



# Auditory stimulus timing influences perceived duration of co-occurring visual stimuli

Vincenzo Romei<sup>1,2\*</sup>, Benjamin De Haas<sup>1,2</sup>, Robert M. Mok<sup>2,3</sup> and Jon Driver<sup>1,2</sup>

<sup>1</sup> Wellcome Trust Centre for Neuroimaging at UCL, University College London, London, UK

<sup>2</sup> UCL Institute of Cognitive Neuroscience, University College London, London, UK

<sup>3</sup> Department of Psychological Sciences, Birkbeck, University of London, London, UK

## Edited by:

Nadia Bolognini, University Milano  
Bicocca, Italy

## Reviewed by:

Simon Baumann, Newcastle  
University, UK

Raymond V. Ee, University Utrecht,  
Netherlands

## \*Correspondence:

Vincenzo Romei, Wellcome Trust  
Centre for Neuroimaging at UCL,  
University College London, 12 Queen  
Square, WC1N 3BG London, UK.  
e-mail: v.romei@ucl.ac.uk

There is increasing interest in multisensory influences upon sensory-specific judgments, such as when auditory stimuli affect visual perception. Here we studied whether the duration of an auditory event can objectively affect the perceived duration of a co-occurring visual event. On each trial, participants were presented with a pair of successive flashes and had to judge whether the first or second was longer. Two beeps were presented with the flashes. The order of short and long stimuli could be the same across audition and vision (audio–visual congruent) or reversed, so that the longer flash was accompanied by the shorter beep and vice versa (audio–visual incongruent); or the two beeps could have the same duration as each other. Beeps and flashes could onset synchronously or asynchronously. In a further control experiment, the beep durations were much longer (tripled) than the flashes. Results showed that visual duration discrimination sensitivity ( $d'$ ) was significantly higher for congruent (and significantly lower for incongruent) audio–visual synchronous combinations, relative to the visual-only presentation. This effect was abolished when auditory and visual stimuli were presented asynchronously, or when sound durations tripled those of flashes. We conclude that the temporal properties of co-occurring auditory stimuli influence the perceived duration of visual stimuli and that this can reflect genuine changes in visual sensitivity rather than mere response bias.

**Keywords:** multisensory integration, crossmodal interactions, response bias, signal-detection theory, audition, vision, time perception

## INTRODUCTION

There is increasing interest within multisensory research in how auditory information can influence visual perception (e.g., Shams et al., 2000; Vroomen and de Gelder, 2000; Calvert et al., 2004; Spence and Driver, 2004; Vroomen et al., 2004; Cappe et al., 2009; Vroomen and Keetels, 2009; Leo et al., 2011). Time is one fundamental dimension for all sensory modalities, and there are now several studies that demonstrate that manipulating the temporal dimension in one modality affect perception for other modalities (e.g., Eagleman, 2008; Freeman and Driver, 2008). Perception of an event's duration can deviate from its physical characteristics (Eagleman, 2008) and in multisensory cases might be most influenced by the sensory modality that carries the most reliable temporal information (see Welch and Warren, 1980; Walker and Scott, 1981; Recanzone, 2003; Wada et al., 2003; Alais and Burr, 2004; Ernst and Bühlhoff, 2004; Witten and Knudsen, 2005; Burr and Alais, 2006), as we studied here for audio–visual cases.

While the visual system typically has a higher spatial resolution than the auditory system (e.g., Witten and Knudsen, 2005) audition is usually more reliable for temporal aspects of perception (Repp and Penel, 2002; Bertelson and Aschersleben, 2003; Morein-Zamir et al., 2003; Guttman et al., 2005; Getzmann, 2007; Freeman and Driver, 2008). Accordingly vision can dominate audition in determining spatial percepts, as in the classic “ventriloquist effect” (Howard and Templeton, 1966; Thurlow and Jack, 1973; Bertelson

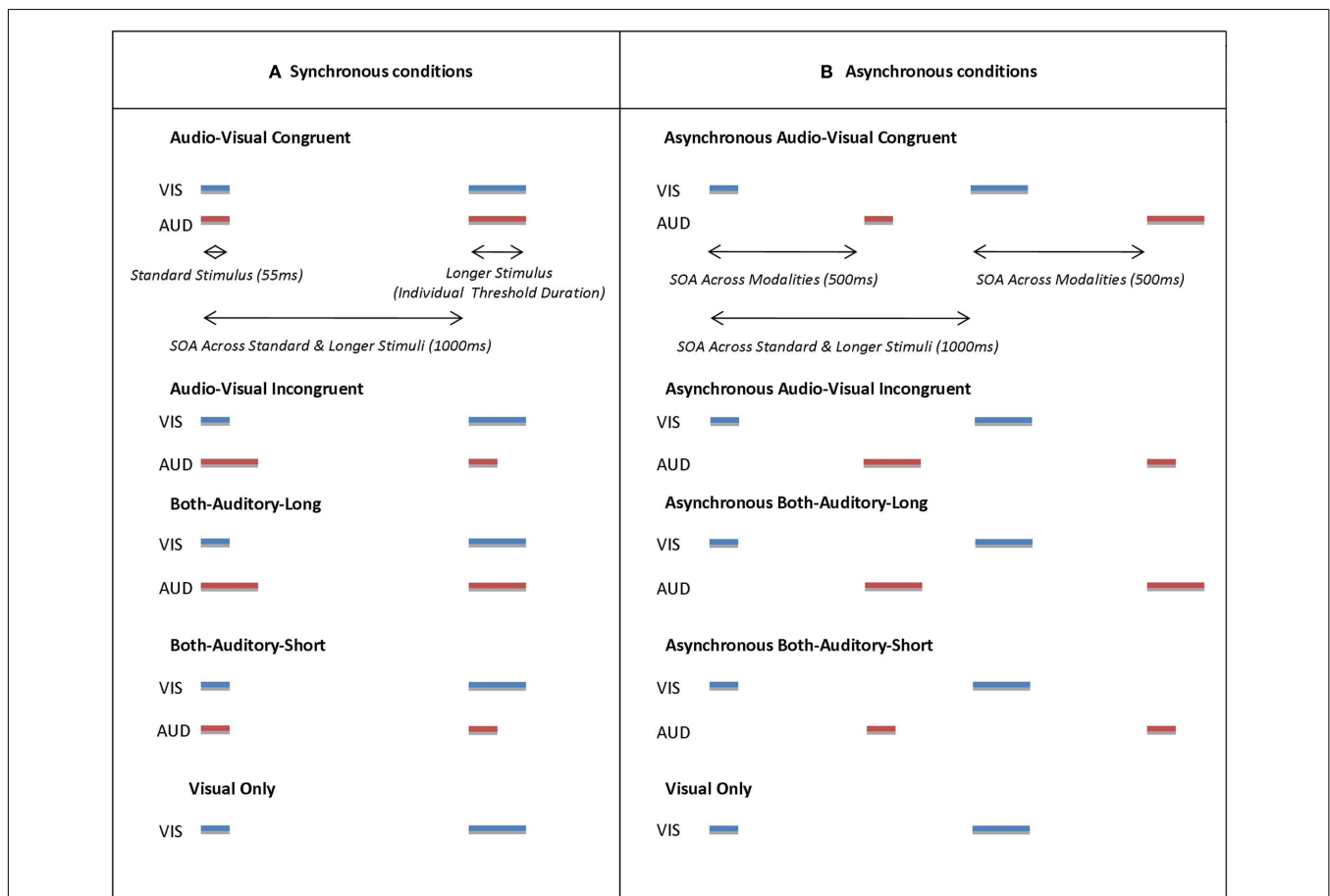
and Radeau, 1981). Conversely, audition may dominate vision in the temporal domain (Welch and Warren, 1980; Repp and Penel, 2002; Bertelson and Aschersleben, 2003; Morein-Zamir et al., 2003; Guttman et al., 2005; Getzmann, 2007; Freeman and Driver, 2008; Kanai et al., 2011) leading to so-called “temporal ventriloquism” (e.g., Gehard and Mowbray, 1959; Bertelson and Aschersleben, 2003). Freeman and Driver (2008) found that timing of a static sound can strongly influence spatio-temporal processing of concurrent visual apparent motion. Shams et al. (2000) found that illusory percepts of multiple flashes can be induced when a single flash is accompanied by a sequence of multiple beeps. Shipley (1964) showed that changes in the physical flutter rate of a clicking sound induce simultaneous changes in the apparent flicker rate of a flashing light.

Several crossmodal effects on subjective time perception in particular have been described (e.g., Walker and Scott, 1981; Donovan et al., 2004; Chen and Yeh, 2009; Klink et al., 2011). Chen and Yeh (2009) reported in an oddball paradigm that auditory stimuli can apparently extend reported visual duration, while visual stimuli had no such impact on reported auditory duration (see also Donovan et al., 2004; Klink et al., 2011; but see also van Wassenhove et al., 2008; Aen-Stockdale et al., 2011 for alternative accounts). But despite such suggestions of auditory influences on visual duration perception, to date it has typically been hard to establish whether such influences reflect response biases or instead genuine

changes in visual sensitivity, in signal-detection terms (Macmillan and Creelman, 1991).

We sought to address this issue directly here. On each trial subjects were presented with two visual stimuli in succession (see **Figure 1**) and had to make a force-choice about which was longer, which had an objectively correct answer. The multisensory manipulation was that we could also present two sounds on each trial. In Experiment 1a each sound was presented simultaneously with a flash (see **Figure 1**). These were possible two durations for the flashes, both used on every single trial so that one flash (either the first or the second) was longer. The beep durations on a given trial could potentially agree with those for the two successive flashes (congruent condition). Alternatively the two sounds could have the reverse order of durations (incongruent condition), or else the same duration as each other (both sounds short, or both sounds long). We measured whether manipulating the auditory durations had an impact on objective performance in the visual duration discrimination task, analyzing this in terms of signal-detection theory.

To determine whether any influence of auditory durations on visual duration judgments depended on synchrony between the multisensory events, in a control study (Experiment 1b) we misaligned the onsets of auditory and visual events (by 500 ms) to produce asynchronous control conditions. If the impact of auditory durations on visual duration perceptions reflects multisensory binding (e.g., see Meredith et al., 1987; Colonius and Diederich, 2011), it should be eliminated or reduced in the asynchronous condition; whereas if instead it were simply to reflect a response bias (similar to the response that a blind observer might give when asked to report visual durations when only hearing sounds) the auditory influence should remain the same even in the new asynchronous case of Experiment 1b. Finally, in a further control situation (Experiment 2) we tripled the length of auditory durations relative to visual durations, reasoning that if auditory and visual durations mismatch sufficiently, there should be less genuine perceptual binding between them (whereas once again, the response bias of a strictly blind observer who only hears sounds, misreporting them as if they were seen, should remain the same).



**FIGURE 1 | Schematic timelines representing conditions in Experiment 1A and 1B.** In Experiment 1A all corresponding auditory (red lines marked with label “AUD”) and visual (blue lines marked with label “VIS”) stimuli had the same onset (synchronous conditions) while in Experiment 1B visual stimuli always preceded auditory stimuli by 500 ms (asynchronous conditions). In both situations, each pair of stimuli within one modality was

separated by a 1000-ms interval; the order of short and long stimuli could be the same across auditory and visual modalities (congruent) or reversed between them (incongruent). In the both-auditor-short condition, both successive sounds had the shorter visual duration (and vice versa for both-auditor-long). Finally a visual-only condition served as a baseline measure.

## EXPERIMENT 1

Seventeen participants with a mean age of 26.29 (range 19–35) took part in the first experiment, nine female, one left-handed. All reported normal or corrected visual acuity and normal hearing. All gave written informed consent in accord with University College London ethics approval, were naïve to the purpose of the study and were paid for their time. One participant was excluded because she showed an inconsistent pattern in the visual titration (see below). Two more were excluded because their performance in the visual-only condition was at ceiling in the main experiments, leaving 14 participants in the sample.

### APPARATUS

Stimuli were presented on a 21" CRT display (Sony GDM-F520) in a darkened room. Participants sat with their head in a chin rest at 65 cm viewing distance. Video resolution was 1600 × 1200, with screen refresh rate of 85 Hz. Two small stereo PC speakers were placed just in front of the monitor immediately on either side of it. Stimulus control and data recording were implemented on a standard PC, running E-Prime 2 Professional (Psychology Software tools, Inc., www.pstnet.com). Unspeeded manual two-choice responses were made using a standard PC keyboard.

### STIMULI

Each visual stimulus comprised a white disk extending 1.2° in visual angle with its midpoint at 3° below a central fixation cross on a gray background. On each trial a pair of disks was flashing consecutively with varying durations from 55 to 165 ms.

The auditory stimulus was a 900-Hz pure tone sampled at 44.100 kHz with durations also varying from 55 to 165 ms. Sound level was measured with an audiometer and set to ~70 dB(A).

### PROCEDURE

#### Visual titration

Only visual stimuli were presented during this part of the experiment. On each trial participants were presented with a pair of disks flashing in two consecutive time windows separated by an SOA of 1000 ms. While one of the two visual stimuli had a constant standard duration of 55 ms, the other was slightly longer, with its duration varying between 66 and 165 ms (10 possible incremental steps of one frame at 85 Hz, i.e., ~11 ms). The latter stimulus type will be referred to as the “longer” stimulus. Each of the resulting 10 pairings of standard and longer stimuli was repeated 10 times per block. Each participant completed two to three blocks. The pairwise order of standard and longer stimuli was counterbalanced between trials, with standard-longer or longer-standard pairwise sequences being equiprobable.

On each trial participants were instructed to indicate whether the first or the second flash lasted longer, by pressing a corresponding button on the keyboard (“1” or “2”). This allowed us to identify the visual duration discrimination threshold for each participant individually. Threshold was defined as corresponding to the increase in duration for the longer stimulus whose duration allowed correct identification of it as longer in ~75% of cases. As it turned out, for the selected threshold stimulus participants were able to discriminate differences of durations correctly in 73.78 (±1.6 SE)% of cases for those stimulus pairings containing the

longer stimulus that was identified as threshold. The average duration of the longer stimuli identified as threshold was 103.4 (±3.97 SE) ms duration, for the visual disks used.

### MAIN EXPERIMENT 1

In each trial of the main experiment, participants were presented with the pair of visual stimuli previously identified as around threshold from the titration task. Again, the order of standard and longer visual stimulus was counterbalanced and equiprobable, with participants again asked to indicate which of the two consecutive flashes lasted longer. But the main experiment now consisted of 10 conditions (5 in Experiment 1a, and 5 in Experiment 1b, with these 10 all intermingled but presented separately here for ease of exposition). These 10 conditions differed with regard to whether, when, and how any sounds were presented with the flashes. Participants were emphatically instructed to ignore all sounds played during the experiment and to judge only the duration of the visual stimuli.

Two pure tone durations were selected for each participant – one lasting 55 ms and thus matching the standard visual stimulus in duration, the other auditory duration matching the participant-specific longer visual stimulus identified as threshold during the preceding visual titration task. These two pure tones were then combined with the flashes according to condition. There were two main classes of conditions: potentially synchronous (Experiment 1a) or asynchronous (Experiment 1b). In the potentially synchronous conditions, tone onset was temporally aligned with the visual onsets; whereas in the potentially asynchronous conditions, the onset of tones was delayed for 500 ms (thus 180° out of “phase” if one considers the pair of visual stimuli as a cycle, for which 180° yields the maximum possible phase offset) relative to flash onsets. In either of the potentially synchronous or asynchronous situations, there were five possible conditions: audio–visual congruent (same order of durations in the two modalities), audio–visual incongruent (opposite orders of durations in the two modalities), both-long auditory stimuli, both-short auditory stimuli, or a purely visual condition (c.f. **Figure 1**). The purely visual condition was of course actually equivalent for “synchronous” and “asynchronous” conditions, corresponding to the same condition arbitrarily divided into two separate datasets (random halves of the visual-only trials per participant), so as to a 5 × 2 factorial analysis of variance (ANOVA) on the data; see below.

Each block contained 10 repetitions for each of the 10 conditions in a randomized order. Every participant repeated three to four of these blocks.

### DATA ANALYSIS

For each participant we computed visual sensitivity ( $d'$ ) and criterion ( $c$ ) for the duration discrimination task, for each condition, using standard formulae as in Macmillan and Creelman (1997), namely:

$$d' = z(H) - z(F)$$

and

$$c = - [z(H) + z(F)]/2$$

where  $z(H)$  stands for the  $z$ -transform of the hit rate, while  $z(F)$  stands for the  $z$ -transform of the false-alarm rate. For any cases in which false-alarm rates were zero, we followed the conservative convention (as recommended by Snodgrass and Corwin, 1988; Macmillan and Creelman, 1991; c.f. Sarri et al., 2006) of adding a count of 0.5 to all cells within a single analysis.

$d'$  and criterion were analyzed using repeated-measures ANOVA, with SYNCHRONY (synchronous/asynchronous) and audio-visual CONDITION (Congruent; Incongruent; Short sounds; Long sounds, Purely visual) as within-subjects factors; followed up by pairwise  $t$ -tests where appropriate.

## RESULTS

The sensitivity ( $d'$ ) results are shown in **Figure 2A** (synchronous conditions) and **Figure 2B** (asynchronous conditions), as group means with SE. Recall that synchronous/asynchronous is a dummy factor solely for the visual-only condition, which was split randomly into two separate datasets. Note the higher sensitivity specifically in the synchronous audio-visual congruent condition (**Figure 2A**, leftmost bar). The overall  $5 \times 2$  ANOVA showed a main effect of SYNCHRONY [ $F(1,13) = 12.09, p < 0.01$ ], a significant main effect of audio-visual CONDITION [ $F(4,52) = 6.68, p < 0.001$ ] and critically a significant interaction between these two factors [ $F(4,52) = 5.96, p < 0.001$ ].

To identify the source of the interaction, first two separate one-way ANOVAs were performed for synchronous or asynchronous datasets, with the five-level factor of condition. While none of asynchronous conditions differed from each other [ $F(4,52) = 0.56, p = 0.69$  for the main effect], the synchronous conditions did [ $F(4,52) = 10.91, p < 0.00001$ ]. Exploratory pairwise  $t$ -tests for the asynchronous conditions confirmed no significant differences between any (all  $p > 0.20$ ). Pairwise  $t$ -tests for the synchronous conditions showed that sensitivity in the synchronous audio-visual congruent condition ( $d' = 1.93 \pm 0.23$  SE) was significantly higher than in all the other conditions,

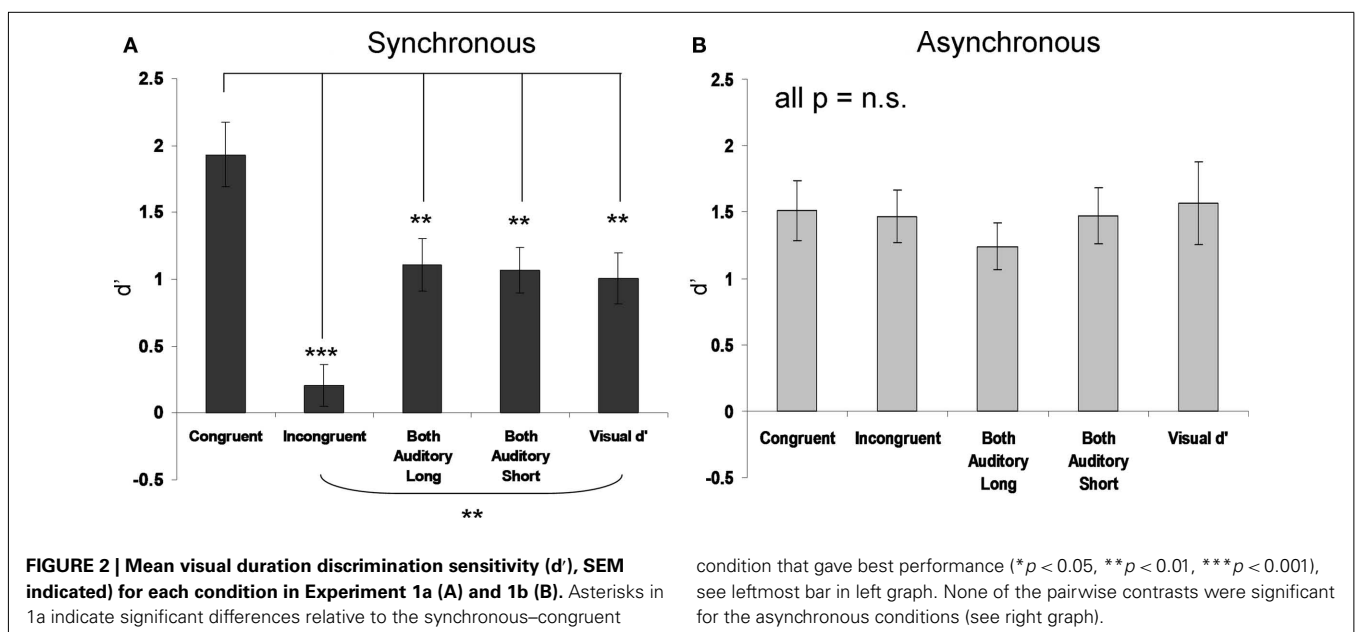
as follows: (i) versus the synchronous audio-visual incongruent condition [ $d' = 0.20 \pm 0.15$  SE;  $t(13) = 4.93; p < 0.001$ ]; (ii) versus the both-auditory-short condition [ $d' = 1.07 \pm 0.16$  SE;  $t(13) = 3.63; p < 0.01$ ]; (iii) versus the both-auditory-long condition [ $d' = 1.11 \pm 0.19$  SE;  $t(13) = 3.0, p = 0.01$ ]. The trend for the somewhat lower  $d'$  overall in the asynchronous than synchronous experiment did not approach significance.

When compared to the visual-only baseline measure ( $d' = 1.01 \pm 0.18$  SE), we found that: (i) visual duration discrimination was significantly enhanced in the synchronous audio-visual congruent condition [ $t(13) = -3.38; p < 0.01$ ]; (ii) was significantly impaired in the synchronous audio-visual incongruent condition [ $t(13) = 3.44, p < 0.01$ ]; (iii) was not significantly affected in the both-auditory-long or short conditions (all  $p > 0.71$ ).

A comparable two-way ANOVA on criterion scores instead revealed no significant results [main effect of SYNCHRONY ( $F(1,13) = 4.17, p = 0.07$ ); main effect of audio-visual CONDITION ( $F(4,52) = 0.35, p = 0.83$ ); interaction between the two ( $F(4,52) = 1.10, p = 0.36$ ); all n.s.].

## DISCUSSION

In Experiment 1 we found that objective duration discrimination for visual stimuli was objectively modulated by the duration of co-occurring auditory stimuli, but only when those auditory stimuli were synchronous with the visual events, rather than being delayed by 500 ms. Specifically, visual duration discrimination sensitivity ( $d'$ ) was enhanced for congruent-duration, synchronous auditory stimuli but decreased for incongruent-duration, synchronous auditory stimuli. Neither a sensitivity enhancement nor a sensitivity decrease could be observed for the asynchronous conditions, in which the sounds were now delayed by 500 ms so as to be (maximally) out-of-phase with the flashes. This elimination of the effect for the asynchronous case, together with the observed impact on  $d'$  (rather than criterion) for the synchronous cases, indicates a



genuine multisensory impact on visual perception, rather than a mere response bias as might be evident if a blind observer had to guess the response based solely on the sounds.

We suggest that the strong effect of auditory duration on perception of synchronous visual events reflects crossmodal binding between them (e.g., see also Meredith et al., 1987; Vroomen and Keetels, 2010; Colonius and Diederich, 2011), plus weighting of the auditory duration when bound due to the higher precision of temporal coding for audition than vision (see Introduction). A visual event is evidently perceived as longer when co-occurring with a slightly longer auditory event that is parsed as part of the same, single multisensory event. But we reasoned that if this is indeed a genuine perceptual effect as we document, rather than merely a response bias, then there should presumably be a limit to how far a longer auditory event can “stretch” perception of a visual event. If the auditory event were to endure much, much longer than the corresponding visual event, it should become less plausible that they arise from the same single external crossmodal event, and the auditory influence should begin to wane. We tested this in Experiment 2.

## EXPERIMENT 2

This study repeated the critical conditions where effects had been apparent in Experiment 1 (i.e., synchronous–congruent, synchronous–incongruent, plus visual-only baseline) except that now either the sounds had the same possible durations as the flashes, or else the sounds were tripled in duration such that they would no longer plausibly correspond to the flashes. A mere response bias, akin to a blind observer simply reporting the durations of the sounds, should lead to a similar outcome in either case; whereas the genuine perceptual effect we have documented should reduce (or even potentially disappear) in the mismatching tripled situation.

## METHODS

### Participants

Fifteen new participants with a mean age of 24.53 (range 18–34) took part, seven females, one left-handed, all reporting normal or corrected visual acuity, and normal hearing.

### Apparatus and stimuli

The setup was as for Experiment 1, but with less conditions and two new ones. We repeated the synchronous–congruent, synchronous–incongruent, and visual-only conditions. We also added two new conditions that were as for the synchronous–congruent and synchronous–incongruent except with tripled auditory durations. In these two new tripled conditions, each sound still onset simultaneously with each flash, but the sounds now lasted three times as long, so that with their much later offset they did not match the visual events so well.

## PROCEDURE

### Visual titration

This aspect of the procedure was the same as in Experiment 1. On average, participants were able to discriminate durations correctly for 76.33 ( $\pm 1.5$  SE)% of cases for those stimulus pairings that contained the longer visual stimulus identified as threshold. The

average duration of the longer visual stimuli identified as threshold was 94.5 ( $\pm 3.49$  SE) ms duration, so 49.5 ms longer than the standard stimulus, with the visual disks used.

## MAIN EXPERIMENT

The procedure resembled Experiment 1, but with only five conditions, two of which were new. The purely visual baseline, synchronous–congruent, and synchronous–incongruent conditions were as before. The two new conditions were tripled-sound-synchronous–congruent and tripled-sound-synchronous–incongruent condition. These were exactly like their untriple counterpart, except that the duration of each sound was three times as long.

## DATA ANALYSIS

For each participant and condition we computed sensitivity ( $d'$ ) and criterion ( $c$ ) for each stimulus condition as in Experiment 1.  $d'$  or criterion ( $c$ ) for the four audio–visual conditions were analyzed using repeated-measures two-way ANOVA, with SOUND LENGTH (tripled or untriple) and audio–visual duration CONGRUENCY (congruent versus incongruent) as factors. In addition pairwise  $t$ -tests compared performance in the purely visual baseline against the remaining conditions.

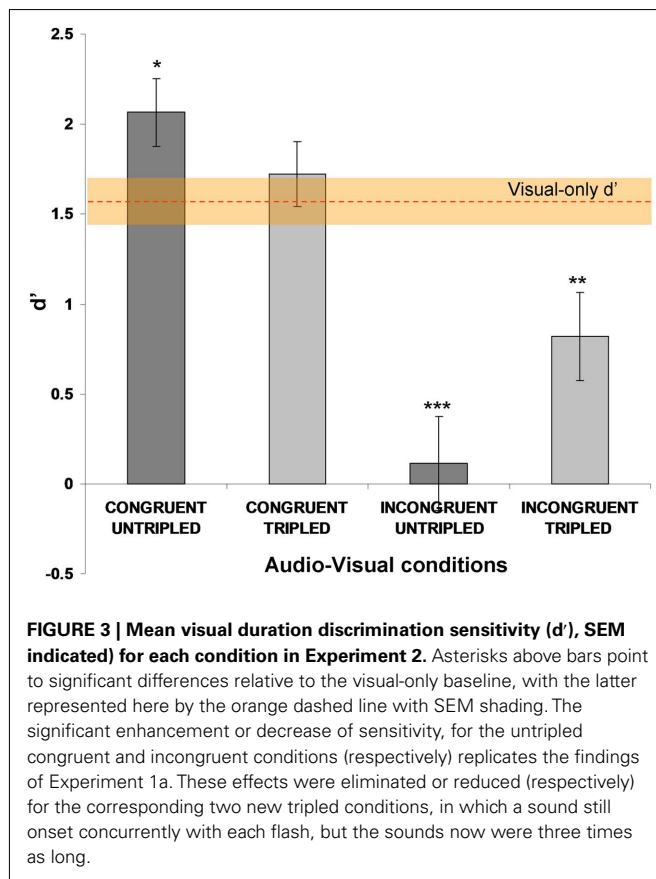
## RESULTS

The sensitivity ( $d'$ ) results are shown in **Figure 3**, as group means with SE. Note the higher sensitivity in the audio–visual untriple synchronous–congruent condition (leftmost bar), replicating the effects obtained in Experiment 1a (c.f. **Figure 2A**).

The two-way ANOVA showed no main effect of SOUND LENGTH [ $F(1,14) = 1.94, p = 0.18$ ], a significant main effect of CONGRUENCY [ $F(1,14) = 49.33, p < 0.00001$ ] and a significant interaction between these two factors [ $F(1,14) = 10.99, p < 0.01$ ], because the congruent/incongruent difference was larger in the untriple than tripled case. Sensitivity in the audio–visual untriple synchronous–congruent condition ( $d' = 2.07 \pm 0.16$  SE) was significantly higher than in all the other conditions, as follows: (i) versus the visual-only duration condition [ $d' = 1.58 \pm 0.16$  SE;  $t(14) = 2.86; p = 0.013$ ]; (ii) versus the tripled congruent condition [ $d' = 1.72 \pm 0.18$  SE;  $t(14) = 2.82; p = 0.014$ ]; (iii) versus the untriple incongruent condition [ $d' = 0.11 \pm 0.26$  SE;  $t(14) = 8.85; p < 0.000001$ ]; (iv) versus the tripled incongruent condition [ $d' = 0.82 \pm 0.25$  SE;  $t(14) = 4.8, p < 0.001$ ].

Thus, by prolonging the duration of the auditory stimuli to triple that of the visual stimuli, the significant enhancement (relative to visual-only baseline) obtained for the congruent audio–visual durations was abolished. While even triple-duration auditory stimuli still produced some sensitivity decrease for incongruent stimuli, even this remaining decrease was still significantly reduced for the tripled versus untriple case [ $t(14) = -2.69, p = 0.018$ ].

A comparable analysis of criterion scores instead revealed no significant results [e.g., for the two-way ANOVA, main effects of SOUND LENGTH ( $F(1,14) = 3.29, p = 0.09$ ); main effect of CONGRUENCY ( $F(1,14) = 0.002, p = 0.96$ ); interaction between the two ( $F(1,14) = 2.15, p = 0.16$ ), all n.s.].



## DISCUSSION

We replicated the results of Experiment 1a for the shared conditions in Experiment 2, showing significant enhancement of objective visual duration discrimination sensitivity ( $d'$ ) by congruent auditory stimuli and a significant decrease of sensitivity for incongruent stimuli. The new finding was that the sensitivity enhancement for congruent stimuli (i.e., same pairwise successive order of longer and shorter) relative to the visual-only baseline was completely abolished for prolonged auditory stimuli that onset concurrently with the visual flashes but endured three times longer. Some sensitivity decrease for incongruent audio-visual pairings remained but at a significantly reduced level in the tripled-sound-duration case. These results show that the impact of auditory durations on visual duration discrimination is larger when the sounds and flashes endure for a similar order of magnitude, being significantly diminished when the sounds endure three times longer than the flashes.

## GENERAL DISCUSSION

In two experiments we tested the influence of sound duration on objective discrimination of the duration of visual events. Our results accord with but go beyond previous reports (e.g., Walker and Scott, 1981; Donovan et al., 2004; Chen and Yeh, 2009; Klink et al., 2011) that auditory stimuli can impact significantly on judgments for the duration of co-occurring visual stimuli. The new findings extend previous work by showing that: (1) not only can incongruent auditory stimuli significantly impair objective visual

performance, but congruent auditory stimuli can benefit visual duration judgments; (2) this applies for visual sensitivity ( $d'$ ) to visual duration in signal-detection terms, rather than affecting mere response bias or criterion; (3) this impact of auditory duration on perception of visual duration depends on whether the audio-visual onsets are synchronous, being eliminated when the sounds lagged here; (4) it also depends on whether the auditory and visual events are similar in length, being reduced, or eliminated when the sounds are triple the duration of corresponding visual events.

A previous study by Donovan et al. (2004) used a similar approach to ours. They investigated the influence of task-irrelevant auditory information on a visual task, but with participants judging whether two sequential visual events were presented for the same or different lengths of time. They found a similar trend to the present pattern for congruent versus incongruent audio-visual conditions, and their effect was abolished for asynchronous conditions, reminiscent of the present Experiment 1b. But Donovan et al. (2004) did not titrate the duration of visual stimuli to a threshold level; did not present a visual-only condition to provide a baseline for assessing any multisensory benefit or cost; did not calculate signal-detection scores.

More recently, Klink et al. (2011) extensively tested cross-modal influences on duration perception, confirming auditory over visual dominance for time perception. These authors adopted a duration discrimination task, similar to the one presented here. In line with our results, they found a reduction in visual duration discrimination accuracy with incongruent auditory stimuli. But they did not test the effect of audio-visual congruent stimuli, and their control for possible response biases was very different to our own (a grouping experiment, Experiment 5 in Klink et al., 2011). Our findings accord with such prior work (see also Introduction) in showing a clear influence of audition on visual duration judgments. We show in signal-detection terms that visual  $d'$  for duration discrimination can not only be impaired by incongruent auditory timing information, but for the first time, also significantly enhanced by congruent auditory information; plus we document some of the boundary conditions for this [in terms not only of synchronous onset across the modalities (see Experiment 1 here) but also in terms of fairly well-matching duration scale (see Experiment 2 here)].

How could such auditory influences over visual duration perception arise? An extensive literature on the many possible mechanisms for time perception has built up (for recent reviews, see Ivry and Schlerf, 2008; Grondin, 2010). One long tradition in the field of time perception posits the presence of central mechanisms for time estimation, such as an internal clock or clocks (e.g., Ivry and Richardson, 2002; Rosenbaum, 2002; Schönner, 2002; Wing, 2002). Some research suggests that such central-clock(s) might operate supramodally. In apparent support of this view are findings that several brain areas (e.g., see Ivry and Keele, 1989; Harrington et al., 1998; Rao et al., 2001; Leon and Shadlen, 2003; Coull et al., 2004; Buetti et al., 2008) are implicated in estimation and representation of time independently from the sensory modality of the stimuli in question, although it should be noted that some of the time judgments used were on longer scales than here.

Other authors have argued that there may be no need to invoke internal “clocks” to describe some timing behaviors (e.g., Zeiler, 1999; Jantzen et al., 2005; Karmarkar and Buonomano, 2007; Ivry and Schlerf, 2008). Recent findings in the field of visual perception, for example, have led to the development of seemingly more modality-specific perspectives (e.g., Yarrow et al., 2001; Ghose and Maunsell, 2002; Morrone, 2005; Johnston et al., 2006; Shuler and Bear, 2006; Xuan et al., 2007), suggesting that estimates for the duration of visual signals could be embedded within the visual modality itself.

Here we showed that visual duration sensitivity can be significantly impaired or enhanced by auditory stimuli that are likely to be parsed as reflecting the same external “event” as the affected visual event. The central-clock perspective might consider this to arise at some internal timing process that is shared between modalities. On the other hand, our results might also be reconciled with visual duration judgments arising within visual cortex itself, provided it is acknowledged that auditory can also impact upon visual cortex (for which an increasing body of evidence now exists; e.g., Martuzzi et al., 2007; Romei et al., 2007, 2009; Wang et al., 2008; Bolognini et al., 2010; Cappe et al., 2010; Noesselt et al., 2010; Bolognini and Maravita, 2011; c.f. Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008 for extensive review). It would be useful to combine the present behavioral paradigm with neural measures in future work; and also to study the impact of neural disruptions, such as transcranial magnetic stimulation (TMS), targeting visual cortex or auditory cortex or heteromodal cortex (see Kanai et al., 2011). Other future extensions of our paradigm could investigate whether lagging auditory events by different parametric amounts in the asynchronous condition would lead to a graded or categorical change in results; and the possible impact of introducing

asynchrony by making auditory events lead instead. Here we had lagged the sounds in our asynchronous condition by a full 500 ms, in order to generate the maximum 180° shift to visual and auditory events being “out-of-phase” in terms of the cycle we used.

A further interesting question for future extensions of our paradigm concerns the possible role of attention in the multisensory effect on sensitivity that we have identified. Indeed there is now a growing literature on the possible role of attention in multisensory integration (see Sanabria et al., 2007; Talsma et al., 2010). Studies of some multisensory phenomena suggesting no role for attention (e.g., Bertelson et al., 2000); while others on different multisensory phenomena suggest a key attentional role (e.g., van Ee et al., 2009). Accordingly it is an empirical issue whether the new multisensory phenomena that we have uncovered may depend on attention or not. The present results already make clear that audio–visual integration can genuinely affect visual sensitivity ( $d'$ ) here, and that this depends on audio–visual synchrony.

Finally, given the evidently perceptual nature of the objective improvements in visual duration discrimination that we observed here due to appropriately timed sounds, it would be intriguing to study whether a slightly longer sound paired with a concurrent visual event can not only extend the apparent duration of that visual event, but actually improve visual perception of its *other* (non-temporal) visual qualities, in a similar manner to the visual improvement found for a genuinely longer visual stimulus (see Berger et al., 2003).

## ACKNOWLEDGMENTS

Vincenzo Romei, Benjamin De Haas, and Jon Driver were supported by the Wellcome Trust. Jon Driver is a Royal Society Anniversary Research Professor.

## REFERENCES

- Aaen-Stockdale, C., Hotchkiss, J., Heron, J., and Whitaker, D. (2011). Perceived time is spatial frequency dependent. *Vision Res.* 51, 1232–1238.
- Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262.
- Berger, T. D., Martelli, M., and Pelli, D. G. (2003). Flicker flutter: is an illusory event as good as the real thing? *J. Vis.* 3, 406–412.
- Bertelson, P., and Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension: 1. evidence from auditory – visual temporal order judgment. *Int. J. Psychophysiol.* 50, 147–155.
- Bertelson, P., and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys.* 29, 578–584.
- Bertelson, P., Vroomen, J., de Gelder, B., and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept. Psychophys.* 62, 321–332.
- Bolognini, N., and Maravita, A. (2011). Uncovering multisensory processing through non-invasive brain stimulation. *Front. Psychol.* 2:46. doi: 10.3389/fpsyg.2011.00046
- Bolognini, N., Senna, I., Maravita, A., Pascual-Leone, A., and Merabet, L. B. (2010). Auditory enhancement of visual phosphene perception: the effect of temporal and spatial factors and of stimulus intensity. *Neurosci. Lett.* 477, 109–114.
- Bueti, D., Bahrami, B., and Walsh, V. (2008). The sensory and association cortex in time perception. *J. Cogn. Neurosci.* 20, 1–9.
- Burr, D., and Alais, D. (2006). Combining visual and auditory information. *Prog. Brain Res.* 155, 243–258.
- Calvert, G. A., Spence, C., and Stein, B. E. (eds). (2004). *The Handbook of Multisensory Processes*. Cambridge: MIT Press.
- Cappe, C., Thut, G., Romei, V., and Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia* 47, 1045–1052.
- Cappe, C., Thut, G., Romei, V., and Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.* 30, 12572–12580.
- Chen, K., and Yeh, S. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychol. (Amst.)* 130, 225–234.
- Colonus, H., and Diederich, A. (2011). Computing an optimal time window of audiovisual integration in focused attention tasks: illustrated by studies on effect of age and prior knowledge. *Exp. Brain Res.* 212, 327–337.
- Coull, J. T., Vidal, F., Nazarian, B., and Macar, F. (2004). Functional anatomy of the attentional modulation of time estimation. *Science* 303, 1506–1508.
- Donovan, C. L., Lindsay, D. S., and Kingstone, A. (2004). Flexible and abstract resolutions to crossmodal conflicts. *Brain Cogn.* 56, 1–4.
- Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron* 57, 11–23.
- Eagleman, D. M. (2008). Human time perception and its illusions. *Curr. Opin. Neurobiol.* 18, 131–136.
- Ernst, M. O., and Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci. (Regul. Ed.)* 8, 162–169.
- Freeman, E., and Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Curr. Biol.* 18, 1262–1266.
- Gehard, J. W., and Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *Am. J. Psychol.* 72, 521–529.
- Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception* 36, 1089–1103.
- Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci. (Regul. Ed.)* 10, 278–285.
- Ghose, G. M., and Maunsell, J. H. R. (2002). Attentional modulation in visual cortex depends on task timing. *Nature* 419, 616–620.
- Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and

- theoretical directions. *Atten. Percept. Psychophys.* 72, 561–582.
- Guttman, S. E., Gilroy, L. A., and Blake, R. (2005). Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychol. Sci.* 16, 228–235.
- Harrington, D. L., Haaland, K. Y., and Hermanowicz, N. (1998). Temporal processing in the basal ganglia. *Neuropsychology* 12, 3–12.
- Howard, I. P., and Templeton, W. B. (1966). *Human Spatial Orientation*. London: Wiley.
- Ivry, R., and Keele, S. (1989). Timing functions of the cerebellum. *J. Cogn. Neurosci.* 1, 36–152.
- Ivry, R. B., and Richardson, T. C. (2002). Temporal control and coordination: the multiple timer model. *Brain Cogn.* 48, 117–132.
- Ivry, R. B., and Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends Cogn. Sci. (Regul. Ed.)* 12, 273–280.
- Jantzen, K. J., Steinberg, F. L., and Kelso, J. A. S. (2005). Functional MRI reveals the existence of modality and coordination-dependent timing networks. *Neuroimage* 25, 1031–1042.
- Johnston, A., Arnold, D. H., and Nishida, S. (2006). Spatially localised distortions of event time. *Curr. Biol.* 16, 472–479.
- Kanai, R., Lloyd, H., Buetti, D., and Walsh, V. (2011). Modality-independent role of the primary auditory cortex in time estimation. *Exp. Brain Res.* 209, 465–471.
- Karmarkar, U. R., and Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron* 53, 427–438.
- Klink, P. C., Montijn, J. S., and van Wezel, R. J. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Atten. Percept. Psychophys.* 73, 219–236.
- Leo, F., Romei, V., Freeman, E., Ladavas, E., and Driver, J. (2011). Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds. *Exp. Brain Res.* 213, 193–201.
- Leon, M. I., and Shadlen, M. N. (2003). Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* 38, 317–327.
- Macmillan, N. A., and Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.
- Macmillan, N. A., and Creelman, C. D. (1997). d'Plus: a program to calculate accuracy and bias measures from detection and discrimination data. *Spat. Vis.* 11, 141–143.
- Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J. P., Maeder, P. P., Clarke, S., and Meuli, R. A. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb. Cortex* 17, 1672–1679.
- Meredith, M. A., Nemitz, J. W., and Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.* 7, 3215–3229.
- Morein-Zamir, S., Soto-Faraco, S., and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Brain Res. Cogn. Brain Res.* 17, 154–163.
- Morrone, M. C. (2005). Saccadic eye movements cause compression of time as well as space. *Nat. Neurosci.* 8, 950–954.
- Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H. J., and Driver, J. (2010). Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *J. Neurosci.* 30, 13609–13623.
- Rao, S. M., Mayer, A. R., and Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nat. Neurosci.* 4, 317–323.
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *J. Neurophysiol.* 89, 1078–1093.
- Repp, B. H., and Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1085–1099.
- Romei, V., Murray, M. M., Cappe, C., and Thut, G. (2009). Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Curr. Biol.* 19, 1799–1805.
- Romei, V., Murray, M. M., Merabet, L. B., and Thut, G. (2007). Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. *J. Neurosci.* 27, 11465–11472.
- Rosenbaum, D. A. (2002). Time, space, and short-term memory. *Brain Cogn.* 48, 52–65.
- Sanabria, D., Soto-Faraco, S., and Spence, C. (2007). Spatial attention and audiovisual interactions in apparent motion. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 927–937.
- Sarri, M., Blankenburg, F., and Driver, J. (2006). Neural correlates of cross-modal visual-tactile extinction and of tactile awareness revealed by fMRI in a right-hemisphere stroke patient. *Neuropsychologia* 44, 2398–2410.
- Schöner, G. (2002). Timing, clocks, and dynamical systems. *Brain Cogn.* 48, 31–51.
- Shams, L., Kamitani, Y., and Shimojo, S. (2000). What you see is what you hear. *Nature* 408, 788.
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science* 145, 1328–1330.
- Shuler, M. G., and Bear, M. F. (2006). Reward timing the primary visual cortex. *Science* 311, 1606–1609.
- Snodgrass, J. G., and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia, and amnesia. *J. Exp. Psychol. Gen.* 117, 34–50.
- Spence, C., and Driver, J. (eds). (2004). *Crossmodal Space and Crossmodal Attention*. Oxford: Oxford University Press.
- Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci. (Regul. Ed.)* 14, 400–410.
- Thurlow, W. R., and Jack, C. E. (1973). Certain determinants of the “ventriloquism effect.” *Percept. Mot. Skills* 36, 1171–1184.
- van Ee, R., van Boxtel, J. J., Parker, A. L., and Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *J. Neurosci.* 29, 11641–11669.
- van Wassenhove, V., Buonomano, D. V., Shimojo, S., and Shams, L. (2008). Distortions of subjective time perception within and across senses. *PLoS ONE* 3, e1437. doi: 10.1371/journal.pone.0001437
- Vroomen, J., and de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1583–1590.
- Vroomen, J., de Gelder, B., and Vroomen, J. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 513–518.
- Vroomen, J., and Keetels, M. (2009). Sounds change four-dot masking. *Acta Psychol. (Amst.)* 130, 58–63.
- Vroomen, J., and Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Atten. Percept. Psychophys.* 72, 871–884.
- Wada, Y., Kitagawa, N., and Noguchi, K. (2003). Audio-visual integration in temporal perception. *Int. J. Psychophysiol.* 50, 117–124.
- Walker, J. T., and Scott, K. J. (1981). Auditory–visual conflicts in the perceived duration of lights, tones and gaps. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 1327–1339.
- Wang, Y., Celebrini, S., Trotter, Y., and Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci.* 9, 79. doi: 10.1186/1471-2202-9-79
- Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 88, 638–667.
- Wing, A. M. (2002). Voluntary timing and brain function: an information processing approach. *Brain Cogn.* 48, 7–30.
- Witten, I. B., and Knudsen, E. I. (2005). Why seeing is believing: merging auditory and visual worlds. *Neuron* 48, 489–496.
- Xuan, B., Zhan, D., He, S., and Chen, X. (2007). Larger stimuli are judged to last longer. *J. Vis.* 7, 1–5.
- Yarrow, K., Haggard, P., Heal, R., Brown, P., and Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature* 414, 302–305.
- Zeiler, M. D. (1999). Time without clocks. *J. Exp. Anal. Behav.* 71, 288–291.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 July 2011; accepted: 18 August 2011; published online: 08 September 2011.

Citation: Romei V, De Haas B, Mok RM and Driver J (2011) Auditory stimulus timing influences perceived duration of co-occurring visual stimuli. *Front. Psychology* 2:215. doi: 10.3389/fpsyg.2011.00215

This article was submitted to *Frontiers in Perception Science*, a specialty of *Frontiers in Psychology*.

Copyright © 2011 Romei, De Haas, Mok and Driver. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.