Published in final edited form as:

Atten Percept Psychophys. 2014 July ; 76(5): 1485–1502. doi:10.3758/s13414-014-0663-x.

Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability

Laura Ortega¹, Emmanuel Guzman-Martinez¹, Marcia Grabowecky^{1,2}, and Satoru Suzuki^{1,2}

¹Department of Psychology, Northwestern University, Evanston, Illinois

²Interdepartmental Neuroscience Program, Northwestern University, Evanston, Illinois

Abstract

Whereas the visual modality tends to dominate over the auditory modality in bimodal spatial perception, the auditory modality tends to dominate over the visual modality in bimodal temporal perception. Recent results suggest that the visual modality dominates bimodal spatial perception because spatial discriminability is typically greater for the visual than auditory modality; accordingly, visual dominance is eliminated or reversed when visual-spatial discriminability is reduced by degrading visual stimuli to be equivalent or inferior to auditory spatial discriminability. Thus, for spatial perception, the modality that provides greater discriminability dominates. Here we ask whether auditory dominance in duration perception is similarly explained by factors that influence the relative quality of auditory and visual signals. In contrast to the spatial results, the auditory modality dominated over the visual modality in bimodal duration perception even when the auditory signal was clearly weaker, when the auditory signal was ignored (i.e., the visual signal was selectively attended), and when the temporal discriminability was equivalent for the auditory and visual signals. Thus, unlike spatial perception where the modality carrying more discriminable signals dominates, duration perception seems to be mandatorily linked to auditory processing under most circumstances.

Send correspondence to: Satoru Suzuki or Laura Ortega Northwestern University Department of Psychology 2029 Sheridan Rd., Evanston, IL 60208 Phone: (847) 467-1271 Fax: (847) 491-7859 satoru@northwestern.edu, or lauraortegat@gmail.com.

¹ In Figure 3A it may seem puzzling that the PSEs were not overall shorter in the blocks with visually presented reference durations (left panel) relative to the blocks with auditorily presented reference durations (right panel) because visually presented reference durations should have been perceived to be shorter than auditorily presented ones. Because the visual and auditory presentations of the reference durations were randomly intermixed across the 18 relatively short blocks of trials, it is possible that participants pooled the reference information from the two modalities in memory (e.g., Penney et al., 1998). Regardless, because the unimodal and bimodal trials were randomly intermixed within each block, durations of all unimodal and bimodal stimuli would have been judged relative to the same internal reference durations within each block. Thus, although it is difficult to interpret the absolute levels of PSEs, the comparison of PSEs across the visual-alone, auditory-alone, and bimodal conditions is valid, so that we can make conclusions regarding whether the visual or auditory modality dominated duration perception for the bimodal stimuli.

² Digital signage monitors are very bright.

³ Although the modality in which the reference durations were presented did not significantly influence the PSEs, the PSEs were overall shorter with the visually presented references (690 ms for the visual-alone, 634 ms for the auditory-alone, and 620 ms for the bimodal stimulus) than with the auditorily presented references (713 ms for the visual-alone, 655 ms for the auditory-alone, and 659 ms for the bimodal stimulus). This trend is consistent with the expectation that visually presented references would be perceived to be shorter than auditorily presented references, thus increasing the perceived test durations (i.e., reducing the PSEs) overall.

⁴ The fact that our luminance-decrement detection task yielded about 80% accuracy indicates that the task was indeed attention demanding. This, however, does not rule out the possibility that a more radical manipulation of visual selective attention might diminish auditory dominance in bimodal duration perception.

Keywords

Time perception; Modality effect; Auditory dominance; Selective attention

Introduction

Auditory-visual interactions have been extensively studied in relation to the perception of spatial location (e.g., Stein, Meredith, Huneycutt, & McDade, 1989; Andersen, Snyder, Bradley, & Xing, 1997; Stein, 1997; Driver & Spence, 1998; Frassinetti, Bolognini, & Làdavas, 2002; Alais & Burr, 2004; Bolognini, Frassinetti, Serino, & Làdavas, 2005), of objects (e.g., Beauchamp, Lee, Argall, & Martin, 2004; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005; Molholm, Martinez, Ritter, Javitt, & Foxe, 2005; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005; Smith, Grabowecky, & Suzuki, 2007; Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008), of motion (e.g., Sekuler, Sekuler, & Lau, 1997; Lewis, Beauchamp, & DeYoe, 2000; Meyer & Wuerger, 2001; Grassi & Casco, 2009), and of timing of events (e.g., Shimojo & Shams, 2001; McDonald, Teder-Salejarvi, Di Russo, & Hillyard, 2005). However, relatively little is known about auditory-visual interactions in the perception of duration.

In particular, processing of short durations is crucial for fine motor behavior as well as for the perception of motion, rhythm, and speech (see Mauk & Buonomano, 2004 for a review). Furthermore, short durations up to about a second are directly perceived (e.g., Johnston, Arnold, & Nishida, 2006; Ortega, Guzman-Martinez, Grabowecky, & Suzuki, 2012), independently of cognitive strategies such as counting (e.g., Grondin, Meilleur-Wells, & Lachance, 1999). We thus investigated how auditory and visual modalities influence the perceived duration of a short bimodal event.

We took advantage of the well-known phenomenon where an auditory stimulus is perceived to be longer than a visual stimulus of the same physical duration (e.g., Droit-Volet, Meck, & Penney, 2007; Goldstone & Goldfarb, 1964; Grondin, Meilleur-Wells, Ouellette, & Macar, 1998; Ortega, Lopez, & Church, 2009; Penney, 2003; Wearden, Edwards, Fakhri, & Percival, 1998; Wearden, Todd, & Jones, 2006). It is thought that auditory stimuli are perceived to be longer partly because an internal clock in the auditory modality may run at a faster rate than a visual clock, generating a greater number of accumulated “ticks.” (e.g., Penney, Allan, Meck, & Gibbon, 1998; Penney, Gibbon, & Meck, 2000; Wearden, Edwards, Fakhri, & Percival, 1998; Wearden, Todd, & Jones, 2006). This modality difference allows us to determine how auditory and visual components contribute to the perceived duration of a bimodal stimulus by comparing the perceived duration of the bimodal stimulus with those of the auditory and visual components presented alone. Is a bimodal stimulus perceived to be as long as the auditory component (indicating auditory dominance in bimodal duration perception), as short as the visual component (indicating visual dominance in bimodal duration perception), or somewhere in between (indicating weighted contributions from both modalities)?

Two previous studies (Goldstone, Boarman, & Lhamon, 1959; Walker & Scott, 1981) investigated this question, and suggested that increased stimulus intensity of one modality increases the contribution of that modality (at least for some durations) and that selective attention has little effect. However, these results are not definitive partly because they assumed rather than measured the relative perceived strength of the auditory and visual stimuli, and partly because, although Goldstone et al. (1959) instructed participants to attend to either the auditory or visual component of the bimodal stimulus, they did not verify that participants indeed allocated attention as instructed. In addition, an important factor that might influence the relative contribution of auditory and visual signals in bimodal duration perception is temporal discriminability. For spatial perception, the perceived location of a bimodal stimulus is dominated by the modality that provides greater spatial discriminability, with the weighting of the visual and auditory signals close to being statistically optimal (Alais & Burr, 2004). The perceived duration of a bimodal stimulus might be similarly dominated by the modality that provides greater duration discriminability. Neither the method of duration measurement used by Goldstone et al. (1959) (judging perceived durations relative to participants' own representation of one second) nor that used by Walker and Scott (1981) (motor reproduction of perceived duration) provided a measure of duration discriminability.

The goal of the current study was to systematically investigate the dependence of bimodal duration perception on (1) the relative intensity of auditory and visual signals, (2) selective attention to visual signals, and (3) the relative duration discriminability in the two modalities. In particular, if auditory dominance in duration perception occurs regardless of relative intensity, visual selective attention, and duration discriminability, this would support the recent view that duration information is extracted from auditory signals automatically and preattentively (e.g., Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001; Guttman, Gilroy, & Blake, 2005).

Experiment 1: Effects of sound intensity on the perceived duration of auditory-visual bimodal stimuli

We examined how the relative intensity of auditory and visual components influenced the perceived duration of a bimodal stimulus. We used a constant visual intensity, and psychophysically determined three auditory intensities relative to the visual intensity: (1) an intensity subjectively matched to the visual intensity, (2) an intensity clearly lower than the visual intensity, and (3) an intensity clearly higher than the visual intensity. We used a standard "temporal-bisection" task that yielded the Point of Subjective Equality (PSE) as a measure of perceived duration, and the Just Noticeable difference (JND) as a measure of duration discriminability. The PSE is the physical duration perceived to be half way between the short and long reference durations, and the associated JND is the amount of deviation from the PSE required to be reliably perceived as closer to the short or the long reference duration. A longer PSE (i.e., a longer physical duration is required to be perceived as half way between the short and long reference durations) indicates shorter perceived duration, and a longer JND indicates poorer duration discrimination.

We measured the perceived durations of three bimodal stimuli, the visual stimulus combined with the lower-intensity auditory stimulus (V+lower_A), the visual stimulus combined with the intensity-matched auditory stimulus (V+matched_A), and the visual stimulus combined with the higher-intensity auditory stimulus (V+higher_A), and two unimodal stimuli, the visual stimulus presented alone (V), and the intensity-matched auditory stimulus presented alone, (matched_A). If duration perception depends on the modality that receives the stronger perceptual signal, the perceived duration of a bimodal V+lower_A stimulus should be similar to the perceived duration of a unimodal visual stimulus, and the perceived duration of a bimodal V+higher_A stimulus should be similar to the perceived duration of a unimodal auditory stimulus.

Method

Participants: Nine Northwestern University undergraduate students (6 female) gave informed consent to participate for partial course credit. In all experiments, all participants had normal or corrected-to-normal visual acuity, normal color vision, and normal hearing, were naïve to the purpose of the experiment, and were tested individually in a quiet and normally lit room.

Apparatus: Visual stimuli were displayed on a 19" color CRT monitor (75 Hz, 1152 by 870 resolution) at 60 cm viewing distance. Auditory stimuli were presented via HD 265 Sennheiser headphones (10 Hz - 25 kHz frequency response). The experiment was controlled with Apple MacBook OS X using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, & Pelli, 2007) for MATLAB.

Stimuli: The visual stimulus was a blue circle (17.2 cd/m^2 ; CIE[.175, .119]) subtending 4.90° of visual angle, and it was presented at the center of the screen against a white background (100.8 cd/m^2). The auditory stimulus was a 500 Hz tone with one of 3 different sound levels: an intensity clearly lower than the visual intensity (54 dB SPL-A), an intensity matched to the visual intensity (59 dB SPL-A), or an intensity clearly higher than the visual intensity (64 dB SPL-A).

These sound levels were determined in a psychophysical calibration experiment with four participants who did not participate in the main experiment. On each trial of the calibration experiment, the blue circle and a tone (randomly chosen from 12 sound levels ranging from 46 to 65 dB SPL-A) were simultaneously presented for one second. The participant responded as to whether the circle or the sound was more intense by pressing a corresponding button on the computer keyboard. The next trial began 500 ms after the participant's response. A total of 144 trials were run in four blocks of 36 trials with each sound level presented three times per block. We obtained a psychometric function for each participant, and determined the subjectively matched sound level (59 dB SPL-A) corresponding to the PSE that was perceived to be equivalent to the visual intensity of the circle, the lower-intensity sound level (54 dB SPL-A) that was clearly audible but judged as weaker than the visual circle by all observers nearly 100% of the time, and the higher-intensity sound level (64 dB SPL-A) that was judged as stronger than the visual circle by all observers nearly 100% of the time (Figure 1). Because of the way in which the 12 sound

levels were distributed, the PSE of 59 dB SPL-A (58.9 with $SD=0.8$) was close to the median value. There was thus a concern that participants used the strategy of choosing the median value of the presented sound levels as the one that matched the visual intensity of the circle rather than crossmodally comparing the auditory and visual intensities. We thus repeated the calibration experiment (with two of the original participants and two new participants) using a different range of sound levels so that the median level was 55 dB SPL-A. If the participants used the median matching strategy, the PSE should have decreased to 55 dB, whereas if participants perceptually matched 59 dB to the visual intensity of the circle in the original experiment, the PSE should have remained near 59 dB. The PSE was essentially unchanged at 58.2 dB SPL-A ($SD=0.6$), with 54 dB SPL-A still consistently judged as lower intensity than the visual circle, and 64 dB SPL-A still consistently judged as higher intensity than the visual circle.

For the main experiment to measure bimodal duration perception, durations of the visual, auditory, and bimodal stimuli were, 200, 300, 400, 500, 600, 700 and 800 ms. The 200 ms duration served as the short reference and the 800 ms duration served as the long reference for the temporal-bisection task (see below).

Procedure: The 2-alternative-forced-choice (2AFC) temporal-bisection task consisted of 18 blocks of trials with each block consisting of a reference phase and a test phase.

In the reference phase, the short and long reference durations were alternately presented across six trials (the alternate presentation helped participants to contrast the short and long reference durations). The reference duration was presented either visually or auditorily (intensity matched to the visual stimulus) in a given block; that is, nine randomly chosen blocks began with the visual reference trials; the remaining nine blocks began with the auditory reference trials. At the beginning of each reference trial, the initial display indicated which reference (short or long) would be presented, followed by a presentation of the corresponding reference duration, and a prompt to respond by pressing the appropriate button (Figure 2A). The next trial began one second after the participant's response.

In the test phase, each of the five stimulus types, (1) visual, V, (2) matched-intensity auditory, A_matched, (3) visual with lower-intensity auditory, V+lower_A, (4) visual with matched-intensity auditory, V+matched_A, and (5) visual with higher-intensity auditory, V+higher_A, was presented for the seven durations including the two reference durations. Each test trial began with a blank screen lasting one second, followed by the stimulus of the selected type and duration (both stimulus type and duration were randomly intermixed across trials). The participant then responded as to whether the stimulus duration was closer to the short or long reference by pressing the corresponding button (Figure 2B). The next trial began one second after the participant's response.

Each block consisted of six reference trials followed by 35 test trials (5 stimulus types $\times$ 7 durations), and each participant was tested in 18 such blocks. The regularly inserted reference trials were intended to maintain stable memory representations of the reference durations.

Results—We analyzed the proportion of “long” responses as a function of stimulus duration for the five stimulus conditions. We fit a logistic function to each individual psychometric function to compute the PSE (the stimulus duration for which the responses were equally often “long” and “short”) and the JND. The logistic function we used has the form,

$$y = \frac{1}{1 + e^{-(x-a)/b}},$$

where x represents the stimulus duration (in ms), a represents the PSE (in ms), and b represents the JND (in ms) around the PSE (we use $1.0986 \times b$ as JND because this value indicates the deviation from the PSE that yields 75% correct “long” or “short” classification). Logistic functions produced good fits to individual participants’ data (mean $\omega^2 = 0.95$ across all experiments).

Note that the relationship between the PSE and the perceived stimulus duration is inverted. A longer PSE indicates that the perceived stimulus duration was shorter because a longer physical duration was necessary to be perceived as half way between the short and long reference durations. In contrast, a shorter PSE indicates that the perceived stimulus duration was longer because a shorter physical duration was sufficient to be perceived as half way between the short and long reference durations.

The PSE was the longest for the visual-alone stimulus compared with all other stimuli whether the reference durations were presented visually or auditorily (Figure 3A).¹ For both reference types, the PSE was significantly longer for the visual-alone stimulus compared with the auditory-alone and all bimodal (V+A) stimuli ($t[8]s > 4.194$, $ps < 0.03$ for visual reference and $t[8]s > 6.734$, $ps < 0.002$ for auditory reference) and the PSEs were equivalent for the auditory-alone and bimodal stimuli ($t[8]s < 2.701$, $n.s.$ for auditory reference and $t[8]s < 2.947$, $n.s.$ for the visual reference, except that the PSE for the V+matched_A stimulus was shorter than that for the auditory-alone stimulus, $t[8] = 4.215$, $p < 0.03$, in the visual reference condition). All p values have been Bonferroni corrected for multiple pairwise comparisons (10 possible comparisons across the five conditions). Importantly, none of the PSEs for the bimodal stimuli were longer than that for the auditory-alone stimulus, indicating that none of the PSEs for the bimodal stimuli were “pulled” in the direction of the visual PSE. Because PSEs are inversely related to perceived durations, this means that the visual-alone stimulus was perceived to be the shortest in duration while all bimodal stimuli were perceived to be about the same duration as the auditory-alone stimulus.

The JNDs for the bimodal stimuli are also more similar to the JND for the auditory-alone stimulus than to the JND for the visual-alone stimulus. In the visual reference condition (Figure 3B, left), the JND was significantly greater for the visual-alone stimulus compared with the auditory-alone and bimodal stimuli ($t[8]s > 4.251$, $ps < 0.03$, except that the difference was marginal for the V+lower_A stimulus $t[8] = 3.488$, $p < 0.083$). The JNDs were equivalent for the auditory-alone and bimodal stimuli ($t[8]s < 1.291$, $n.s.$, except that JND was slightly but significantly greater for the V+lower_A stimulus than for the auditory-alone stimulus, $t[8] = 3.985$, $p < 0.04$). In the auditory reference condition, we obtained the

same pattern of JND results, but none of the comparisons were statistically significant, $t(8)s < 2.798$, *n.s.* All *p* values have been Bonferroni corrected for multiple pair-wise comparisons. Overall, duration discriminability was superior for the auditory than visual modality, and bimodal discriminability was similar to auditory discriminability.

In summary, we replicated the previous results that a visually presented duration is perceived to be shorter than an auditorily presented duration. Importantly, we showed that the perceived durations of the bimodal stimuli were equivalent to the perceived duration of the auditory stimulus irrespective of the relative intensity of the auditory and visual components, and even when the visual component was clearly more intense than the auditory component.

Experiment 2: Effects of visual selective attention on auditory dominance in duration perception

We investigated whether selectively attending to a visual component could overcome auditory dominance in duration perception. When auditory and visual stimuli are synchronously presented and perceived to originate from the same spatial location, the combined bimodal stimuli are experienced as a single perceptual event with a single duration. This makes it difficult to separately attend to the duration of the visual component. It is possible that selective attention to the visual component might overcome auditory dominance if we facilitate perceptual segregation of the visual component from the auditory component. We accomplished this in two ways.

In Experiments 2A and 2B, we asynchronously presented the visual and auditory components to perceptually segregate their onsets and offsets. In Experiment 2C, we perceptually segregated the visual and auditory components by presenting them at widely separated spatial locations. We also ensured that participants attended to the visual stimulus by using a demanding visual task (Experiments 2B and 2C).

Thus, by facilitating perceptual segregation between the auditory and visual components as well as by enforcing attention to the visual component, we rigorously tested the hypothesis that selective attention to the visual component might overcome auditory dominance in duration perception.

Experiment 2A: Selective attention to the visual component facilitated by auditory-visual asynchrony

Method

Participants: A new group of eighteen Northwestern University undergraduate students (10 female) gave informed consent to participate for partial course credit.

Apparatus and stimuli: The apparatus and stimuli were the same as those used in Experiment 1 with the following exceptions. We used the visual-alone stimulus and the bimodal stimuli with either the intensity-matched or lower-intensity sounds because we did not expect visual attention to overcome auditory dominance when the auditory component was substantially more intense than the visual component. The durations for the visual-alone

stimulus and the visual component of the bimodal stimuli were 400, 500, 600, 700, 800, 900, and 1000 ms, with 400 ms and 1000 ms serving as the short and long reference durations, respectively; these durations were each 200 ms longer than those used in Experiment 1 to permit the asynchronous auditory-visual presentation. The bimodal stimuli were presented in two types of temporal asynchrony: (1) the auditory component longer than the visual component, V+longer_A, where the auditory component started 150 ms earlier and ended 150 ms later relative to the visual component (Figure 4A), and (2) the auditory component shorter than the visual component, V+shorter_A, where the auditory component started 150 ms later and ended 150 ms earlier relative to the visual component (Figure 4B). In this way, we perceptually segregated the visual and auditory components without changing the center of the temporal location of the visual and auditory components. Previous results suggest that asynchronous presentations do not eliminate auditory influences on visual duration judgments so long as the auditory and visual components are temporally aligned at the center and the temporal offset between them is less than 500 ms (Klink, Montijn, & van Wezel, 2011; Romei, De Haas, Mok, & Driver, 2011).

Procedure: The 2AFC temporal-bisection task was the same as in Experiment 1 except that the reference durations were always visually presented. Each block consisted of 21 trials (3 stimulus types [V, V+longer_A, and V+shorter_A] $\times$ 7 durations), and the total number of blocks was 12. The intensity-matched auditory stimulus was used for half of the participants (N=9), and the lower-intensity auditory stimulus was used for the remaining participants (N=9). The participants were instructed to judge the duration of the visual component while ignoring the auditory component.

Results—If attention to the visual stimulus was 100% effective, the duration judgments for the visual component of the bimodal stimuli should be identical to those for the visual-alone stimuli. In contrast, if the auditory component still dominated duration perception despite the fact that participants attempted to judge the duration of the visual component while ignoring the auditory component, the duration judgments should be longer for the V+longer_A stimulus and shorter for the V+shorter_A stimulus relative to those for the visual-alone stimulus. This is exactly what happened when the auditory intensity was matched to the visual intensity; the PSE for the V+longer_A stimulus was significantly shorter ($t[8] = 4.695, p < 0.002$) and the PSE for the V+shorter_A stimulus was significantly longer ($t[8] > 10.030, p < 0.0001$) than the PSE for the visual-alone stimulus (Left panel in Figure 5A). Thus, when the stimulus intensities were matched between the two modalities, perceived durations of bimodal stimuli were strongly influenced by the auditory component despite the fact that the auditory-visual asynchrony facilitated perceptual segregation of the visual and auditory components and participants were instructed to attend the visual component and ignore the auditory component.

Whereas visual selective attention did not overcome auditory dominance when the visual and auditory intensities were matched, attention had some effect when the auditory intensity was substantially weaker than the visual intensity. The PSE for the V+longer_A stimulus was significantly shorter than the PSE for the visual-alone stimulus ($t[8] = 7.459, p <$

0.0001), but the PSE for the V+shorter_A stimulus was not significantly different from the PSE for the visual-alone stimulus ($t[8] = 2.016, n.s.$) (Right panel in Figure 5A).

Selective attention had little effect on JND (Figure 5B). For both the intensity-matched and lower-intensity sounds, JNDs were statistically equivalent for the visual-alone, V+longer_A and V+shorter_A stimuli ($t[8]s < 1.176, n.s.$, and $t[8]s < 1.192, n.s.$, for the intensity-matched and lower-intensity sounds, respectively).

Thus, visual selective attention can overcome auditory dominance in duration perception, but only when the auditory intensity is weaker than the visual intensity and when the auditory component is temporally enclosed within the visual component (see Figure 4B). It is reasonable to expect that the temporal enclosure further reduced the salience of the auditory component because the visual component marked both the initial onset and the final offset, thereby forward masking the onset and backward masking the offset of the weaker auditory component. Taken together, our results suggest that visual selective attention can overcome auditory dominance in bimodal duration perception but only when the auditory component is substantially weaker than the visual component.

Experiment 2B: Selective attention to the visual component facilitated by auditory-visual asynchrony and enforced by a luminance-decrement detection task

In the preceding experiment, attention was manipulated by instructing participants to judge the duration of the visual component while ignoring the auditory component, as in Godstone et al. (1959). However, simply instructing participants may not be sufficient to ensure that they actually attend to the visual stimulus. Here, we added a luminance-decrement detection task to ensure that participants attended to the visual component. We used only the lower-intensity sounds because in Experiment 2A visual attention was only effective when the visual component was more intense than the auditory component.

Method

Participants: A new group of nine Northwestern University undergraduate students (7 female) gave informed consent to participate for partial course credit.

Stimuli: The stimulus conditions were the same as in Experiment 2A with the following exceptions. Only the lower-intensity sounds were used. On a luminance-decrement trial, the blue circle became slightly darker (i.e., luminance decreased from 17.2 cd/m^2 to 16.9 cd/m^2) at a randomly chosen time (excluding the first and last 150 ms of the stimulus duration).

Procedure: We used the same procedure as in Experiment 2A except that luminance-decrement trials were randomly inserted, comprising 20% of the trials. After each stimulus presentation (visual-alone, V+longer_A, or V+shorter_A), the participant first performed the 2AFC temporal-bisection task, responding as to whether the presented stimulus was closer in duration to the short or long reference. The participant then responded as to whether the disk had darkened by pressing the “Y” or “N” key. The next trial began one second after the response.

Chi-squared analyses on individual participants indicated that all participants performed the luminance-decrement detection task significantly above chance ($\chi^2_s > 62.980$, $ps < 0.0001$), yielding a mean accuracy of 78% (about a half way between the floor [50%] and ceiling [100%] suggesting that the level of difficulty of the attention task was optimal). Because luminance decrements occurred unpredictably on randomly selected trials, we can reasonably assume that participants attended to the visual stimulus on all trials. The duration-judgment responses were analyzed only for the trials on which luminance decrements did not occur because a luminance decrement itself and/or attending to it may affect duration perception (e.g., Tse, Rivest, Intriligator, & Cavanagh, 2004).

Results—We replicated the lower-intensity-sound results from Experiment 2A (Figure 6). The PSE for the V+longer_A stimulus was significantly shorter than the PSE for the visual-alone stimulus ($t[8] = 4.585$, $p < 0.002$), whereas the PSE for the V+shorter_A stimulus was not significantly different from the PSE for the visual-alone stimulus ($t[8] = 0.717$, *n.s.*) (Figure 6A). The JND was significantly greater for the visual-alone stimulus compared with both bimodal stimuli ($t[8] > 2.569$, $p < 0.033$), but it did not differ between the two bimodal stimuli ($t[8] = 1.572$, *n.s.*).

Thus, even when the auditory component was substantially weaker than the visual component and selective attention was engaged to the visual component (verified by the luminance-decrement detection task), the auditory component still dominated duration perception except when it was temporally enclosed within the visual component.

Experiment 2C: Selective attention to the visual component facilitated by large spatial separation of auditory and visual components and enforced by a luminance-decrement detection task

In Experiments 2A and 2B we facilitated selective attention to the visual component by asynchronously presenting the visual and auditory components. To further confirm that auditory dominance in duration perception is difficult to overcome with attention, we also facilitated selective attention to the visual component by spatially separating the visual and auditory components. Auditory-visual integration is reduced when the locations of the visual and auditory stimuli are separated by more than about 20° of visual angle (for a stimulus configuration largely similar to ours; e.g., Godfroy, Roumes, & Dauchy, 2003; Lewald & Guski, 2003). Using a large (52") display monitor we presented the blue circle either near the sound source (12° visual angle from the audio speaker on the same side of the monitor) or far from the sound source (88° visual angle from the speaker on the opposite side of the monitor). A green circle was presented on the opposite side from the blue circle. Participants were instructed to always attend to and judge the duration of the blue circle (whose location was pre-cued on each trial). When the tone was played through the speaker near the blue circle, the blue circle and the tone would perceptually integrate, so that it would be difficult to attend to the blue circle and to ignore the tone. In contrast, when the tone was played through the speaker near the green circle, the tone would perceptually integrate with the green circle, helping participants to ignore the tone while attending to the blue circle.

As in Experiment 2B, we used an auditory tone that was substantially weaker than the visual stimulus to make the visual component more salient, and used a luminance-decrement detection task to enforce visual attention to the blue circle. If selective attention, aided by the considerable spatial separation between the visual target and the tone, could overcome auditory dominance in duration perception, the PSE for the bimodal stimulus with the distant tone should be close to the PSE for the visual-alone stimulus. Alternatively, if the auditory component still dominated duration perception, the PSEs for the bimodal stimuli should be equivalent to the PSE for the auditory-alone stimulus whether or not the tone was near or far from the visual target.

Method

Participants: A new group of ten Northwestern University undergraduate students (5 female) gave informed consent to participate for partial course credit.

Apparatus: Visual stimuli were displayed on a Viewsonic 52" color LCD monitor (60 Hz, 1920 by 1080 resolution) located at 60 cm in front of the participant. Auditory stimuli were presented via a pair of JBL DUET speakers (60 Hz – 20 kHz frequency response), located at the left and right sides of the monitor.

Stimuli: The target visual stimulus was a blue circle (16.5° of visual angle in diameter, 39.8 cd/m^2 , CIE[.145, .062]), presented 37.9° of visual angle to the left or right (equiprobable and intermixed across trials) of a central fixation marker (1.1° of visual angle in diameter, 8.0 cd/m^2). On a luminance-decrement trial, the blue circle became slightly darker (34.7 cd/m^2) as in Experiments 2B. A green circle (the same diameter as the blue circle, 71.0 cd/m^2 , CIE[.270, .527]) was simultaneously presented on the opposite side of the fixation marker. These stimuli were presented against a white background (565 cd/m^2)². The left speaker was placed 12.2° of visual angle peripheral to the left visual disk and the right speaker was placed 12.2° of visual angle peripheral to the right visual disk. Note that the contrast of the blue circle was greater and the loudness of the tone (50 db SPL-A) was weaker in this experiment than in any of the preceding experiments. In other words, the visual target was more salient than the tone in this experiment than in any of the preceding experiments. For bimodal stimuli, the visual target and the tone were temporally coincident as in Experiment 1. The durations of the visual-alone and bimodal stimuli were the same as the durations of the visual stimuli in Experiments 2A and 2B.

Procedure: The 2AFC temporal-bisection task was the same as in the preceding experiments, except (1) that the reference durations were always visually presented and (2) that we manipulated the relative location of the visual target and auditory tone. There were three conditions: (1) visual-alone, (2) V+near_A, in which the blue circle (visual target) and the auditory tone were presented on the same side of the monitor, and (3) V+far_A, in which the auditory tone was presented on the opposite side of the blue circle.

Participants were instructed to maintain eye fixation at the fixation marker. A video camera was placed 25 cm in front of the participant, capturing his or her left eye, to monitor central fixation. We did not detect any systematic deviations of eye fixation from the center. One

second after the appearance of the fixation marker, an arrow cue appeared above the fixation marker for 500 ms, indicating the side on which the blue circle would appear. Following a 250 ms inter-stimulus interval, the blue and green circles accompanied by a tone (except on the visual-alone trials) were presented for a given duration. Participants were instructed to judge the duration of the blue circle and to detect its luminance decrement, if any, while ignoring the green circle and the tone. As in Experiment 2B, participants first responded as to whether the duration of the blue circle was closer to the short or long reference, and then responded as to whether the blue circle had darkened. The next trial began one second after the response. As in the preceding experiment, the data were analyzed only for the 80% of the trials on which the luminance of the blue circle did not decrease.

Each block consisted of six reference trials (identical to those in Experiments 2A and 2B except that the arrow cue and the two disks were presented on each trial), followed by 21 test trials (3 stimulus types [visual-alone, V+near_A, and V+far_A] $\times$ 7 durations), and the total number of blocks was 12.

To compare the perceived visual durations in the visual-alone and bimodal conditions with the perceived auditory duration, we subsequently administered an auditory-alone condition. Each auditory-alone block consisted of six reference trials, followed by 21 test trials (7 durations $\times$ 3 repetitions), and the total number of blocks was 4. We tested the visual-alone and bimodal stimuli first so that participants always attended to and judged the visual duration of the blue circle on all visual and bimodal trials before they were asked to respond to the tone.

Chi-squared analyses on individual participants indicated that all participants performed the luminance-decrement detection task significantly above chance ($\chi^2_s > 7.371$, $ps < 0.001$), yielding a mean accuracy of 81% (again, about half way between the floor [50%] and ceiling [100%] suggesting an optimal level of difficulty of the attention task).

Results—Selective visual attention did not affect auditory dominance in duration perception. The PSEs for both types of bimodal stimuli (V+near_A and V+far_A) were significantly shorter than the PSE for the visual-alone stimulus ($t[9]s > 5.192$, $ps < 0.001$), and equivalent to the PSE for the auditory-alone stimulus ($t[9]s < 1.253$, $n.s.$) (Figure 7A). Similarly, the JNDs for both types of bimodal stimuli were significantly less than the JND for the visual-alone stimulus ($t[9]s > 2.862$, $ps < 0.02$), and equivalent to the JND for the auditory-alone stimulus ($t[9]s < 2.234$, $n.s.$) (Figure 7B).

Thus, even when the auditory stimulus was weak, perceptually bound to a proximate non-target visual stimulus, and the attended visual target was 88° of visual angle away from the sound source, the perceived duration and temporal discriminability of the attended visual target were still determined by the concurrent sound.

Experiment 3: Is it auditory dominance or dominance by the modality with superior temporal discriminability?

So far we have provided evidence suggesting that when an auditory stimulus and a visual stimulus are concurrently presented, the auditory stimulus dominates duration perception

even when the visual stimulus is selectively attended and the auditory stimulus is perceptually weaker and is ignored. It is thus possible that the auditory modality is uniquely dominant in duration perception. Alternatively, it is possible that duration signals from each modality may be weighted in proportion to their temporal discriminability (see the Introduction). When the auditory-alone stimulus was included in Experiments 1 and 2C, the JND was smaller for the auditory-alone stimulus than for the visual-alone stimulus, indicating that temporal discriminability was greater for the auditory than visual modality in the preceding experiments. Thus, our results so far are consistent with both the auditory-dominance hypothesis and the discriminability-dependent-weighting hypothesis.

In this experiment, we temporally degraded the auditory stimulus by ramping its onset and offset, that is, we gradually increased the sound intensity to reach a maximum value and then gradually decreased the intensity to silence. A prior study showed that gradual onsets degraded duration discrimination compared to abrupt onsets (Schlaugh, Ries, & DiGiovanni, 2001). Although the same study did not find any effect of gradual offsets on duration discrimination, we used auditory stimuli with gradual onsets and offsets so that they sounded temporally symmetric. In this way, we attempted to make the JND of the auditory-alone stimulus similar to that of the visual-alone stimulus. The auditory-dominance hypothesis predicts that the perceived duration of a bimodal stimulus should still be the same as the perceived duration of the auditory component even when the visual and auditory components are equivalently discriminable. In contrast, the discriminability-dependent-weighting hypothesis predicts that the perceived duration of a bimodal stimulus should then be the average of the perceived durations of the visual and auditory components.

Method

Participants: Fourteen new Northwestern University undergraduate students (7 female) gave informed consent to participate for partial course credit.

Apparatus and stimuli: The apparatus and stimuli were the same as those in Experiment 1 with the following changes. Only tones with the matched intensity were presented. Each tone started 50 ms before and ended 50 ms after the concurrent visual stimulus. The tone intensity was monotonically increased in the first half following the function, $a(t) = e^{5(t/T-1)}$, and monotonically decreased in the second half following the mirror-reversed version of the same function, where $a(t)$ is the amplitude of the tone at time t (measured from tone onset) and T is the duration of the tone (Grassi & Darwin, 2006).

Procedure: The test durations were the same as in Experiments 2A-2C. As in Experiment 2C, we first tested the visual-alone and bimodal stimuli (all intermixed), and subsequently tested the auditory-alone stimulus. In the former, each block consisted of six reference trials followed by 14 test trials (2 stimulus types [visual-alone and bimodal] $\times$ 7 durations), and the total number of blocks was 9. In the latter, each block consisted of six reference trials followed by 21 test trials (1 stimulus type [auditory-alone] $\times$ 7 durations $\times$ 3 repetitions), and the total number of blocks was 3. The reference durations were always visually presented for half of the participants, and were always auditorily presented for the remaining participants.

Results—The main effect of reference modality on the PSE was not significant ($F[1,13] = 1.658$, *n.s.*), and it did not interact with stimulus condition (visual-alone, auditory-alone, or bimodal), $F(2,24) = 0.182$, *n.s.*; we therefore pooled the data from all participants in the analysis³.

The critical manipulation for this experiment was that we temporally “smeared” the auditory stimuli so that auditory temporal discriminability would be more similar to visual temporal discriminability. Although we designed the smeared temporal profile of the auditory stimuli based on a pilot study, the JND was still significantly greater for the visual-alone stimulus than for the auditory-alone stimulus, $t(13) = 3.154$, $p < 0.008$. Nevertheless, the difference in the JND between the visual-alone and auditory-alone stimuli was much reduced in this experiment (25 ms) relative to that in Experiment 1 (90 ms) and Experiment 2C (45 ms).

To evaluate whether auditory dominance depended on the relative temporal discriminability of auditory and visual modalities, we rank-ordered the 14 participants according to their difference in JND between the visual-alone and auditory-alone stimuli, from positive (i.e., higher temporal discriminability in audition than vision) to negative (i.e., higher temporal discriminability in vision than audition), and divided them into two groups at the median (i.e., a median split). In the higher auditory-discriminability group the JND was significantly greater for the visual-alone than auditory-alone and bimodal stimuli ($t[6]s > 3.988$, $ps < 0.01$; the auditory-alone and bimodal conditions did not significantly differ, $t[6] = 1.733$, *n.s.*) (Figure 8A, left). In the lower auditory-discriminability group, the JNDs were statistically equivalent among the visual-alone, auditory-alone and the bimodal stimuli ($t[6]s < 0.769$, *n.s.*) (Figure 8A, right); we call this group, the equivalent discriminability group. Thus, the discriminability-dependent-weighting hypothesis predicts that the bimodal PSE should be equivalent to the auditory-alone PSE for the higher auditory-discriminability group, but that the bimodal PSE should be the average between the auditory-alone and visual-alone PSEs for the equivalent discriminability group. In contrast, the auditory-dominance hypothesis predicts that the bimodal PSE should be the same as the auditory-alone PSE for both groups.

It is clear from Figure 8B that the bimodal PSE was the same as the auditory-alone PSE for both groups. This observation is confirmed by a two-factor ANOVA with group and stimulus condition (visual-alone, auditory-alone, or bimodal) as the independent variables and PSE as the dependent variable, yielding a main effect of stimulus condition ($F[2,24] = 9.385$, $p < 0.001$) with the visual-alone PSE significantly longer than both the auditory-alone and bimodal PSEs ($t[13]s > 3.353$, $ps < 0.01$) and no reliable difference between the auditory-alone and bimodal PSEs ($t[13] = 0.533$, *n.s.*), and no significant interaction between group and stimulus condition ($F[2,24] = 0.181$, *n.s.*).

To further confirm that the bimodal PSE was unaffected by the relative temporal discriminability of the visual and auditory components, we evaluated the correlation between the relative discriminability of the visual component and the relative contribution of the visual component to the bimodal PSE. We quantified the relative visual discriminability as the auditory-alone JND minus the visual-alone JND, where a larger value indicates greater visual discriminability relative to auditory discriminability. To quantify the relative

contribution of the visual component to the bimodal PSE, we computed the visual-PSE-contribution index = [bimodal PSE – auditory-alone PSE]/[visual-alone PSE – auditory-alone PSE]; the index assumes a value of 1 if the bimodal PSE is the same as the visual-alone PSE (visual dominance), 0.5 if the bimodal PSE is the average of the visual-alone and auditory-alone PSEs (equal weighting of the visual and auditory signals), and 0 if the bimodal PSE is the same as the auditory-alone PSE (auditory dominance). The discriminability-dependent-weighting hypothesis predicts a positive correlation between the visual-PSE-contribution index and the relative visual discriminability, whereas the auditory dominance hypothesis predicts no correlation. We found no evidence of any such correlation, $r = -0.082$, *n.s.*

We thus found no evidence that the contributions of the auditory and visual signals in bimodal duration perception depend on the relative discriminability of the auditory and visual components. These results suggest that even when auditory discriminability is equivalent to visual discriminability, the auditory modality completely dominates duration perception for bimodal stimuli.

General Discussion

For the perception of duration up to about 10 seconds an auditory stimulus is perceived to be longer than a visual stimulus of the same physical duration (e.g., Droit-Volet, Meck, & Penney, 2007; Goldstone & Goldfarb, 1964; Grondin et al., 1998; Ortega et al., 2009; Penney, 2003), suggesting that an auditory internal clock ticks faster than a visual clock (e.g., Penney et al., 1998, 2000; Wearden et al., 1998, 2006). Which clock would determine the perceived duration of a bimodal event? Specifically, would the duration be perceived according to the auditory clock (i.e., perceived as the same duration as the auditory duration), to the visual clock (i.e., perceived as the same duration as the visual duration), or to some combination of both clocks (i.e., perceived as somewhere between the auditory and visual durations)? We focused on judgments of short durations (up to about a second), which are likely to involve perceptual mechanisms (e.g., Johnston et al., 2006; Ortega et al., 2012) rather than cognitive strategies such as counting (e.g., Grondin et al., 1999). Earlier studies suggested that the perceived duration of a bimodal stimulus depends more strongly on the auditory component (Goldstone et al., 1959; Walker & Scott, 1981), but the extent of auditory dominance has been unclear. Our results suggest that the auditory component nearly completely dominates the perceived duration of a bimodal stimulus even when (1) the auditory component is clearly less salient than the visual component, (2) the visual component is selectively attended, and (3) auditory temporal discriminability is reduced to approximately match visual temporal discriminability.

Goldstone et al. (1959) and Walker and Scott (1981) reported some effects of relative intensity. However, their results are difficult to interpret partly because the relative intensities of the auditory and visual stimuli were not psychophysically calibrated, and partly because their measurements of time perception depended on subjective standards of one second or manual reproduction of perceived durations. It is possible that in those studies different intensities of the visual and auditory components might have produced differential effects on the subjective standards and/or motor responses without (or in addition to)

affecting duration perception. Our results with a temporal-bisection task using psychophysically calibrated auditory and visual intensities suggest that the auditory modality dominates bimodal time perception even when the auditory component is clearly weaker than the visual component.

Goldstone et al. (1959) reported that selective attention had little effect on the perceived duration of a bimodal stimulus. Their null result, however, was inconclusive because attention was manipulated only by instructions, and their coincident presentations of the visual and auditory stimuli allowed the visual and auditory components to be perceived as a single integrated event with a single duration, making it difficult to selectively attend to the duration of the visual component. We facilitated selective attention to the visual component by temporally or spatially segregating the auditory and visual components of the bimodal stimulus, and by using a demanding luminance-decrement detection task (yielding about 80% accuracy) to enforce attention to the visual component. Even when the auditory and visual components were asynchronously presented, auditory dominance in bimodal time perception was generally impervious to effects of visual selective attention unless the auditory component was clearly weaker and its onset and offset were masked by the visual component. Even when the attended visual target was spatially separated from the concurrent auditory stimulus by as much as 88° of visual angle and the auditory stimulus, clearly weaker than the visual target, was perceptually grouped with a proximate non-target visual stimulus, the perceived duration of the attended visual target was still determined by the auditory stimulus. Our results thus suggest that under typical circumstances⁴ visual selective attention cannot overcome auditory dominance in bimodal duration perception.

When an inter-sensory conflict arises, the modality that provides greater discriminability for a given perceptual task tends to dominate perception (e.g., Welch & Warren, 1980; Recanzone, 1998; Welch, 1999; Kitagawa & Ichihara, 2002). For example, the visual modality typically yields greater spatial discriminability than the auditory modality (Julesz & Hirsh, 1972; Bertelson et al., 2000), so that the visual modality tends to dominate over the auditory modality when there is spatial conflict. For instance, when visual and auditory signals originate from different locations, the apparent source of the sound tends to be captured by the location of the visual stimulus (e.g., Howard & Templeton, 1966). In contrast, the auditory modality typically yields greater temporal discriminability than the visual modality (Grondin, 1993; Grondin, Roussel, Gamache, Roy, & Ouellet, 2005; Hirsh & Sherrick, 1961; Shipley, 1964), so that the auditory modality tends to dominate over the visual modality when there is a temporal conflict. For instance, when two rapid beeps are simultaneously presented with a flash of light, the single flash often appears to be two flashes (Shams, Kamitani, & Shimojo, 2000). In addition, when visual and auditory signals are slightly shifted in time, the perceived temporal location of the visual stimulus is shifted toward the auditory stimulus (e.g., Aschersleben & Bertelson, 2003; Morein-Zamir, Soto-Franco, & Kingstone, 2003; Burr, Banks, & Morrone, 2009).

For the perception of spatial location, the idea that the modality with higher-quality signals dominates bimodal perception has been directly confirmed by a demonstration that visual and auditory spatial signals are optimally combined by weighting them according to the spatial discriminability afforded by each modality (Alais & Burr, 2004). For perception of

temporal location, however, there is evidence suggesting that auditory signals (relative to visual signals) are weighted more strongly than predicted by optimum signal combination (Burr et al., 2009). This may suggest that the auditory modality generally provides stronger input to the processing of temporal location.

For duration perception, our results suggest that the auditory modality dominates over the visual modality irrespective of their relative temporal discriminability. When we degraded auditory stimuli so that temporal discriminability was on average equalized for the auditory and visual modalities, the auditory modality still completely dominated duration perception in the sense that the perceived duration of the bimodal stimulus was equivalent to the perceived duration of the auditory stimulus. When we analyzed the results based on individual differences in auditory temporal discriminability (relative to visual temporal discriminability), we confirmed that there was virtually no correlation between the degree of auditory dominance in bimodal duration perception and the degree of auditory superiority in temporal discriminability.

When visual and auditory stimuli are presented together, specific brain structures that mediate time perception in each modality may be activated, potentially including MT/V5 for the visual stimuli (Buetti, Bahrami, & Walsh, 2008; Buetti, Walsh, Frith, & Rees, 2008) and the right superior temporal gyrus for the auditory stimuli (Buetti, van Dongen, & Walsh, 2008). The fact that auditory signals dominate over visual signals even when the auditory signals are perceptually weaker, the visual signals are selectively attended, and auditory temporal discriminability is equivalent to visual temporal discriminability, may suggest that visual duration mechanisms are automatically entrained by auditory duration mechanisms. This interpretation appears to be consistent with a finding that disruption of the auditory cortex with TMS (transcranial magnetic stimulation) impairs time estimation for both auditory and visual stimuli (Kanai, Lloyd, Buetti, & Walsh, 2011).

Note that we examined a simple form of time perception where participants estimated the duration of a single auditory-visual event. There is some evidence suggesting that auditory dominance over vision in bimodal time perception may be diminished or reversed in a more complex situation where multiple bimodal stimuli are sequentially presented and the perceived duration of each stimulus depends critically on the dynamics of changes in auditory and visual features (e.g., van Wassenhove, Bunomano, Shimojo, & Shams, 2008).

In summary, we have provided evidence suggesting that auditory processing dominates over visual processing in the perception of simple durations up to about a second even when the visual component is clearly more salient, the visual component is selectively attended, and visual temporal discriminability is no worse than auditory temporal discriminability.

Acknowledgments

This research was supported by National Institutes of Health grants R01 EY018197 and EY021184, National Science Foundation grant BCS0643191, and CONACyT grant EP 0094258.

References

- Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*. 2004; 14:257–262. [PubMed: 14761661]
- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ. Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*. 2005; 166:559–571. [PubMed: 16028028]
- Andersen RA, Snyder LH, Bradley DC, Xing J. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*. 1997; 20:303–330.
- Aschersleben G, Bertelson P. Temporal ventriloquism: crossmodal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology*. 2003; 50:157–163. [PubMed: 14511843]
- Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*. 2004; 41:809–823. [PubMed: 15003179]
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*. 2004; 7:1190–1192.
- Bertelson P, Vroomen J, de Gelder B, Driver J. The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*. 2000; 62:321–332. [PubMed: 10723211]
- Bolognini N, Frassinetti F, Serino A, Làdavas E. “Acoustical vision” of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental Brain Research*. 2005; 160:273–282. [PubMed: 15551091]
- Brainard DH. The Psychophysics Toolbox. *Spatial Vision*. 1997; 10:433–436. [PubMed: 9176952]
- Bueti D, Bahrami B, Walsh V. Sensory and association cortex in time perception. *Journal of Cognitive Neuroscience*. 2008; 20:1054–1062. [PubMed: 18211231]
- Bueti D, van Dongen EV, Walsh V. The role of superior temporal cortex in auditory timing. *PLoS ONE*. 2008; 3:e2481. [PubMed: 18575615]
- Bueti D, Walsh V, Frith C, Rees G. Different brain circuits underlie motor and perceptual representations of temporal intervals. *Journal of Cognitive Neuroscience*. 2008; 20:204–214. [PubMed: 18275329]
- Burr D, Banks MS, Morrone MC. Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*. 2009; 198:49–57. [PubMed: 19597804]
- Driver J, Spence C. Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*. 1998; 2:254–262. [PubMed: 21244924]
- Droit-Volet S, Meck WH, Penney TB. Sensory modality and time perception in children and adults. *Behavioural Processes*. 2007; 74:244–250. [PubMed: 17084041]
- Frassinetti F, Bolognini N, Làdavas E. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*. 2002; 147:332–343. [PubMed: 12428141]
- Godfroy M, Roumes C, Dauchy P. Spatial variations of visual-auditory fusion areas. *Perception*. 2003; 32:1233–1245. [PubMed: 14700258]
- Goldstone S, Boorman WK, Lhamon WT. Intersensory comparisons of temporal judgments. *Journal of Experimental Psychology*. 1959; 57:243–248. [PubMed: 13641599]
- Goldstone S, Goldfarb JL. Direct comparison of auditory and visual durations. *Journal of Experimental Psychology*. 1964; 67:483–485. [PubMed: 14171241]
- Grassi M, Casco C. Audiovisual bounce-inducing effect: Attention alone does not explain why the discs are bouncing. *Journal of Experimental Psychology: Human Perception and Performance*. 2009; 35:235–243. [PubMed: 19170485]
- Grassi M, Darwin CJ. The subjective duration of ramped and damped sounds. *Perception & Psychophysics*. 2006; 68:1382–1392. [PubMed: 17378424]
- Grondin S. Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perceptual Psychophysics*. 1993; 54:383–393.

- Grondin S, Meilleur-Wells G, Lachance R. When to start explicit counting in a time-intervals discrimination task: a critical point in the timing process of humans. *Journal of Experimental Psychology: Human Perception and Performance*. 1999; 25:993–1004.
- Grondin S, Meilleur-Wells G, Ouellette C, Macar F. Sensory effects on judgments of short-time intervals. *Psychological Research*. 1998; 61:261–268. [PubMed: 9870293]
- Grondin S, Roussel M-E, Gamache P-L, Roy M, Ouellet B. The structure of sensory events and the accuracy of time judgments. *Perception*. 2005; 34:145–158.
- Guttman SE, Gilroy LA, Blake R. Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological Science*. 2005; 16:228–235. [PubMed: 15733204]
- Hirsh IJ, Sherrick CE Jr. Perceived order in different sense modalities. *Journal of Experimental Psychology*. 1961; 62:423–432. [PubMed: 13907740]
- Howard, IP.; Templeton, WB. *Human Spatial Orientation*. Wiley; New York, NY: 1966.
- Iordanescu L, Guzman-Martinez E, Grabowecky M, Suzuki S. Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*. 2008; 15:548–554. [PubMed: 18567253]
- Johnston A, Arnold DH, Nishida S. Spatially localized distortions of event time. *Current Biology*. 2006; 16:472–479. [PubMed: 16527741]
- Julesz, B.; Hirsh, IJ. Visual and auditory perception—An essay of comparison.. In: David, EE., Jr.; Denes, PB., editors. *Human communication: A unified view*. McGraw-Hill; New York: 1972. p. 283-340.
- Kanai R, Lloyd H, Buetti D, Walsh V. Modality-independent role of the primary auditory cortex in time estimation. *Experimental Brain Research*. 2011; 209:465–471. [PubMed: 21318347]
- Kitagawa N, Ichihara S. Hearing visual motion in depth. *Nature*. 2002; 416:172–174. [PubMed: 11894093]
- Kleiner M, Brainard D, Pelli D. What's new in Psychtoolbox-3? *Perception*. 2007; 36 ECVF Abstract Supplement.
- Klink PC, Montijn RJ, van Wezel RJA. Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention, Perception & Psychophysics*. 2011; 73:219–236.
- Lewald L, Guski R. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*. 2003; 16:468–478. [PubMed: 12706226]
- Lewis JW, Beauchamp MS, DeYoe EA. A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*. 2000; 10:873–888. [PubMed: 10982748]
- Mauk MD, Buonomano DV. The neural basis of temporal processing. *Annual Review of Neuroscience*. 2004; 27:304–340.
- McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA. Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience*. 2005; 8:1197–1202.
- Meyer GF, Wuerger SM. Cross-modal integration of auditory and visual motion signals. *Neuroreport*. 2001; 12:2557–2560. [PubMed: 11496148]
- Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ. The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex*. 2005; 15:545–551. [PubMed: 15342438]
- Morein-Zamir S, Soto-Franco S, Kingstone A. Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*. 2003; 17:154–163. [PubMed: 12763201]
- Ortega L, Lopez F, Church RM. Modality and intermittency effects on time estimation. *Behavioural Processes*. 2009; 81:270–273. [PubMed: 19429221]
- Ortega L, Guzman-Martinez E, Grabowecky M, Suzuki S. Flicker adaptation of low-level cortical visual neurons contributes to temporal dilation. *Journal of Experimental Psychology: Human Perception and Performance*. 2012; 38:1380–1389. [PubMed: 22866761]
- Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*. 1997; 10:437–442. [PubMed: 9176953]
- Penney, TB. Modality differences in interval timing: Attention, clock speed, and memory.. In: Meck, WH., editor. *Functional and neural mechanisms of interval timing*. CRC Press; Boca Raton, FL: 2003. p. 209-234.

- Penney T, Gibbon J, Meck W. Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26:1770–1787. [PubMed: 11129373]
- Penney, T.; Allan, L.; Meck, W.; Gibbon, J. Memory mixing in duration bisection.. In: Rosenbaum, DA.; Collyer, CE., editors. *Timing of behavior: Neural, computational, and psychophysical perspectives*. MIT Press; Cambridge, MA: 1998. p. 165-193.
- Recanzone GH. Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences*. 1998; 95:869–875.
- Ries DT, Schlauch RS, DiGiovanni JJ. The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *Journal of the Acoustical Society of America*. 2008; 124:3772–3783. [PubMed: 19206803]
- Romei V, De Hass B, Mok RM, Driver J. Auditory stimulus timing influences perceived duration of co-occurring visual stimuli. *Frontiers in Psychology*. 2011; 2(215):1–8. [PubMed: 21713130]
- Schlauch RS, Ries DT, DiGiovanni JJ. Duration discrimination and subjective duration for ramped and damped sounds. *Journal of the Acoustical Society of America*. 2001; 109:2880–2887. [PubMed: 11425130]
- Sekuler R, Sekuler AB, Lau R. Sound alters visual motion perception. *Nature*. 1997; 385:308. [PubMed: 9002513]
- Shams L, Kamitani Y, Shimojo S. Illusions. What you see is what you hear. *Nature*. 2000; 408:788. [PubMed: 11130706]
- Shimojo S, Shams L. Sensory modalities are not separate modalities: Plasticity and interactions. *Current Opinion Neurobiology*. 2001; 11:505–509.
- Shipley T. Auditory flutter-driving of visual flicker. *Science*. 1964; 145:1328–1330. [PubMed: 14173429]
- Smith E, Grabowecky M, Suzuki S. Auditory–visual crossmodal integration in perception of face gender. *Current Biology*. 2007; 17:1680–1685. [PubMed: 17825561]
- Stein BE. Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research*. 1997; 123:124–135. [PubMed: 9835401]
- Stein BE, Meredith MA, Huneycutt WS, McDade L. Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*. 1989; 1:12–24. [PubMed: 23968407]
- Tse PU, Rivest J, Intriligator J, Cavanagh P. Attention and the subjective expansion of time. *Perception & Psychophysics*. 2004; 66:1171–1189. [PubMed: 15751474]
- van Wassenhove V, Buonomano DV, Shimojo S, Shams L. Distortions of subjective time perception within and across senses. *PLoS ONE*. 2008; 3(1):e1437. [PubMed: 18197248]
- von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*. 2005; 17:367–376. [PubMed: 15813998]
- Vroomen J, Bertelson P, de Gelder B. The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*. 2001; 63:651–659. [PubMed: 11436735]
- Walker JT, Scott KJ. Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology Human Perception and Performance*. 1981; 7:1327–1339. [PubMed: 6458656]
- Wearden J, Todd N, Jones L. When do auditory–visual differences in duration judgments occur? *Quarterly Journal of Experimental Psychology*. 2006; 59:1709–1724.
- Wearden J, Edwards H, Fakhri M, Percival A. Why sounds are judged longer than lights: Application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology B*. 1998; 51:97–120.
- Welch, RB. Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perceptions.. In: Aschersleben, Bachmann G.; Musseler, TJ., editors. *Cognitive Contributions to the Perception of Spatial and Temporal Events*. Elsevier; Amsterdam: 1999. p. 371-387.
- Welch RB, Warren DH. Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*. 1980; 88:638–667. [PubMed: 7003641]

Figure 1.

Results of auditory-visual intensity calibration. Proportion of “Auditory more intense than visual” responses as a function of sound intensity. The data from the four participants are shown separately. For each participant, the continuous curve represents the logistic-function fit; the vertical line that intersects the curve at the 50% point represents the point of subjective equality (PSE), that is, the auditory intensity subjectively matched to the visual intensity. Note that the psychometric functions and the PSEs are nearly identical for the four participants. In the experiments, we used 59 dB as the sound intensity subjectively matched to the visual stimulus intensity, 50 dB or 54 dB as the sound intensity clearly lower than the visual stimulus intensity, and 64 dB as the sound intensity stimulus intensity.

Figure 2.

Illustration of a trial sequence for the 2-alternative-forced-choice (2AFC) temporal-bisection task. (A) Reference trial. Either an auditory or visual stimulus was presented for the short or long reference duration. (B) Test trial. Either an auditory, visual, or auditory-visual bimodal stimulus was presented for a variable duration between the short and long reference durations (inclusive), and participants judged the test duration as being closer to the short or long reference duration (see the text for details).

Figure 3.

Results of Experiment 1. (A) Mean values of the point of subjective equality, PSE, when a visual stimulus was presented alone (V), a matched-intensity auditory stimulus was presented alone (matched_A), a visual stimulus was presented with a clearly lower clearly lower-intensity auditory stimulus (V+lower_A), a visual stimulus was presented with a matched -intensity auditory stimulus (V+matched_A), and when a visual stimulus was presented with a clearly higher-intensity auditory stimulus (V+higher_A). (B) Mean values of the just noticeable difference, JND (defined as the deviation from the PSE that yields 75% correct “long” or “short” classification) in the corresponding stimulus conditions. In both A and B, the left panel shows the results when the short and long reference durations were visually presented and the right panel shows the results when the reference durations were auditorily presented. The error bars represent ± 1 SEM adjusted for within-subjects comparisons.

Figure 4.

Illustration of the configuration of the bimodal stimuli used in Experiments 2A and 2B. (A) The auditory component started later and ended earlier relative to the visual component — V+shorter_A. (B) The auditory component started earlier and ended later relative to the visual component — V+longer_A. The duration of the visual component was the same in both cases.

Figure 5.

Results of Experiment 2A. Participants were instructed to attend to and judge the duration of the visual component duration of the visual component, when the visual and auditory components were asynchronously presented. (A) Mean values of the point of subjective equality, PSE, when a visual stimulus was presented alone (V), a visual stimulus was presented with a longer auditory stimulus (V+longer_A), and when a visual stimulus was presented with a shorter auditory stimulus (V+shorter_A). (B) Mean values of the just noticeable difference, JND (defined as the deviation from the PSE that yields 75% correct “long” or “short” classification) in the corresponding stimulus conditions. The left column shows the PSE and JND when the auditory intensity was subjectively matched to the visual intensity, and the right column shows the PSE and JND

when the auditory intensity was substantially lower than the visual intensity. The error bars represent ± 1 *SEM* adjusted for within-subjects comparisons. comparisons

Figure 6.

Results from Experiment 2B. The participant's attention was engaged to the visual component using a demanding luminance-decrement detection task. (A) Mean component using a demanding luminance values of the point of subjective equality, PSE, when PSE a visual stimulus was presented alone (V), a visual stimulus was presented with a longer auditory stimulus (V+longer_A), and when a visual stimulus was presented with a shorter auditory stimulus (V+shorter_A). (B) Mean values of the just noticeable difference, JND (defined as the deviation from the PSE that yields 75% correct “long” or “short” classification) in the corresponding stimulus conditions classification . The auditory

intensity was clearly lower than the visual intensity in all cases. than th The error bars represent ± 1 *SEM* adjusted for within- subjects comparisons.

Figure 7.

Results of Experiment 2C. The participant's attention was engaged to the visual component using a demanding luminance-decrement detection task, when the component using a demanding luminance visual target was either spatially near or far from the auditory stimulus. (A) Mean values of the point of subjective equality, PSE, when visual stimuli were presented alone (V), an auditory stimulus was presented alone (A), the visual target and an auditory stimulus were presented on the same side of the screen (V+near_A), and when the visual target and an auditory stimulus were presented on opposite sides of the screen (its sides of the screen (V+far_A). (B) Mean values of the just noticeable difference, JND (defined as the deviation from the PSE that yields 75% correct “long” or “short” classification) in the corresponding stimulus conditions. in the corresponding stimulus conditions The auditory intensity was clearly lower than the visual intensity in all cases (*

the intensity difference was larger than in any other experiments in this study) larger than in any other experiments in this study). The error bars represent ± 1 *SEM* adjusted for within-subjects comparisons. comparisons

Figure 8.

Results of Experiment 3. Auditory discriminability was reduced by temporally “smearing” the auditory stimuli. The participants were median split into two groups according to their values of visual JND minus auditory JND. The “Higher auditory- discriminability group” (left column) had longer visual JNDs compared to their auditory JNDs as a group. The “Equivalent discriminability group” (right column) had equivalent visual and auditory JNDs as a group. (A) Mean values of just noticeable difference, JND (defined as the deviation from the PSE that yields 75% correct “long” or “short” classification) when visual stimuli were presented alone (V), an auditory stimulus was presented alone (A), and when they were presented together (V+A). (B) Mean values of the point of subjective equality, PSE, in the corresponding stimulus conditions. Note that the PSE for the bimodal stimulus was equivalent to the PSE for the auditory stimulus for both groups. The auditory intensity was

subjectively matched to the visual intensity in all cases. The error bars represent ± 1 *SEM* adjusted for within- subjects comparison.