1883-1891
DOI 10.3758/s13414-013-0527-9

Atten Percept Psychophys (2013) 75:1883–1891
DOI 10.3758/s13414-013-0527-9

Effects of temporal asynchrony and stimulus magnitude on competitive audio–visual binding

Jonathan M. P. Wilbiks · Benjamin J. Dyson

Published online: 17 August 2013

Abstract When making decisions as to whether or not to bind auditory and visual information, temporal and stimulus factors both contribute to the presumption of multimodal unity. In order to study the interaction between these factors, we conducted an experiment in which auditory and visual stimuli were placed in competitive binding scenarios, whereby an auditory stimulus was assigned to either a primary or a secondary anchor in a visual context (VAV) or a visual stimulus was assigned to either a primary or secondary anchor in an auditory context (AVA). Temporal factors were manipulated by varying the onset of the to-be-bound stimulus in relation to the two anchors. Stimulus factors were manipulated by varying the magnitudes of the visual (size) and auditory (intensity) signals. The results supported the dominance of temporal factors in auditory contexts, in that effects of time were stronger in AVA than in VAV contexts, and stimulus factors in visual contexts, in that effects of magnitude were stronger in VAV than in AVA contexts. These findings indicate the precedence for temporal factors, with particular reliance on stimulus factors when the to-be-assigned stimulus was temporally ambiguous. Stimulus factors seem to be driven by high-magnitude presentation rather than cross-modal congruency. The interactions between temporal and stimulus factors, modality weighting, discriminability, and object representation highlight some of the factors that contribute to audio–visual binding.

Keywords Multisensory processing · Auditory perception · Visual perception

We are constantly exposed to information from multiple sources and in many different modalities. Making good decisions as to the assignment of modality-specific information to common sources is an important contribution to functioning properly in our world. Welch and Warren (1980) conceptualized the decision-making process associated with assigning a modality to a source as the *unity assumption*, wherein a perceiver must make a decision as to whether the two (or more) sensory inputs experienced are different-modality expressions of the same source or of two (or more) separate sources. Recent reviews (e.g., Koelewijn, Bronkhorst, & Theeuwes, 2010) have focused on the factors that contribute to the perception of unity between multiple sensory inputs, including task demands (e.g., Stein & Stanford, 2008), spatial and temporal coincidence (e.g., Calvert, Spence, & Stein, 2004), and congruency between the composite signals (e.g., Molholm, Ritter, Javitt, & Foxe, 2004). In the following experiment, we present novel competitive-binding paradigms by which congruency relations are pitted against temporal variation, in order to assess how these factors interact in making decisions regarding audio–visual unity.

The first influence on the binding of sensory inputs that is under consideration here is temporal coincidence. Despite a reasonable assumption that audio–visual binding should be optimal when sensory components are presented simultaneously, research has revealed that binding can also occur on either side of simultaneous presentations (Vroomen, Keetels, de Gelder, & Bertelson, 2004). Consequently, a time range within which audio–visual integration is most likely (a *temporal window of integration*) is often described. For example, van Wassenhove, Grant, and Poeppel (2007) found that participants reported audio and visual speech elements as being fused (defined as an illusory McGurk combination) if they were presented within an approximately 200-ms window, from the auditory information being presented 30 ms before the visual information (hereafter, *audio lead*) to the auditory information being presented 170 ms after the visual information (hereafter, *audio lag*). Such a temporal asymmetry is apparent in much of the previous data (see Vatakis & Spence, 2010, for a discussion of various stimulus types other than speech), in that auditory and visual forms of information are more likely to bind under conditions of auditory lag than of auditory lead. A general guideline for unity appears to be that the auditory stimulus

J. M. P. Wilbiks (✉) · B. J. Dyson
Department of Psychology, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada
e-mail: jwilbiks@psych.ryerson.ca

should be presented somewhere from 100 ms before to 200 ms after the visual stimulus (Dixon & Spitz, 1980; Lewald & Guski, 2003; Soto-Faraco & Alsisus, 2009; Spence & Squire, 2003), although there is evidence that neurons in the superior colliculus can also tolerate audio–visual integration at an asynchrony of 600 ms (Meredith, Nemitz, & Stein, 1987; Stein & Stanford, 2008, cited in Koelewijn et al., 2010). One reason for this asymmetry may be due to differences in the transmission speeds of auditory and visual information, both in the air and in the cortex. As was reported by Fujisaki, Shimojo, Kashino, and Nishida (2004), when visual information and auditory information arrive simultaneously at their respective sensory transducers, cortical responses to audition are approximately 30 to 70 ms faster than responses to vision. Previous research on the temporal window of integration guided the range of temporal variation deployed in the present design and helped anticipate the preference for auditory stimuli to be bound to preceding rather than subsequent visual stimuli.

The second influence under consideration is stimulus variation, particularly with reference to magnitude. The manipulation of size in the visual domain and the manipulation of intensity in the auditory domain help test the idea that common effects of magnitude may arise as a result of shared coding between modalities (the ATOM theory; Walsh, 2003), in addition to offering inroads into the study of (in)congruency across magnitude values in the context of audio–visual integration. In addition to the idea that larger-magnitude stimuli may promote binding assignment, certain congruent combinations of audio–visual information may also contribute to the binding process. The notion of congruency between auditory and visual information can be expressed through a number of stimulus dimensions and can be observed at a number of levels (see Spence, 2011, and, Walker, 2012, for reviews). For example, the relationships between auditory pitch and visual brightness, form, position, size, spatial frequency, and contrast (Evans & Treisman, 2010), and between auditory intensity and visual brightness (Marks, 1987), have all been studied in terms of the privileged nature of certain value combinations. For example, multimodal objects generating higher pitches are also more readily associated with lighter rather than darker illumination (Spence & Deroy, 2012). In terms of the implications of stimulus congruency for multimodal unity, Parise and Spence (2009) confirmed that for audio–visual combinations of visual (circle) size and auditory (tone) pitch, congruent pairings (i.e., large circles being associated with low pitch and small circles with high pitch) were more likely to be judged as being synchronous at larger time lags, relative to incongruent pairings (i.e., large circles being associated with high pitch and small circles with low pitch). Thus, the congruency between auditory and visual values appears to relax the temporal window of integration and increase the likelihood of unity at times when neutral or incongruent pairings would not be considered elements of the same source. In using magnitude variation in the present study, we hoped to express congruency between auditory and visual information in terms of their structural relationships (Spence, 2011): Namely, the magnitude of a visual signal was represented by size, and the magnitude of an auditory signal by intensity. Our hope was that this would facilitate correspondences between modalities (i.e., quiet/loud intensities and small/large sizes representing congruent audio–visual pairings, and quiet/loud intensities and large/small sizes representing incongruent audio–visual pairings), as well as offering the opportunity to examine these correspondences through both the repetition and change of magnitude within modalities.

In contrast to the potential importance of certain congruent magnitude relations between modalities, certain incongruent combinations of stimulus magnitude within the same modality are also thought to be differently processed (e.g., Maier, Neuhoff, Logothetis, & Ghazanfar, 2004). Specifically, the increase of stimulus magnitude across time appears to be preferentially processed such as the visual shape of a predator getting larger in size, or, an ambulance siren getting louder in intensity. For example, Cappe, Thut, Romei, and Murray (2009) examined auditory and visual magnitude change by presenting an image of a disk that got either larger or smaller, and sounds that got either louder or quieter. They found that multisensory larger and louder pairs showed the fastest reaction times when compared to smaller and quieter pairs, or, ambiguous pairings (larger + quieter, smaller + louder). Such data are traditionally discussed in the context of "looming," although in the present investigation other explanations are available and will be considered in the Discussion section.

In considering the relationship between temporal and stimulus factors, the closest study to our proposed work has been, to our knowledge, Gallace and Spence (2006). Here, two disks of different sizes were sometimes accompanied by a sound simultaneous with the second disk. They found that responses were slowest when no sound was presented, and fastest when the sound was congruent with the second disk (e.g., small–quiet). The overall finding that the "synesthetically" congruent sound facilitated responding shows that participants were unable to avoid processing auditory information, even though they were asked to ignore it and attend only to the visual stimuli. Although they produced evidence to suggest that auditory information contributed to visual decision making, it is not currently clear whether the reverse obtains—namely, whether visual information will contribute to auditory decision making. Moreover, the Gallace and Spence (2006) study contained some aspects of visual magnitude change (i.e., big and small sizes of the visual disks) but only studied the congruency between the auditory tone and the second visual disk. The present research elaborates on the procedures of Gallace and Spence (2006) to consider the interaction of both within- and cross-modal congruency and temporal factors on audio–visual binding. To consider all of these factors within the same design, a novel competitive-binding paradigm was established in which an auditory stimulus was assigned to either a primary or a secondary anchor in a visual context (VAV), or a visual stimulus was assigned to either a primary or a secondary anchor in an auditory context (AVA).

Given what is known about temporal presentation (Calvert et al., 2004; Scholl & Nakayama, 2000), we predicted that there would be a general tendency in VAV to attribute the to-be-bound auditory stimulus to the primary visual anchor, with a corresponding tendency in AVA to attribute the to-be-bound visual stimulus to the secondary auditory anchor. This pattern of responding follows the expected ecological combination of pairing an auditory stimulus with a visual stimulus that precedes rather than follows it (i.e., tolerance for auditory lag; van Wassenhove et al., 2007). Moreover, we expected temporal effects to be stronger in AVA contexts, given the dominance of auditory information (two auditory stimuli to one visual stimulus), but magnitude effects to be stronger in VAV contexts, given the dominance of visual information (two visual stimuli to one auditory stimulus; after Alais & Burr, 2004; Burr, Banks, & Morrone, 2009).

With respect to magnitude effects, specific predictions can also be made both between and within modalities. In terms of between-modality relations, we expected small visual stimuli to more likely be bound with quiet auditory stimuli, and large visual stimuli to more likely be bound with loud auditory stimuli (Spence, 2011; Walsh, 2003). It was hypothesized that these congruent combinations would yield greater evidence of binding outside of the "normal" range of temporal integration, relative to incongruent combinations (after Parise & Spence, 2009). In this way, specific *congruent between-modality* combinations were expected to play a role in perceptual assignment. In terms of within-modality relations, specific *incongruent within-modality* combinations were expected to play a role in perceptual assignment, with the use of small-to-large but not large-to-small magnitude changes across the two visual anchors (cf. Cappe et al., 2009; Neuhoff, 2001).

The final critical aspect of the design was to consider cases in which stimulus magnitude was put in competition with temporal factors. In doing so, we anticipated some of the future research recently proposed by Van der Burg, Awh, and Olivers (2013) in their study of the capacity of audio–visual integration: "If only one of multiple visual candidates is going to be associated with a sound, which one is integrated with it? One possibility is that the auditory signal is integrated with that happens to be the most dominant or salient synchronized visual event at that moment" (p. 351). Take, for example, in a VAV context, the case of a large primary visual anchor–loud auditory to-be-bound–small secondary visual anchor, where the to-be-bound stimulus was (a) congruent with the primary anchor but (b) presented simultaneously with the secondary anchor. Temporal coincidence dictates that it should be paired with the secondary anchor, but between-modality congruency relationships suggest a pairing with the primary anchor. This experiment will shed light on how stimulus (magnitude) and temporal factors work in collaboration or competition with one another in the binding of auditory and visual stimuli.

Method

Participants

Informed consent was obtained from 29 participants prior to the experiment. The only exclusion criterion was the failure to observe a strong positive correlation between second anchor attributions and time, since the likelihood of attributing the auditory signal to the second visual signal should increase as the delivery of the auditory signal moved toward the second visual signal. Two participants were excluded for poor correlations with time (*r*s = –.73, .28), with the correlation of the final sample being *r* = .97. The 27 participants making up the final sample had a mean age of 20.2 years (*SD* = 2.0) and included 24 females and 27 right-handed individuals. All participants self-reported normal or corrected-to-normal vision and hearing. The experimental procedure was approved by the Research Ethics Board at Ryerson University.

Stimuli and apparatus

A series of 100-ms 1-kHz sounds, with 5-ms linear onset and offset ramps, were created using SoundEdit 16 (MacroMedia). All sounds were played binaurally from free-field speakers (Harman-Kardon) positioned on either side of a computer monitor viewed approximately 57 cm away, to encourage magnitude coincidence between the auditory and visual signals (Calvert et al., 2004). All sounds were calibrated using a Scosche SPL100 sound level meter to approximately 56 or 71 dB(C), to represent quiet and loud sounds, respectively. The visual stimulus consisted of a yellow asterisk, presented in the center of a black screen in either 24- or 96-point Chicago font, to represent small and large sizes, respectively.[1] Stimulus presentation was controlled by PsyScope (Cohen, MacWhinney, Flatt,

[1] These values were based on the descriptive data of a preceding experiment in which participants (*n* = 36) were presented with various combinations of loud and quiet sounds (ranging across 71, 66, 61, and 56 dB(C)) and large and small sizes (ranging across 96-, 48-, 24-, and 12-point Chicago font). At each trial, an audio–visual stimulus was presented, and participants were prompted to respond either to the auditory or the visual value of the composite stimulus. Congruency effects were computed by subtracting reaction times (and error rates) for congruent pairings from those for incongruent pairings, and modality differences were calculated by taking the differences between auditory and visual responding. The values selected for the present experiment were those combinations of auditory intensity and visual size that produced the highest congruency effect (indicating a high level of between-modality congruency), combined with the smallest difference between auditory and visual responding (indicating a close fit between auditory and visual processing).

& Provost, 1993), and responses were recorded using a PsyScope Button Box.

Design and procedure

Experimental blocks of 240 trials were developed for each condition (VAV and AVA), involving the orthogonal combination of first-stimulus magnitude (V1: small, large; A1: quiet, loud), second-stimulus magnitude (A: quiet, loud; V: small, large), and third-stimulus magnitude (V2: small, large; A2: quiet, loud). These eight (2 × 2 × 2) sets of stimuli were further varied by changing the temporal presentation of the second stimulus (S2) with respect to the first (S1) and third (S3) stimuli (see Fig. 1). S1 and S3 were always presented with a 300-ms SOA and for 100 ms each. S2 could occur simultaneously with the onset of S1 (0 ms) and at 100-ms intervals until 100 ms after the onset of S3 (400 ms), for a total of five possible temporal presentations. These time points were chosen for the onset of the to-be-bound stimulus to coincide with the respective onsets (0 and 300 ms) and offsets (100 and 400 ms) of each anchor, plus a fifth time point (200 ms), at which only the to-be-bound offset and the second anchor onset were associated. These conditions were subject to a variable initial delay of 100, 200, 300, 400, 500, or 600 ms between trials, which was not considered in any further analysis.

Each trial began with the presentation of a blank screen for 500 ms, followed by a variable lag. Participants were then presented with the first anchor (S1) for 100 ms, and following a 200-ms interval, the second anchor (S3) was presented for 100 ms. The to-be-bound stimulus (S2) was presented at some time between V1 and V2 presentation, at 100-ms intervals. The stimulus presentation was followed, 500 ms after the offset of S3, by a response prompt saying "FIRST OR SECOND?" Participants were asked to respond by pressing the leftmost button on a PsyScope button box if they thought that the to-be-bound stimulus was caused by the first anchor, and the rightmost button if they thought that the to-be-bound stimulus was caused by the second anchor. As a result of the subjective nature of the task, no feedback was provided.

Participants completed two blocks of 240 trials for the VAV condition and two blocks of 240 trials for the AVA condition, each of which was preceded by a single practice block consisting of 12 trials taken randomly from the experimental blocks. In both the practice and experimental blocks, trial order was randomized. The order in which participants completed the blocks (VAV first or AVA first) was counterbalanced between participants.

Results

The proportions for second-anchor (V2 or A2) responding were subjected to a five-way repeated measures analysis of variance (ANOVA) with the factors Condition (VAV, AVA) × S1 (low magnitude [small or quiet], high magnitude [large or loud]) × S2 (low, high) × S3 (low, high) × Time (0, 100, 200, 300, 400 ms); the full results are shown in Table 1. In order to adjust for Type I errors, Greenhouse–Geisser corrections were applied to all analyses, and subsequent pairwise comparisons were evaluated with Bonferroni tests ($p < .05$). We found a main effect of condition, $F(1, 26) = 11.68$, $p = .002$ (also reported in Wilbiks & Dyson, in preparation), with the AVA paradigm yielding more second responding than the VAV paradigm. A significant Condition × Time interaction, $F(2.09, 54.32) = 15.07$, $p < .001$ (see Fig. 2a), showed that, whereas both conditions showed a clear preference for first responding at early time points, a preference for second responding at late time points, and relative ambiguity at the middle time point, the difference between VAV and AVA was significant at 200, 300, and 400 ms. Thus, the AVA context yielded a much clearer delineation between first- and second-anchor attribution as a function of temporal variation, presumably stemming from the

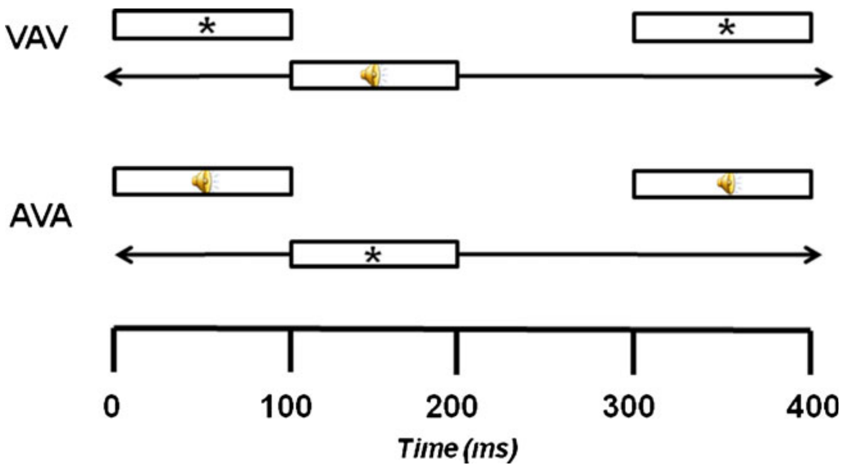

Fig. 1 Schematic representation of our procedure

Table 1 Summary of the five-way analysis of variance with Greenhouse–Geisser corrections applied

Metric	*df*	*F*	*MSE*	*p*
Condition (C)	1, 26	**11.68**	**.151**	**.002**
1st Stimulus (S1)	1, 26	**46.21**	**.044**	**<.001**
2nd Stimulus (S2)	1, 26	0.39	.033	.539
3rd Stimulus (S3)	1, 26	**11.09**	**.028**	**.003**
Time (T)	1.48, 38.48	**194.84**	**.185**	**<.001**
C × S1	1, 26	2.56	.052	.121
C × S2	1, 26	0.09	.023	.770
C × S3	1, 26	2.67	.024	.114
C × T	2.09, 54.32	**15.07**	**.048**	**<.001**
S1 × S2	1, 26	0.51	.011	.482
S1 × S3	1, 26	0.96	.049	.336
S1 × T	3.38, 88.00	**19.31**	**.019**	**<.001**
S2 × S3	1, 26	0.11	.021	.744
S2 × T	2.87, 74.57	**3.67**	**.015**	**.017**
S3 × T	3.08, 79.95	**3.16**	**.014**	**.028**
C × S1 × S2	1, 26	0.23	.018	.639
C × S1 × S3	1, 26	0.17	.054	.685
C × S1 × T	3.33, 86.52	1.39	.016	.250
C × S2 × S3	1, 26	0.03	.013	.862
C × S2 × T	3.51, 91.29	1.35	.011	.262
C × S3 × T	3.79, 98.44	1.48	.016	.218
S1 × S2 × S3	1, 26	0.16	.017	.695
S1 × S2 × T	3.47, 90.31	2.26	.013	.077
S1 × S3 × T	3.29, 85.56	**3.02**	**.012**	**.030**
S2 × S3 × T	3.35, 87.11	0.29	.014	.852
C × S1 × S2 × S3	1, 26	0.28	.011	.605
C × S1 × S2 × T	3.10, 80.51	0.57	.017	.643
C × S1 × S3 × T	3.12, 81.24	**4.40**	**.018**	**.005**
C × S2 × S3 × T	3.78, 98.25	0.91	.014	.455
S1 × S2 × S3 × T	3.61, 93.96	0.42	.013	.775
C × S1 × S2 × S3 × T	3.19, 82.93	0.51	.013	.687

Statistical significance is indicated by bold.

higher resolution of temporal information as a result of using two auditory, rather than visual, anchors. To support the idea of increased temporal certainty in AVA relative to VAV, the absolute values of the differences between the average proportions of secondary responding and .5 were taken for each time bin. Thus, a certainty score of 0 would indicate *total uncertainty* (i.e., secondary responding at .5 – .5 = 0), whereas a certainty score of .5 would indicate *total certainty* (i.e., proportion of secondary responding at either 0.0 or 1.0 – .5 = |.5|).[2] Certainty measures were calculated and submitted to a Condition (VAV, AVA) × Time (0, 100, 200, 300, 400 ms) ANOVA (see Fig. 2b). We found a main effect of condition, $F(1, 26) = 7.93$, $p = .009$, which indicated that the AVA paradigm showed higher levels of binding certainty than did the VAV paradigm. An effect of time, $F(4, 104) = 41.15$, $p < .001$, was significant, and Bonferroni comparisons indicated that the 0- and 400-ms time points showed significantly higher levels of certainty than the other time points (100, 200, and 300 ms). This is in accordance with expectations, as a to-be-bound stimulus temporally coinciding with an anchor stimulus should be more certainly bound to that anchor. A Condition × Time interaction, $F(4, 104) = 10.76$, $p < .001$, showed that the AVA was significantly more certain than the VAV paradigm at the 300- and 400-ms time points. Thus, response certainty was higher for AVA than for VAV, higher at time points coincident with anchor presentation, and particularly high for AVA responding during the presentation of the second rather than the first auditory anchor.

To return to the main ANOVA, an S1 × Time interaction, $F(3.38, 88.00) = 19.31$, $p < .001$, revealed that the presentation of low-magnitude first stimuli (small, quiet) increased the likelihood of second responding relative to presentation of high-magnitude first stimuli (large, loud). A complementary effect

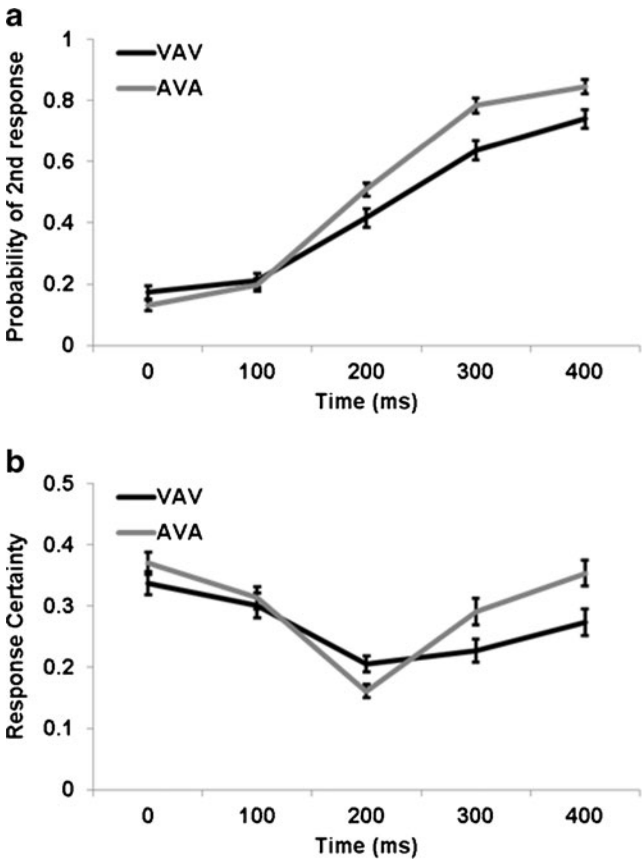

Fig. 2 (**a**) Graph showing the probabilities of second-anchor responding as a function of time and paradigm (VAV, AVA), with AVA conditions leading to increased probability of binding to the second anchor, relative to VAV conditions. Error bars depict standard errors. (**b**) Graph showing the probabilities of response certainty as a function of time and paradigm (VAV, AVA). VAV contexts produced less response certainty than AVA contexts during second-anchor presentation. Error bars depict standard errors

[2] We thank an anonymous reviewer for this suggestion

was found in the S3 × Time interaction, $F(3.08, 79.95) = 3.16$, $p = .028$, wherein high-magnitude third stimuli (large, loud) increased the likelihood of second responding relative to low-magnitude third stimuli (small, quiet). An S1 × S3 × Time interaction, $F(3.29, 85.56) = 3.02$, $p = .030$, confirmed that such stimulus effects were more likely during the presentation of the to-be-bound stimulus at 200 ms, a temporal point at which only to-be-bound offset and second-anchor onset were temporally contiguous (see Fig. 1). A four-way Condition × S1 × S3 × Time interaction, $F(3.12, 81.24) = 4.40$, $p = .005$, allowed for resolution of these lower-level interactions (see Fig. 3). Interestingly, beyond the effects already discussed, we saw that a low-magnitude S1 (small) in combination with a high-magnitude S3 (large) increased second responding relative to all other stimulus pairings only in the VAV condition, and only at 300 ms. The AVA paradigm failed to show this within-modality effect of stimulus incongruency.

An interaction between S2 and Time, $F(2.87, 74.57) = 3.67$, $p = .017$, failed to reveal any significant differences between magnitudes, and evidence for between-modality stimulus effects failed to reach traditional levels of statistical significance (all ps > .078).

Discussion

This experiment was conducted to investigate how temporal and stimulus factors work with one another when reaching a decision on the relationship between auditory and visual signals (Welch & Warren, 1980). Novel competitive binding paradigms were established: In the case of a visual-rich context (VAV), a to-be-bound auditory stimulus had to be ascribed to one of two visual anchors, whereas in the case of an auditory-rich context (AVA), a to-be-bound visual stimulus had to be ascribed to one of two auditory anchors. Alongside temporal variation of the to-be-bound stimulus, all auditory and visual signals also varied with respect to stimulus magnitude: Small visual and quiet auditory signals were deemed to be of low

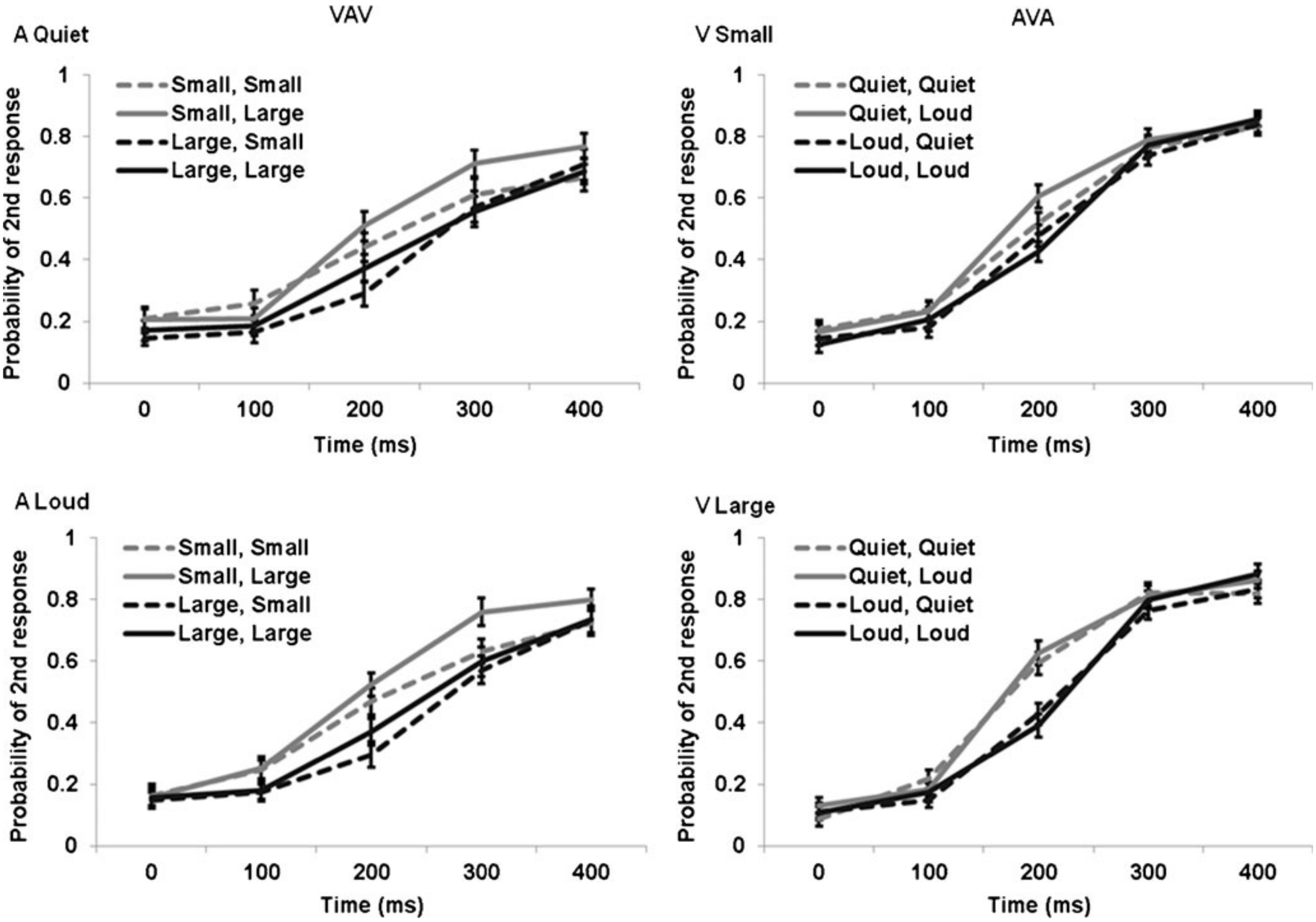

Fig. 3 Graph showing the the probabilities of second-anchor responding as a function of visual-rich (VAV) and auditory-rich (AVA) contexts, temporal asynchrony, and stimulus magnitude. Error bars depict standard errors

magnitude, whereas large visual and loud auditory signals were deemed to be of high magnitude (Walsh, 2003). This allowed for the evaluation of *between-modality congruency* effects (e.g., Gallace & Spence, 2006), wherein signals of shared magnitude values might increase the likelihood of binding, but also the evaluation of *within-modality incongruency* effects (e.g., Cappe et al., 2009), wherein specific changes in magnitude (small to large) might also influence the likelihood of anchor attribution. Although we believed that temporal and stimulus factors would contribute in both VAV and AVA contexts, it was predicted that different contexts would show differential temporal and stimulus effects due to the modality of the anchor. Given the association between the auditory modality and temporal analyses (e.g., Burr et al., 2009), temporal effects were expected to be stronger in the AVA than in the VAV case. Conversely, stimulus effects were expected to be stronger in the VAV than in the AVA case (e.g., Alais & Burr, 2004).

In terms of temporal effects (e.g., Fig. 2), the likelihood of binding was relatively consistent when the to-be-bound stimulus was presented coincidentally with an event associated with the primary anchor (e.g., at 0 and 100 ms). This contrasted with subsequent binding performance, in which the probability of secondary responding differentially increased for the VAV and AVA paradigms (300 and 400 ms) following the numerical midpoint of the temporal distribution (200 ms): Our comparison confirmed a tendency for the AVA paradigm to show more secondary binding than the VAV paradigm. As indexed by previous research (see Kohlrausch & van de Par, 2000; Roseboom, Nishida, & Arnold, 2009; Scholl & Nakayama, 2000; van Wassenhove et al. 2007), sensory systems are more likely to perceive auditory and visual stimuli as having a common source when the visual leads the auditory. Hence, for the auditory-rich context, the stronger temporal association was between the last two stimuli (A*VA*), whereas for the visual-rich context, the stronger temporal association was between the first two stimuli (*VA*V). Further support for the increased reliance on temporal factors in AVA relative to VAV was derived from a certainty measure. Here, binding attributions were closer to the absolute probabilities of 0 and 1 in AVA, suggesting that temporal resolution in the presence of two auditory anchors was finer-tuned than in the presence of two visual anchors.

In terms of stimulus effects, participants appeared sensitive to the magnitudes of both visual and auditory signals, but in different ways. One basic effect observed was that when the primary anchor was of low magnitude, expressed either by small size in vision or quiet intensity in audition, it was less likely to attract the binding of the to-be-bound stimulus (S1 × Time interaction). The similar behavior of both visual size and auditory intensity supports the ATOM theory (Walsh, 2003), which holds that size and intensity may be different modal metrics of the same global index (i.e., magnitude). Similarly, Smith and Sera (1992) would deem these prothetic (as opposed to metathetic) dimensions, which similarly allow for a mapping onto an amodal representation of magnitude. However, the VAV and AVA paradigms also showed different effects, in that the VAV context showed within-modality effects and no between-modality effects, whereas the AVA context showed neither within- nor between-modality effects. The within-modality stimulus magnitude effects observed in VAV were characterized as incongruent, with increased second-anchor responding occurring for small- followed by large-size visual asterisks. Although this is at least consistent with the previous data of looming stimuli, in that the change from a small visual shape to a larger version of the same shape might give the impression of the same object approaching in depth, concern might be raised over the labeling of such noncontinuous change between visual stimuli as "looming." Indeed, alternative explanations regarding attentional capture also suggest themselves, in that the binding of an auditory signal to a strong visual source is promoted only when the previous visual source is weak (e.g., small) rather than strong (e.g., large). In other words, auditory signals bind themselves to the first attentionally capturing visual source, defined by size. Such an account would appear consistent with the S1 × Time interaction reported above. With the observation of this effect for visual anchors only, we do not seek to deny the data regarding auditory (Neuhoff, 2001) and multimodal (Maier et al., 2004) looming, but simply to underscore that in the present context, size variation in visual stimuli was more pertinent to the observer than intensity variation in auditory stimuli (see also Cappe et al., 2009).[3] Further research is clearly warranted regarding the strength of the "looming" effect between continuous and noncontinuous stimuli, and also as a function of magnitude strength.

In contrast to the significant within-modality stimulus effects for VAV, the present data failed to show significant between-modality stimulus effects for VAV and AVA. Our failure to find a significant effect of audition on vision, as per Gallace and Spence (2006), may be attributable to a number of different design features, including (but not limited to) an increase in

[3] An additional experiment was performed in which the color of visual stimuli was varied, within the VAV paradigmatic framework. This was done in order to look at the effect of color change as an index of object constancy: When the two visual anchors were of the same color, they should be more likely to be perceived as two representations of the same item, whereas if they were of different colors, they would more likely be perceived as single representations of two items. It was believed that, as some of the findings were being interpreted as the perception of looming between the two visual stimuli, having a same-color pairing would support this looming, and having a different-color pairing would break the looming effects. The findings, however, indicate that this was not the case. Rather, the V1 × V2 × Color interaction, $F(1, 26) = 6.62$, $p = .016$, showed that color change enabled a release from primary-anchor binding when the second visual anchor was of high magnitude. That is to say, when V1 and V2 are both large, having a color change increases the probability of secondary binding, allowing release from the initial inclination to bind to V1. It is also possible that this change in color serves as an additional attention-capturing feature of the secondary anchor, and leads to increased secondary binding in that way.

visual variation in our experiment (only the second of the two visual stimuli varied in size in Gallace & Spence, 2006), the temporal predictability of the auditory stimulus (auditory stimuli were only concurrent with the second visual stimuli in Gallace & Spence, 2006), and/or the use of highly discriminable pitches (300 and 4500 Hz in Gallace & Spence, 2006) relative to potentially less discriminable variations in intensity (56 and 71 dB) in our study.

In terms of the interaction between temporal and stimulus (magnitude) factors in establishing audio–visual binding, in the present setup we found evidence to suggest that the consideration of stimulus relationships works very much in the service of temporal factors in making binding decisions. It is apparent from Fig. 3 that the temporal relationship between the anchors and the to-be-bound stimulus was the dominant contribution, with stimulus factors playing a role when the to-be bound stimulus coincided with a second anchor event (e.g., the to-be-bound stimulus offset with the second-anchor onset [200-ms condition]; the to-be-bound onset with the second-anchor onset [300-ms condition]; or the to-be-bound stimulus onset with the second-anchor offset [400-ms condition]). One pragmatic reason for the influence of stimulus factors during secondary- rather than primary-anchor presentation is because at the first (0-ms) time point, when the to-be-bound stimulus onset was simultaneous with the primary-anchor onset, the two paradigms (VAV and AVA) were identical with regard to stimulus presentation. This effect is also consistent with Roseboom et al.'s (2009) suggestion that when initial multimodal binding decisions are made, they are hard to break. As such, binding may be more flexible at the end of a temporal epoch than at the beginning.

One final consideration is that since modality interactions can occur at a variety of different levels of processing (see Spence, 2011, Table 2), the expression of stimulus relations in the present series may have been late (semantic; Walker, 2012) rather than early (structural or statistical), and this may be why stimulus effects were utilized only after temporal information failed to resolve the causal attribution. The use of auditory and visual dimensions that represent magnitude makes this seem less likely, since it could be argued that the correlation between visual size and auditory intensity is largely structural rather than semantic in nature (Spence, 2011; Walsh, 2003). It would be interesting to consider whether the relationship between temporal and stimulus factors is contingent on the use of magnitude and to consider alternative combinations of visual shapes and auditory waveforms (e.g., amorphous vs. sharp shapes, and sinusoidal vs. square waves; Hossain, 2011; Ramachandran & Hubbard, 2001). By maintaining the competitive paradigm, but using contour rather than magnitude manipulations, it would be possible to evaluate the idea of whether temporal factors could ever work in the service of certain types of stimulus variations. The proposed research may also help to allay concerns regarding magnitude change and attentional capture (see above). An additional check of the relationship between temporal and stimulus factors would be the systematic manipulation of the degrees of variation found within each. In the present case, it could be argued that eight examples of stimulus variation were possible (i.e., S1 [2] × S2 [2] × S3 [2]), but only five examples of temporal variation (i.e., 0, 100, 200, 300, or 400 ms). On the basis of variation alone, stimulus change was the more variable factor, and hence could have been prioritized over temporal change. The observation that temporal information was, overall, more influential during audio–visual binding suggests that differences in variation do not respond in the expected way. Nevertheless, as a preliminary step in the present series, we have revealed asymmetries in assigning a to-be-bound auditory stimulus to one of two visual anchors and in assigning a to-be-bound visual stimulus to one of two auditory anchors, with these asymmetries aligning themselves with classically held domains of modality dominance. By allowing both temporal and stimulus factors to collaborate and compete with one another, we have revealed some of the complex interactions in resolving the audio–visual binding problem.

Author Note B.J.D. is supported by an Early Researcher Award granted by the Ontario Ministry of Research and Innovation. Portions of the data were presented at the International Multisensory Research Forum, 19–22 June, 2012, at Oxford, UK.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198,* 49–57.

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processing*. Cambridge, MA: 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.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 25,* 257–271. doi:10.3758/BF03204507

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9,* 719–721.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1):6, 1–12. doi:10.1167/10.1.6.

Fujisaki, W., Shimojo, S., Kashino, S., & Nishida, S. (2004). Recalibration of audio–visual simultaneity. *Nature Neuroscience, 7,* 773–778.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68,* 1191–1203. doi:10.3758/BF03193720

Hossain, S. (2011). *Shapes and sounds: an exploration of audiovisual crossmodality.* Unpublished manuscript, University of Texas at Dallas, Dallas, TX.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audio–visual studies. *Acta Psychologica, 134,* 372–384.

Kohlrausch, A., & van de Par, S. (2000). Experimente zur Wahrnehmbarkeit von Asynchronie in audio–visuellen Stimuli [Experiments on the perception of asynchrony with audio–visual stimuli]. In *Fortschritte der Akustik* (DAGA 2000, pp. 316–317). Oldenburg, Germany: DEGA Geschaftstelle.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16,* 468–478.

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

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology. Human Perception and Performance, 13,* 384–394. doi:10.1037/0096-1523.13.3.384

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors. *Journal of Neuroscience, 7,* 3215–3299.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual–auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex, 14,* 452–465.

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

Parise, C. V., & Spence, C. (2009). "When birds of a feather flock together": Synesthetic correspondences modulate audio–visual integration in non-synesthetes. *PLoS One, 4,* e5664.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies, 8,* 3–34.

Roseboom, W., Nishida, S., & Arnold, D. H. (2009). The sliding window of audio–visual simultaneity. *Journal of Vision, 9*(12):4, 1–8. doi:10.1167/9.12.4.

Scholl, B. J., & Nakayama, K. (2000, November). *Contextual effects of the perception of causality.* Poster presented at the annual meeting of the Psychonomic Society, New Orleans, LA.

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology, 24,* 99–142. doi:10.1016/0010-0285(92)90004-L

Soto-Faraco, S., & Alsisus, A. (2009). Deconstructing the McGurk–MacDonald illusion. *Journal of Experimental Psychology. Human Perception and Performance, 35,* 580–587.

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

Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *i-Perception, 3,* 316–318. doi:10.1068/i0526ic

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology, 13,* R519–R521.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9,* 255–266.

Van der Burg, E., Awh, E., & Olivers, C. N. L. (2013). The capacity of audiovisual integration is limited to one item. *Psychological Science, 24,* 345–351. doi:10.1177/0956797612452865

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception. *Neurophysiologia, 45,* 598–607.

Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli. In M. J. Naumer & J. Kaiser (Eds.), *Multisensory object perception in the primate brain* (pp. 95–121). Berlin Heidelberg: Springer-Verlag. doi:10.1007/978-1-4419-5615-6_7

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audiovisual asynchrony. *Cognitive Brain Research, 22,* 32–35.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics, 74,* 1792–1809. doi:10.3758/s13414-012-0341-9

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space, and quantity. *Trends in Cognitive Sciences, 7,* 483–488. doi:10.1016/j.tics.2003.09.002

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88,* 638–667. doi:10.1037/0033-2909.88.3.638

Wilbiks, J. M. P., & Dyson, B. J. (in preparation). *I can't believe my eyes (but I can believe my ears): Adjusting auditory signals to visual sources but not visual signals to auditory sources*. Manuscript submitted for publication.