Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency

Massimiliano Di Luca   Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Tonja-Katrin Machulla   Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Marc O. Ernst   Max Planck Institute for Biological Cybernetics, Tübingen, Germany

After exposure to asynchronous sound and light stimuli, perceived audio-visual synchrony changes to compensate for the asynchrony. Here we investigate to what extent this audio-visual recalibration effect transfers to visual-tactile and audio-tactile simultaneity perception in order to infer the mechanisms responsible for temporal recalibration. Results indicate that audio-visual recalibration of simultaneity can transfer to audio-tactile and visual-tactile stimuli depending on the way in which the multisensory stimuli are presented. With presentation of co-located multisensory stimuli, we found a change in the perceptual latency of the visual stimuli. Presenting auditory stimuli through headphones, on the other hand, induced a change in the perceptual latency of the auditory stimuli. We argue that the difference in transfer depends on the relative trust in the auditory and visual estimates. Interestingly, these findings were confirmed by showing that audio-visual recalibration influences simple reaction time to visual and auditory stimuli. Presenting co-located stimuli during asynchronous exposure induced a change in reaction time to visual stimuli, while with headphones the change in reaction time occurred for the auditory stimuli. These results indicate that the perceptual latency is altered with repeated exposure to asynchronous audio-visual stimuli in order to compensate (at least in part) for the presented asynchrony.

Keywords: subjective simultaneity, recalibration, time perception, lag adaptation, multisensory, audio-visual-tactile, Bayesian modeling

Citation: Di Luca, M., Machulla, T.-K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision, 9*(12):7, 1–16, http://journalofvision.org/9/12/7/, doi:10.1167/9.12.7.

Introduction

Most events in our surroundings can be perceived through multiple sensory modalities. An approaching train, for instance, can be seen, heard, and possibly even felt through vibrations of the ground. These different sensory signals have to be combined into one coherent multimodal percept of the environment. An important cue that is thought to aid the correct combination of information across separate sensory channels is perceived synchrony (Spence & Squire, 2003). In particular, it has been shown that multimodal interactions are strongest when stimuli are presented simultaneously or close in time (within a temporal window of about 100 ms) and get weaker as the temporal discrepancy between stimuli increases (Bresciani, Dammeier, & Ernst, 2006; Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Shams, Kamitani, & Shimojo, 2002). As is most evident in the case of audio-visual stimuli, perceived synchrony and

physical synchrony differ considerably: light and sound stimuli not only differ in their physical propagation velocity, but they also have different transduction and processing times (see, i.e., Allison, Matsumiya, Goff, & Goff, 1977; King, 2005; Spence & Squire, 2003). This difference is captured by measuring the asynchrony necessary to perceive two stimuli to be synchronous, defined as the Point of Subjective Simultaneity (PSS). It has been shown that the PSS for multimodal stimuli can differ significantly depending on the conditions of stimulation (i.e., Zampini, Shore, & Spence, 2003) and it is also affected by adaptation. Evidence for such a flexibility has recently been provided by a host of studies reporting that the perception of temporal order and simultaneity can be recalibrated for audio-visual (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Hanson, Heron, & Whitaker, 2008; Harrar & Harris, 2005, 2008; Heron, Whitaker, McGraw, & Horoshenkov, 2007; Vatakis, Navarra, Soto-Faraco, & Spence, 2007; Vroomen, Keetels, de Gelder, & Bertelson, 2004), audio-tactile (Hanson et al., 2008), and

visual-tactile (Hanson et al., 2008; Keetels & Vroomen, 2008; Takahashi, Saiki, & Watanabe, 2008) stimuli. The basic paradigm in these studies is to measure observers' perception of cross-modal simultaneity before and after they have been exposed for several minutes to a constant temporal discrepancy between the stimuli in two modalities. The repeated presentation of the same stimulus pair might increase the tendency of the participants to perceive the visual and the auditory signals to be related. In other words, the repeated exposure increases the probability that the two signals were in fact produced by the same external event ("unity assumption," Welch & Warren, 1980) and were therefore produced in synchrony. In this case, the asynchrony between the signals is likely due to some temporal bias (i.e., the difference in propagation velocity) and its effect should be reduced to a minimum. Consistently, results indicate that after exposure to asynchronous stimuli perception of simultaneity changes such that the asynchrony is perceived to be less. In other words, observers tend to perceptually realign the asynchronous stimuli during exposure. It is not clear, however, how such a recalibration is achieved.

Recalibration mechanisms

Two distinct but not mutually exclusive mechanisms for multisensory recalibration are conceivable: (a) recalibration directly affects the processing of the *individual signals* composing the multisensory stimulus pair ("adjustment"), or (b) recalibration is specific to the *multisensory pair* of signals and only affects the mapping between the two ("remapping"). To illustrate these mechanisms it is helpful to consider how temporal judgments in general may be performed (Figure 1, left panel). One of the simplest models that describe the temporal processing of a stimulus is the independent channels model by Sternberg and Knoll (1973), according to which a stimulus can be detected only after a certain time has elapsed from the presentation. This time may be different for every stimulus, so we will refer to this time as the "perceptual latency" of that particular stimulus. The perceived order of two stimuli is determined by evaluating the timing at which stimuli could be detected through a "comparator." With this mechanism in mind, temporal recalibration based on the "adjustment" of perceptual latency would be an increase of the perceptual latency of the leading stimulus in the pair, or a decrease of the perceptual latency of the second stimulus in the pair, or a compromise of both (Figure 1, right panel). For the "remapping" mechanism, instead, the processing of the signals itself would not be affected, but there would be a change in the relationship between the signals (change in comparison operator, Figure 1, middle panel).

The two mechanisms make distinct predictions with respect to the transfer of recalibration to other stimuli pairs. If recalibration is achieved by remapping the stimuli in one modality pair (e.g., audio-visual), there should be
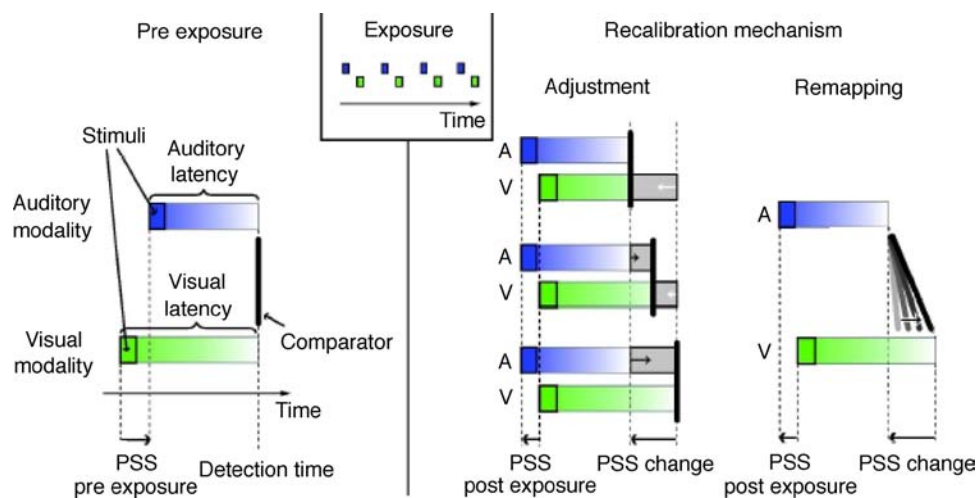

Figure 1. Mechanisms for the recalibration of simultaneity perception of an audio-visual stimulus pair before and after exposure to an audio leading stimulus (adapted from Sternberg & Knoll, 1973). In the experiments we used both exposure with audio leading stimuli and with visual leading stimuli to compare the results. (Left) Before exposure, audio-visual stimulus pairs are perceived to be synchronous when the asynchrony matches the difference in perceptual latency (a comparator detects the synchrony between the two signals). The asynchrony necessary to perceive the auditory and visual stimuli to be synchronous is the Point of Subjective Simultaneity (PSS). (Middle) After exposure to repeated asynchronous audio leading stimuli, perceived synchrony changes. For the hypothesized recalibration mechanism based on "adjustment" of perceptual latency, different combinations of adjustments can produce the same PSS change: 100% visual adjustment, 50% visual and 50% audio adjustment, 100% audio adjustment. (Right) After exposure to repeated asynchronous audio leading stimuli, the hypothesized recalibration mechanism based on "remapping" would not change the sensory latency but only the comparator.

no transfer of the recalibration effect when tested with another modality pair (e.g., visual-tactile, audio-tactile) because the remapping is specific to the recalibrated modality pair (Figure 1, middle panel). Specifically, if recalibration of audio-visual (henceforth AV) simultaneity perception is due to remapping, perceived simultaneity of the audio or visual stimulus combined with a tactile stimulus (AT or VT) should not be affected by the exposure to asynchronous AV stimuli. On the other hand, if recalibration is achieved by adjusting one of the perceptual latencies or both (Figure 1, right panel), transfer to another modality pair should be observed whenever the signal of the adjusted modality is involved (Figure 2). That is, if the mechanism involved in the recalibration of simultaneity is based on the adjustment of perceptual latency, then AV recalibration of simultaneity will transfer to the audio-tactile (AT), visual-tactile (VT), or both stimulus pairs. The underlying rationale for this prediction is that the perceptual latency determines the perceptual availability of a stimulus, therefore influencing all temporal relationships between all the perceived stimuli. This influence should be present even for stimuli that are different than the ones composing the repeated pair.

Ideal observer based on trust

If there is an adjustment of the latency for one or both stimuli of the multisensory pair to recalibrate perceived simultaneity, the question naturally arises whether there is an optimal way for the perceptual system to decide which of the two estimates to adjust and by how much: the visual (Figure 2, top), the auditory (Figure 2, bottom), or both. If the goal of recalibration would be merely to dissolve the intersensory conflict, there would be no preference about which estimate to adjust, as any combination of latency adjustment that counteracts the conflict is equally effective. However, if the goal is to successfully interact with the environment, then the objective of the perceptual system and thus also of the ideal observer must be to establish accurate sensory estimates. In the case of temporal estimations, accuracy means that the processing latency of the individual signals is known as accurately as possible. Consequently, the signal that is less accurate should be recalibrated more. As the sensory estimates do not provide direct information about their accuracy (it can only be estimated using feedback, for example through interactions) we introduce the idea of a "trusted sensory estimate" (see also Backus & Haijiang, 2007). If a sensory estimate had provided accurate information with high probability in the past (i.e., unbiased), it is likely to continue to provide unbiased information and thus can be "trusted" also in simultaneity judgments. According to this definition, the "trust" in a sensory estimate is inversely related to the probability of the estimate to be biased. Trust can only be expressed as a prior belief, as there is no indication in the sensory estimate that could provide direct information about whether this estimate is momentarily biased. Only feedback through past inter-actions can be used to build up or update such a prior belief. For example, feedback from catching a moving object can provide information on the temporal accuracy (bias) of the perceptual estimates involved and thus update the trust on the signals. For the ideal observer, the relative trust in the sensory estimate should determine the extent to which it can be taken as a standard to calibrate the other estimates to achieve accurate perceived simultaneity. Put another way, the amount by which each estimate should be calibrated is determined by a weighting scheme where the weights are inversely proportional to the relative probabilities of the prior belief that the individual estimates have been biased from the calibrated state.[1] Several factors may contribute to this prior belief, such as the accuracy of the estimates (but not their reliabilities; cf. Discussion section), or their causal relationships. For example, if one estimate can be considered the odd one out because it is unnatural and it deviates from all the
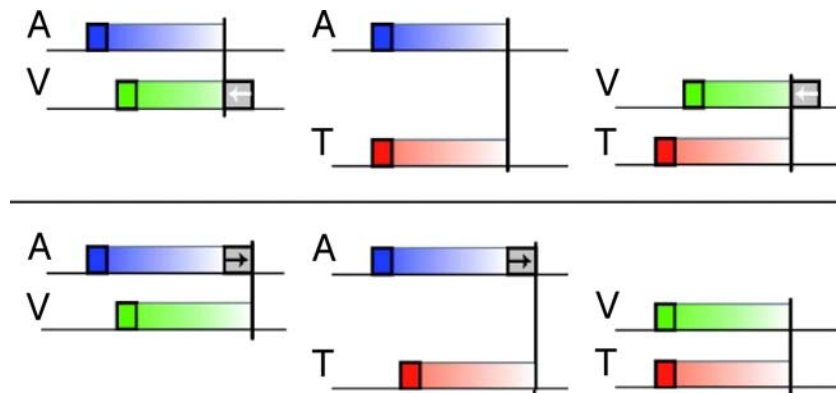

Figure 2. Transfer of perceived simultaneity recalibration due to audio leading AV exposure to stimulus pairs containing tactile signals: (top) adjustment of the visual perceptual latency and (bottom) adjustment of the auditory perceptual latency. For simplicity, perceptual latency pre-exposure is depicted to be equal for the three stimuli (i.e., PSS pre-exposure is 0 for all pairs).

other sensory estimates (e.g., it may be perceived at an odd location compared to all other estimates), then it may be causally deduced that this estimate has a greater probability to be biased.

Taking the concept of trust to the situation investigated here, we can make the following predictions. If during AV exposure the perceptual system trusts the auditory estimate, the latency of visual stimulus will be adjusted (Figure 2, top). In this case we should find transfer of AV recalibration on VT simultaneity perception. AT stimuli will not be affected because the latency of the auditory stimulus will remain the same. On the other hand, if during AV exposure the perceptual system trusts the visual estimate, then the latency of the auditory stimulus will change (Figure 2, bottom) and we should find transfer of AV recalibration on AT but not VT simultaneity perception. Finally, if there is roughly equal trust in the auditory and the visual estimates during AV recalibration, both will be adjusted by the amount reflecting the relative trust.

With the notable exception of Harrar and Harris (2005, 2008), most previous experiments have addressed the problem of simultaneity recalibration by considering only a single stimulus pair at a time (Fujisaki et al., 2004; Hanson et al., 2008; Navarra, Soto-Faraco, & Spence, 2007; Navarra et al., 2005; Vroomen et al., 2004). These studies exposed observers to an asynchronous pair of stimuli and then tested the change in the perception of temporal order for the same pair. In the first experiment we investigated whether the recalibration effect after asynchronous audio leading and visual leading AV stimuli transfers to VT and AT stimuli.

Experiment 1

To test for transfer of temporal recalibration, we exposed observers to repeated asynchronous audio and visual stimuli with either the audio or the visual stimulus leading. After exposure, we collected temporal order judgments for AV, AT, or VT stimuli and fitted the resulting psychometric functions with cumulative Gaussians. We then determined the difference in PSS after visual leading and audio leading AV exposure conditions (ΔPSS).

Methods

Participants

Seven students of the Eberhard-Karls University of Tübingen participated. They provided written consent and were paid 8 €/h. All participants were experienced psychophysical observers who were naive to the purpose of the experiment. They all reported normal hearing, no somatosensory disorders, and normal or corrected-to-normal vision.

Apparatus

Stimuli were generated using a custom-built device that generated co-located sound, vibration, and light events with high temporal accuracy (Figure 3). Subjects sat approximately 50 cm from the device in a dark, sound-attenuated room and were instructed to place their left index finger onto the contact surface of the device and to maintain fixation on this location throughout the entire experiment. Two vertically aligned speakers with a center-to-center distance of 7.5 cm and a 5-cm radius were concealed in the device and produced the auditory stimuli. A vibration device (electro-magnetic shaker, Monacor Bass Rocker BR25) was situated between the speakers. It was mounted on a damping mass, and thus produced tactile stimulation without audible noise. An LED display was mounted on top of the vibration device, serving as a vibrating surface as well as a light source (7 × 5 red LEDs, 1.6 × 1.3 cm). A multi-channel sound card (M-audio 1010LT) and three identical amplifiers were used to generate the stimuli. The accuracy of the physical stimuli
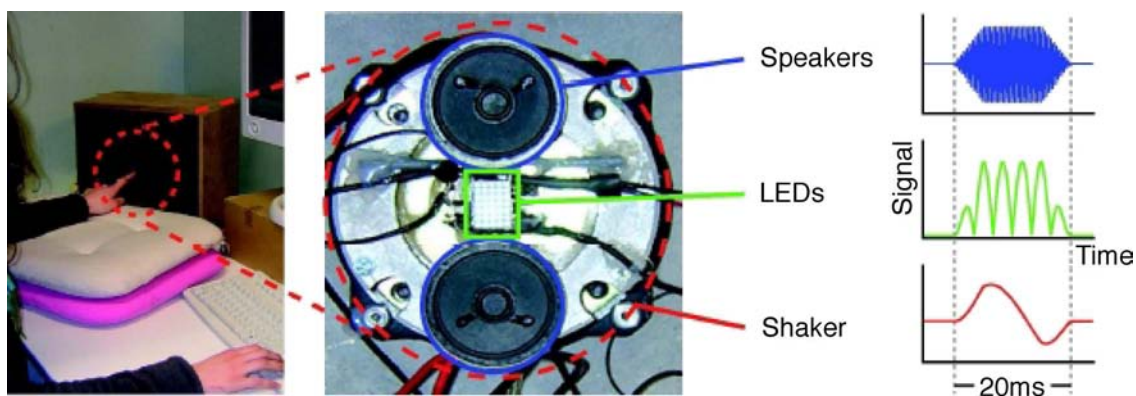

Figure 3. Device employed in the presentation of the stimuli, arrangement of the three type of sensors, and signals produced by the sound card for the three channels.

generated by the device was verified prior to the experiment using an oscilloscope: a microphone captured the signal produced by the speakers and the shaker. This way it could be verified that the auditory and the tactile signals are temporally aligned. Additionally, we used photodiodes to record the light produced by the LED array and the movement of the shaker. When in motion, the shaker intercepted a beam of light cast on the photodiode, which allowed us to record its temporal characteristics. In this way, it could be verified that the visual and the tactile signals are temporally aligned, and thus that the entire stimulation device is temporally calibrated.

Stimuli

The impulse signals used were sinusoids with duration of 20 or 40 ms (see Figure 3) with linearly ramped onset and offset (5 ms). Sounds were 2000-Hz sinusoids where the peak intensity was 61 dB SPL. Lights were 150-Hz sinusoids where negative values were rectified to also cause the LEDs to go on. The 150-Hz visual stimuli were perceived to be continuous flashes with an average luminance of 41 cd/m^2. Tactile stimuli were 40-Hz sinusoids with an intensity that created the sensation of a light "tap" on the finger. The 40-Hz tactile stimuli were not audible. We used the same visual, auditory, and tactile signals throughout the experiment.

Procedure

The experiment consisted of 16 blocks in each of which an "exposure phase" to asynchronous audio-visual stimuli (either audio leading or visual leading) was followed by a "test phase" as described below (see Figure 4).

(1) Exposure phase. During each exposure phase, which lasted 3 minutes, participants were presented with a series of AV stimulus pairs. There was an asynchrony of 150 ms between the visual and auditory stimuli with either audio preceding visual (audio leading) or vice versa (visual leading). Thus there were two types of blocks presented randomized in balanced order. In common for both types of blocks, for the first 30 seconds of the exposure phase the asynchrony between the auditory and the visual stimuli of a pair continuously increased from 0 to 150 ms (with either audio leading or visual leading). The durations of the stimuli within a pair were either 20 ms or 40 ms and there was a random pause (between 250 and 400 ms offset to onset) between successive stimulus pairs. This variation was added to the exposure stimuli in order to render the cross-correlation between the visual and auditory series of events unique. In order to ensure that participants attended to both the auditory and the visual stimuli during the exposure phase, they were asked to count the number of oddball stimulus pairs (between 4 and 9 brighter lights or lower pitch sounds occurred in each exposure phase).

(2) Test phase. In the test phase, directly following the exposure phase, three multimodal stimulus pairs were tested in a randomly intermixed fashion: audio-visual (AV), visual-tactile (VT), and audio-tactile (AT). For each stimulus pair we measured psychometric functions using 11 SOAs between the pair of multisensory stimuli: −240, −120, −90, −60, −30, 0, 30, 60, 90, 120, 240 ms. We defined the SOA to be negative in case the visual stimulus preceded the auditory (AV), the visual stimulus preceded the tactile (VT), and the tactile stimulus preceded the auditory (AT). Within a block each SOA appeared once at random, making up 33 test trials for each block (3 test pairs × 11 SOAs). Each test trial was preceded by a short re-exposure consisting of eight asynchronous stimulus pairs with maximum asynchrony of 150 ms (top-up exposure). The test stimulus pair was identifiable by a longer pause before it. For each test stimulus pair, participants judged which of the two stimuli was first
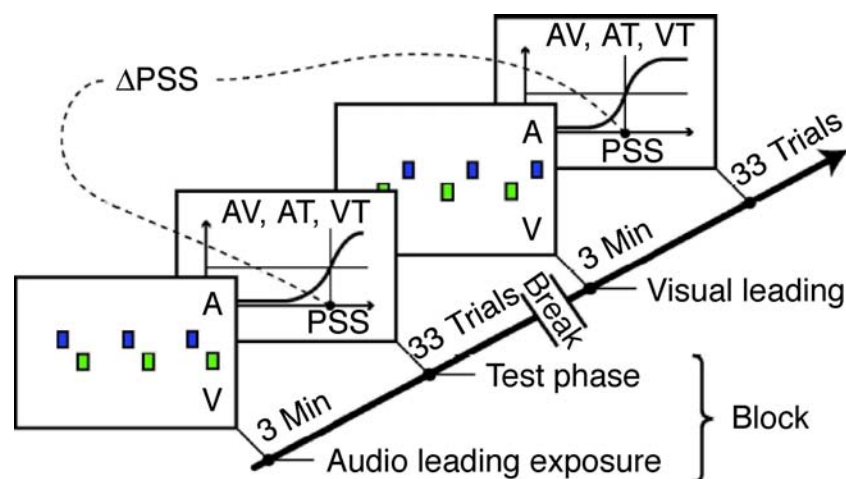

Figure 4. Timeline of experimental protocol including experimental procedure.

(temporal order judgment — TOJ) and entered their response pressing one of three keys for vision, audition, or touch using the right hand.

In summary, throughout the experiment each measurement was repeated 8 times for a total of 528 trials (8 repetitions × 2 audio/visual leading exposure conditions × 3 test pairs × 11 SOAs). These trials were divided into the 16 blocks (8 repetitions × 2 exposure conditions) presented randomly on four consecutive days. Including exposure, a block in the test phase lasted 8 to 10 minutes. After each block (exposure plus test), participants left the experimental room to take a 2-minutes break.

Participants' responses to the test stimuli were fitted with a cumulative Gaussian using psignifit (Wichmann & Hill, 2001). From these fits we determined the Points of Subjective Simultaneity (PSS). Our measure of recalibration was ΔPSS, which was determined by subtracting each subject's PSS in the test phases following audio leading exposure from the PSS in the test phases following visual leading exposure. We chose to have two types of exposure and compare the two PSS after adaptation instead of performing a comparison between the PSS obtained before the exposure with the one after exposure with only one type of exposure asynchrony to avoid order and training effects.

There was no feedback during the experiment. To acquaint participants with the temporal order judgment task, each day there were 60 training trials preceding the experiment (without prior exposure phase).

Results

For each stimulus pair, average values of ΔPSS, the difference in PSS after the audio and visual leading exposure conditions, are shown in Figure 5. Individual data is reported in the Appendix A. In accordance with previous studies (Fujisaki et al., 2004; Hanson et al., 2008; Harrar & Harris, 2005, 2008; Heron et al., 2007; Vatakis et al., 2007; Vroomen et al., 2004), we found that exposure to asynchronous AV stimuli influences the subsequent perception of AV simultaneity. After prolonged exposure, observers adapt to the asynchrony so that they perceive stimulus pairs with the same asynchrony as being temporally less discrepant. For the AV pair, all observers' PSS differed between the two exposure conditions in the predicted direction ($\Delta PSS_{AV} = 26 \pm 6$ ms, mean $\pm$ *SE*; one-tailed *t*-test against 0: $t(6) = 4.55$, $p = 0.0019$). That is, to perceive the auditory and visual stimuli to be simultaneous, the auditory stimulus had to be presented 26 ms earlier after audio leading exposure than after visual leading exposure. The difference in PSS corresponds to 8.9% of the ±150 ms AV asynchrony during exposure. The magnitude of this effect is in line with other reports on AV recalibration of perceived simultaneity (Fujisaki et al., 2004, 12.5%; Keetels & Vroomen, 2007, 6.5%; Vroomen et al., 2004, 6.7%).
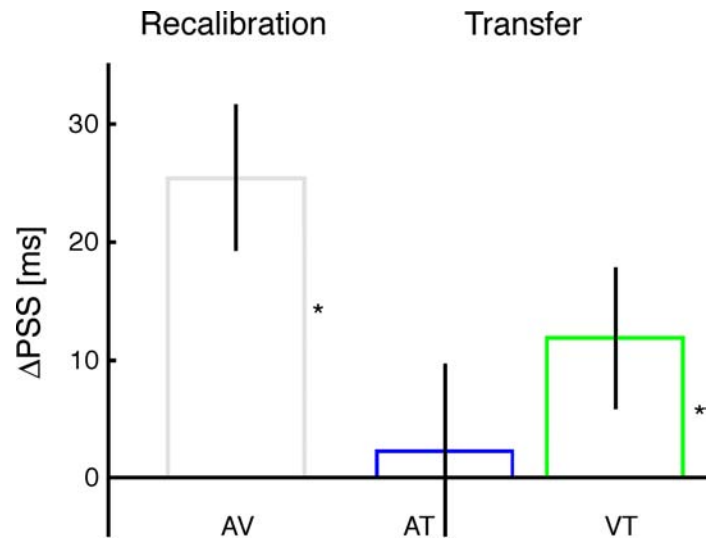

Figure 5. ΔPSS obtained for the three stimulus pairs in Experiment 1. Auditory stimuli were presented without headphones. Error bars represent the standard error of the mean across participants. Significant effects are indicated by * ($p < 0.05$).

More interestingly, we also found an effect of asynchronous AV exposure on ΔPSS in the VT pair ($\Delta PSS_{VT} = 12 \pm 6$ ms; $t(6) = 2.18$, $p = 0.036$). That is, there is a difference in VT synchrony perception following auditory leading AV exposure compared to vision leading AV exposure. This indicates that the visual stimuli had to be presented earlier after visual leading AV exposure than after auditory leading AV exposure in order to perceive the visual and tactile stimuli as simultaneous. There was no significant change in ΔPSS in the AT pair ($\Delta PSS_{AT} = 2$ ms ± 7 ms; $t(6) = 0.34$, $p = 0.37$).

The transfer of AV recalibration to another type of stimulus pair not presented during exposure (VT) indicates that a mechanism based on latency adjustment is involved in the recalibration of perceived simultaneity (see Figure 1). Since the effect on AT stimuli is minimal, it is the perceptual latency of visual stimuli that is prevalently adjusted (Figure 2). This change may indicate that in these experimental conditions the auditory estimates are trusted more.

Does this exclude the possibility of a recalibration mechanism based on remapping? No, it does not. However, because the sum of the transfer effects to VT and AT stimuli does not significantly differ from the AV recalibration effect (one tailed paired-sample *t*-test, $t(6) = 1.00$, $p = 0.17$) and so the transfer is indistinguishable from being complete, these data also provide no evidence that a mechanism based on remapping plays a role for AV recalibration of perceived simultaneity.

Our transfer results are at odds with the findings by of Harrar and Harris (2005) who failed to obtain a change in VT perceived simultaneity following exposure to asynchronous AV stimulus pairs. One key difference between

the two studies is the way auditory stimuli were presented. While in the current experiment visual, tactile, and auditory stimuli were all co-located, in Harrar and Harris (2005) only the visual and tactile stimuli were co-located while the auditory signals were presented using headphones. So there was a spatial discrepancy between the auditory stimuli and the other signals. Although it has been shown that the amount of recalibration of AV simultaneity is not affected by a spatial discrepancy between the auditory and the visual stimuli (Keetels & Vroomen, 2007) or by wearing headphones (Fujisaki et al., 2004), it is not yet determined which of the two signals—the visual or the auditory—is the one that is affected by recalibration. That is, it may be that the way in which the stimuli are presented affects which estimate is trusted more and, hence, which of the signals undergoes a change of latency during recalibration. For example, wearing headphones may result in the auditory signal to be the odd one out (as it is not co-located with vision and touch) and it is the only signal that is somewhat unnatural (as it is not spatialized and does not change with head movements). If the presentation mode really affects which signal undergoes a change, the amount of transfer to the different modality pairs (VT and AT) should be different when wearing headphones without influencing the magnitude of AV recalibration itself. In conclusion, the inconsistency between our current results and the results by Harrar and Harris might be due to wearing headphones as it is reasonable to assume that using headphones causes the auditory signal to be the odd one out and thus may decrease the trust assigned to the auditory estimate (see Arnold, Johnston, & Nishida, 2005). Less trust in the auditory estimate would predict more change in the perceptual latency of the auditory signal and less change in the one of the visual signal during recalibration. Thus, with headphones we might expect a decreased transfer effect to VT and an increased transfer effect to AT perceived simultaneity. To test this prediction, in the next experiment we manipulated the way in which the auditory stimuli were presented in order to determine whether it influences recalibration transfer.

Experiment 2

In Experiment 1 we showed that there is a transfer of recalibration from exposure to AV asynchrony to VT synchrony perception. Thus, with co-located multisensory stimulation the perceptual latency of the visual stimulus changes during exposure to AV asynchrony, whereas the perceptual latency of auditory stimuli stays more or less constant. One reason for this may be that the perceptual system trusts the auditory more than the visual modality under these conditions and therefore uses it as the standard

for recalibrating the perceived timing of the visual stimuli. Is this a general property of the auditory modality to be superior over vision when it comes to temporal recalibration or does the trust in an estimate depend on the particular perceptual situation or task? Whether or not a modality is trusted could be due to many factors. For example, by presenting auditory stimuli over headphones, the perceived location of the auditory stimuli is no longer the same as the visual and the tactile stimuli. It is known that spatial proximity can have a strong effect on perceived simultaneity of multimodal stimuli (Bertelson & Aschersleben, 2003; Calvert, Spence, & Stein, 2004; Driver & Spence, 2004; Spence, Baddeley, Zampini, James, & Shore, 2003; Wallace et al., 2004; Zampini, Guest, Shore, & Spence, 2005). This may suggest that co-location could be relevant to determine which modality provides the more trusted estimate (see Arnold et al., 2005). Changing from co-located stimuli to headphones might decrease the trust in the auditory estimate in favour of the visual. Consequently, we may expect a decrease of transfer of exposure from AV to VT and an increase from AV to AT.

Methods

The apparatus and experimental procedure were identical to Experiment 1.

Stimuli

Auditory stimuli were presented using noise-attenuating headphones (KOSS QZ99). A speaker positioned away from the apparatus produced continuous white noise (74 dB SPL).

Participants

We tested nine naive participants. They had normal hearing, no somatosensory disorders, and normal or corrected-to-normal vision.

Results

For the AV stimulus pair in the test phase, all observers' PSS differed between the two exposure conditions (visual leading and audio leading) in the predicted direction (ΔPSS_{AV} = 28 ms ± 13 ms, mean ± SE; one-tailed t-test against 0, $t(8)$ = 2.27, p = 0.026). That is, to perceive the auditory and visual stimuli as simultaneous, the auditory stimulus had to be presented earlier after audio leading exposure than after visual leading exposure (Figure 6). Individual data is reported in the Appendix A. The magnitude of the AV recalibration effect is comparable
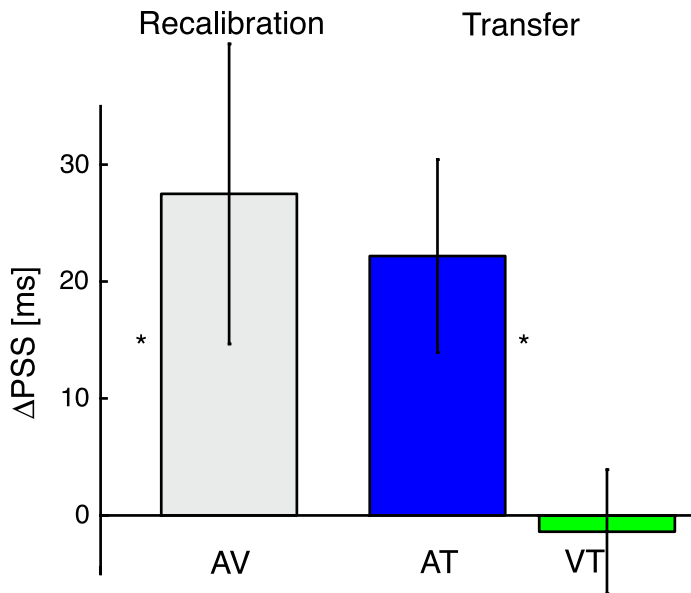
Figure 6. ΔPSS obtained for the three stimulus pairs in Experiment 2. Auditory stimuli were presented with headphones. Significant effects are indicated by * ($p < 0.05$).

to what we found in Experiment 1 without headphones (9.5% of the ±150 ms exposure asynchrony).

Most interestingly, however, in this experiment we now found an effect of transfer in the AT but not in the VT condition. Participants' PSS for the AT stimulus pair differed significantly depending on the AV exposure condition (ΔPSS_{VT} = 22 ms ± 8 ms; $t(8)$ = 2.85, p = 0.011). That is, to perceive the auditory and tactile stimuli as simultaneous in this transfer condition, the auditory stimulus had to be presented earlier after audio leading exposure than after visual leading exposure (see Figure 2). There was no effect of exposure condition on ΔPSS of the VT stimulus pair (ΔPSS_{AT} = 1 ms ± 5 ms; $t(8)$ = 0.28, p = 0.61). These results indicate that with headphones the perceptual latency of the auditory stimulus is recalibrated while perceptual latency of the visual stimulus stays more or less constant. This agrees with the findings of Harrar and Harris (2005) who also found no change of VT perceived simultaneity following exposure to AV asynchrony when the auditory stimuli were delivered via headphones. Unfortunately, they did not test for transfer using AT stimuli.

As in Experiment 1, the sum of the transfer effects to VT and AT stimuli did not differ significantly from the AV recalibration effect (one-tailed paired-sample t-test, $t(8)$ = 0.53, p = 0.33), indicating again that transfer is more or less complete. Thus, also under these experimental conditions there is no evidence that a mechanism based on remapping plays a role for AV recalibration.

In summary, the transfer of AV recalibration to VT and AT differs significantly between the two presentation conditions of Experiments 1 and 2: exposure to AV asynchrony with co-located multisensory stimuli does not

have a significant effect on AT perceived simultaneity, but does transfer to the VT stimulus pair; in contrast, exposure to AV asynchrony with sound stimuli presented via headphones does not have an effect on VT perceive simultaneity, but does transfer to the AT stimulus pair. Thus, with headphones the perceptual latency of the auditory stimulus is recalibrated, while the latency of the visual stimulus changes with co-located stimuli (Figure 2). These results are consistent with the auditory estimates being trusted when the multisensory stimuli are co-located, and the visual estimates being trusted when auditory stimuli are presented via headphones. Said another way, the visual stimuli undergo modification in Experiment 1, and the auditory stimuli undergo modification in Experiment 2. Taken together, these results suggest that a mechanism is required that modifies perceptual latency of one or the other stimuli during temporal recalibration.

Experiment 3

In Experiment 3, we tested the hypothesis that the perceptual latency is altered during recalibration by investigating how AV recalibration affects simple reaction times to audio and visual stimuli. We did this again for the two presentation conditions—co-located and headphones—and predicted a change in the visual reaction times when the stimuli were co-located during AV recalibration and a change in the auditory reaction times when the auditory stimuli were presented via headphones.

The goal of any multisensory recalibration mechanism is to reduce the perceptual disparity between associated stimuli. Here this disparity is the temporal asynchrony between the audio and visual stimuli presented repeatedly during exposure. If such a recalibration is achieved by adjusting the perceptual latency of a sensory stimulus depending on the stimulus presentation (as suggested by Experiments 1 and 2), we expect a change in the visual modality when the experiment is conducted with co-located stimuli, and a change in the auditory modality when it is conducted with headphones (results shown in Figure 7). If the visual stimulus is the one affected, the recalibration of perceptual simultaneity may be obtained with either a decrease in perceptual latency after audio leading exposure or with an increase in perceptual latency after visual leading exposure. Both these modifications of perceptual latency would reduce the perceived asynchrony. Vice versa, if the processing of the auditory stimuli changes with co-located stimuli this may be due to a decrease in latency after visual leading exposure or an increase in latency after auditory leading exposure (results shown in Figure 7). The goal of this experiment is to determine whether the transfer of perceived simultaneity in the previous experiments can also be found using RT.
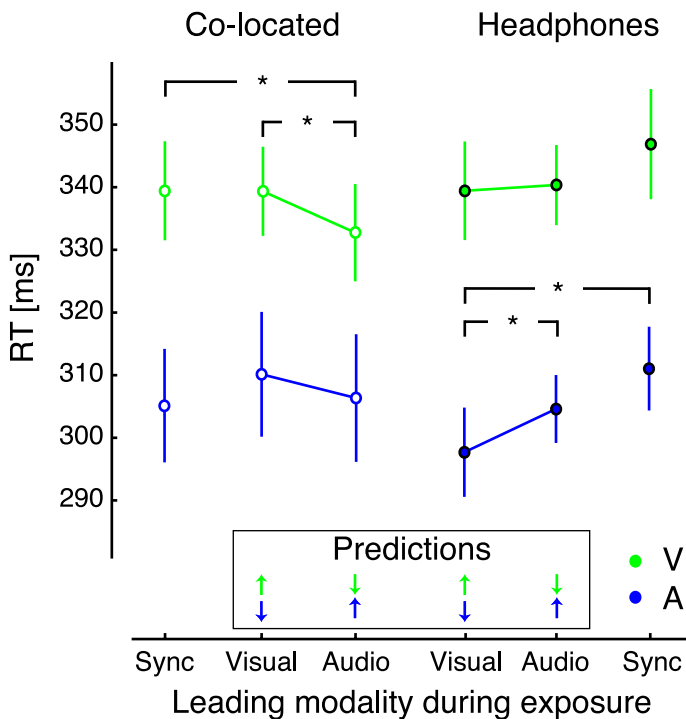
Figure 7. Average reaction time to auditory (blue) and visual (green) stimuli after audio leading, visual leading, and synchronous exposure. The open symbols represent the co-located condition; the filled symbols show the headphone condition. The arrows in the prediction box indicate the predicted direction of changes in reaction time in case the asynchrony during exposure (visual leading or auditory leading) is undone by recalibrating the visual and auditory perceptual latencies. The predictions assume that the reaction times are directly related to the perceptual latencies. The asterisks indicate significant effects. Error bars represent the standard error of the mean across the 14 participants. Significant effects are indicated by * ($p < 0.05$).

We expect that reaction time will capture the difference in perceptual latency for the visual and auditory modalities. To test for consistency of results, we will compare the obtained changes in reaction times for the two presentation conditions (i.e., headphones vs. no headphones) to the transfer effects of the previous experiments.

Methods

The apparatus was the same as in Experiments 1 and 2.

Stimuli

Auditory stimuli were either presented co-located with the visual stimuli via speakers or separated via noise-attenuating headphones (KOSS QZ99). When stimuli were delivered via headphones, an additional speaker distant from the apparatus produced continuous white noise (74 dB SPL).

Participants

We tested fourteen naive participants. They all had normal hearing, no somatosensory disorders, and normal or corrected-to-normal vision.

Procedure

Each experimental block consisted of exposure and test phase. The exposure phase was 4 minutes long and consisted of AV asynchronous or synchronous stimuli. The SOA was −150 ms for the visual leading asynchrony, +150 ms for the auditory leading asynchrony, and 0 ms for synchronous stimulation. Otherwise, the exposure phase was identical to the one in Experiments 1 and 2.

The test phase, directly following exposure, consisted of 60 trials of auditory and visual stimuli, 30 presented to each modality at random. Participants had to press the same key on a PS/2 keyboard as fast as possible when they saw or heard a stimulus. No feedback was provided. Once the response was given, there was a pause of random length (2 to 5 s) before the next trial. There was no top-up exposure between trials. We excluded trials with RTs longer than 2 seconds or more than 3 standard deviations from the participant's average.

In total there were 6 blocks, 3 conducted with headphones and 3 with co-located stimuli. In summary, we used a factorial design with three within-subjects factors: 2 sound presentation conditions (external co-located and headphones); 3 exposure conditions (audio leading, visual leading, and simultaneous); and 2 test stimulus types (audio and visual stimuli). Each combination of exposure condition and sound presentation was presented only once. Prior to each experimental block participants were acquainted with the RT task by responding as quickly as possible to 100 visual and auditory trials (50 in each modality) presented at random.

Results

A three-way repeated-measure ANOVA performed on the RT data showed that there were two expected main effects (Figure 7). First, the reaction times to auditory stimuli (blue symbols) were faster by 34 ms than they were to visual stimuli (green symbols) indicating, consistent with the literature, longer processing time for visual stimuli (stimulus type, $F(1,13) = 4.78$, $p = 0.048$). Second, presentation with headphones shortened reaction times to auditory stimuli by 3 ms, which corresponds to the physical conduction time from the speaker to the ears (presentation condition, $F(1,13) = 5.59$, $p = 0.034$; presentation condition × stimulus type $F(1,13) = 11.53$, $p = 0.005$).

Most importantly, however, the only other significant effect revealed by the ANOVA is the three-way interaction,

which indicates that the exposure condition differentially affected participants' responses to the auditory and visual stimuli depending on whether the auditory stimuli were produced by headphones or speakers ($F(1,13) = 4.98$, $p = 0.044$). To analyze the details of this result, consider that, in general, the recalibration effect should depend on the type of asynchrony during exposure, i.e., whether vision or audition was the leading modality during exposure. That is, to reduce the perceived asynchrony during visual leading exposure, after recalibration either vision needs a longer perceptual latency or audition needs a shorter one. Vice versa, to undo the asynchrony during auditory leading exposure after recalibration either audition needs a longer perceptual latency or vision a shorter one (see predictions in Figure 7). From the transfer effect obtained in Experiment 1 we expect the change during recalibration without headphones to occur in the visual modality. Furthermore, from the transfer effect obtained in Experiment 2 we expect the change during recalibration with headphones to occur in the auditory modality. Consistent with these predictions we find that without headphones the change in reaction time occurs in the visual modality (paired sample t-test one-tailed: V: $t(13) = 2.06$, $p = 0.030$; A: $t(13) = 0.11$, $p = 0.54$), whereas it occurs in the auditory modality with headphones (A: $t(13) = 2.35$, $p = 0.018$; V: $t(13) = 0.65$, $p = 0.73$). Specifically, by comparing the synchronous and asynchronous exposure conditions, we can confirm that with co-located stimuli there is a faster RT to visual stimuli after visual leading exposure, whereas with headphones the significant effect is due to faster RT to auditory stimuli after audio leading exposure (see Table 1 for statistical tests).

Taken together, the difference in RT to auditory and visual stimuli after asynchronous exposure is in qualitative agreement with the change in PSS measured in Experiments 1 and 2: vision changes with co-locates presentation, whereas audition changes with headphones. The change in RT to visual stimuli for the co-located presentation condition (6.4 ms) corresponds to 4.3% of the exposure asynchrony and the RT change to auditory stimuli for the headphone presentation condition (13.3 ms) corresponds to 8.9% of the exposure asynchrony. Even though the magnitude of the change in RT doubles with headphone presentation, the two values do not differ significantly from each other (paired-sample t-test:

$(13) = 1.09$, $p = 0.29$) and roughly correspond to what has been found in Experiments 1 and 2 (8.9% and 9.5%, respectively).

In summary, these results support our earlier conclusions, that with co-located stimuli the perceptual system trusts auditory estimates while it modifies the perceptual latency of visual stimuli, in order to decrease the perceived asynchrony between AV stimuli. Consistent with the idea of relative trust, when the auditory stimuli are affected by wearing headphones, the perceptual system seems to shift the trust towards the visual estimate and the auditory stimuli undergo a change in perceptual latency. These results are in agreement to what we found in the first two experiments: different transfer of the recalibration effect due to the presentation conditions of the auditory stimuli, which may indicate a change in the relative trust.

Discussion

We investigated the mechanism involved in temporal recalibration. In particular we tested whether and under what circumstances audio-visual recalibration transfers from the exposure stimulus to different pairs of stimuli involving the tactile modality. The results indicate that exposure to asynchronous AV stimulus pairs influences not only AV simultaneity perception, but they also indicate that there is a transfer to the perceived simultaneity of VT and AT stimuli. The amount of transfer depends on the presentation conditions: with speakers, the multisensory stimuli were co-located and we found that AV perceived synchrony transfers to the VT stimulus pairs (Experiment 1). In contrast, with headphone presentation—and therefore non-co-located stimuli—we found transfer of AV recalibration to the AT stimuli (Experiment 2). Consistent with the effect of transfer, we found that reaction times were also affected after exposure to AV asynchrony (Experiment 3): with speakers RT to visual stimuli changed, whereas with headphones RT to auditory stimuli changed. In both cases, RT decreased. Specifically, changes in RT without headphones are consistent with a decrease in the perceptual latency of visual stimuli after audio leading exposure, whereas

Stimulus	A				V			
Sound presentation	Co-located		Headphones		Co-located		Headphones	
Leading during exposure	Audio	Visual	Audio	Visual	Audio	Visual	Audio	Visual
$t(13)$	0.17	1.79	1.09	2.48	2.24	0.07	1.72	1.19
p	0.86	0.096	0.29	0.028*	0.043*	0.95	0.11	0.25

Table 1. Paired sample, two-tailed t-test for RT in the asynchronous vs. synchronous exposure conditions. *Note*: Significant effects are indicated by * ($p < 0.05$).

changes in RT with headphones are consistent with a decrease in the perceptual latency of auditory stimuli after visual leading exposure. Taken together, both the transfer of simultaneity recalibration and the changes in RT indicate that a reduction in perceptual latency causes the temporal recalibration of perceived simultaneity.

Previous studies have shown that AV temporal recalibration occurs whether auditory stimuli are presented via loudspeakers that are placed in the vicinity of the light source, or from speakers placed at 90 degrees from the light source, or via headphones (Fujisaki et al., 2004; Keetels & Vroomen, 2007). Even though the *magnitude* of the recalibration is independent of the spatial separation of the stimuli, our findings suggest that recalibration is achieved differently in these conditions. When the light source and the loudspeakers are co-located, recalibration is caused by a temporal shift of the visual stimulus toward the auditory percept. In contrast, the opposite occurs with headphone presentation.

It is well established that the distribution of attention across modalities as well as across space has an impact on the speed of processing (for reviews see Driver & Spence, 2004; Shore & Spence, 2005; Spence & McDonald, 2004). In particular, attended stimuli are perceived faster than unattended stimuli, an effect that has been termed "prior entry" (Titchener, 1908; for a review see Shore, Spence, & Klein, 2001). This effect can also affect the PSS (Schneider & Bavelier, 2003; Shore et al., 2001; Shore & Spence, 2005; Spence, Shore, & Klein, 2001; Zampini et al., 2005) although Fujisaki et al. (2004) showed that recalibration is a distinct phenomena and not an artifact of attention switching (when observers were instructed to attend to either one modality or the other, the prior entry effect was measured in addition to the recalibration shift). However since such a shift is indistinguishable from temporal recalibration, one could assume that in our study a different deployment of attention between audition and vision in the different sound presentation conditions (co-located vs. headphones) could have led to an adjustment of the perceptual latency of either the visual or the auditory stimulus therefore explaining the pattern of results found. In the following we will show that there is no plausible consistent argument based on attention that could explain all of our results. Our argument against a possible effect of attention is twofold.

(1) Although it is in principle impossible to exclude the presence of prior entry effects, a very particular distribution of attention across leading stimulus conditions and sound presentation conditions would be necessary to consistently explain all our findings. In particular, observers would have had to attend: (i) to visual stimuli after audio leading exposure with co-located presentation (visual prior entry), (ii) to auditory stimuli after visual leading with headphones presentation (auditory prior entry), and (iii) equally to both the visual and auditory stimuli in the other conditions (no prior entry). We see no

plausible reason to assume such an intricate pattern of behaviour.

(2) Another possibility is that observers could have attended to different spatial locations in the two sound presentation conditions: either to the location of the loudspeaker or to the perceived location of sound presented with headphones (at or near the head). Previous research suggests that the spatial separation of stimuli can have an impact on the speed of sensory processing within and across sensory modalities (e.g., Jonides, 1981; Spence & Driver, 1996; Spence & McGlone, 2001; for a review on multimodal stimuli, see Spence & McDonald, 2004). If two stimuli are presented in short succession, the perceptual latency of the second stimulus is shortened if it is presented at the same location as the first (Kato & Kashino, 2001). This benefit is lost with increasing distance between the stimuli (e.g., e.g. Chong & Mattingley, 2000; Driver & Spence, 1998; Frassinetti, Bolognini, & Làdavas, 2002). The generally accepted explanation for this effect is that attention is involuntarily drawn to the first stimulus and this shift results in enhanced processing at the cued location. However, if the second stimulus appears at a different location, it then takes time for the focus of attention to shift (Spence, 2001), resulting in slower responses. Applying this account to our data should lead us to expect faster processing of the second stimulus in the co-located presentation condition, but not in the headphones presentation condition since auditory stimuli appear to be spatially separated. However, our data clearly indicates a decrease in perceptual latency in both cases: RTs are reduced for visual signals after audio leading exposure in the co-located condition, but RTs are also reduced for audio signals after visual leading exposure in the headphone presentation condition. So, even though the second signal of the pair is always the one affected by the exposure, this does not happen exclusively in the co-located condition.

Having excluded attention as a possible explanation for the observed pattern of results, we now direct our attention to the original motivation of this research. Our original goal was to distinguish whether the mechanism involved in temporal recalibration is based on the adjustment of the perceptual latency of individual signals or whether it is based on remapping (Figure 1). These two mechanisms make opposite predictions with respect to the transfer of recalibration to other modality pairs. Because perceptual latency adjustments are specific to an individual sensory signal, transfer is to be expected. Remapping effects, on the other hand, are specific to an individual pair of stimuli, so there should be no transfer. In Experiments 1 and 2 we found the sum of the transfer effects to AT and VT to be statistically indistinguishable from the AV recalibration effect. Thus, there is nearly complete transfer, which indicates that the recalibration of perceived simultaneity can be explained almost entirely by a mechanism based on perceptual latency adjustment. For the particular perceptual situation studied here, this does not exclude, but

leaves little room for a temporal recalibration mechanism based on remapping.

Our results of transfer are in agreement with what was previously found in the literature: in particular, Harrar and Harris (2005) also exposed participants to asynchronous AV stimuli and they subsequently determined perceived simultaneity of either AV stimuli or VT stimuli (AT stimuli were not tested). Stimuli were delivered through headphones. Results indicate no change of VT perceived simultaneity following exposure to AV asynchrony. Consistently, in our study we also found no significant transfer effect on the VT stimuli when auditory stimuli were presented using headphones (Experiment 2), whereas we found a transfer effect on the AT stimuli. In a subsequent study, Harrar and Harris (2008) presented auditory stimuli without headphones and exposed participants to asynchronous stimuli of different combinations. Although they did not find significant transfer effects to VT or AT perceived simultaneity after VA exposure, the pattern of result is not much different from what we obtained here in Experiment 1. Harrar and Harris (2008) also measured RT before and after exposure to asynchronous AV stimulation. Pre-post test comparison, however, might create order effects (e.g., causing reaction times in the post-test to be generally slower than in the pre-test) that make the comparison between the two sets of results difficult. Recently, Navarra, Hartcher-O'Brien, Piazza, and Spence (2009) studied the effect of AV exposure on RT to audio and visual stimuli by comparing change in RT after audio leading and synchronous exposure (similar to Experiment 3). They found a change in the RT to auditory stimuli and no change to visual stimuli, which is similar to our findings in Experiment 3 with non-co-located stimuli. Navarra et al. did not use headphones, but the stimuli were not co-located. Sound stimuli were, in fact, presented with speakers placed away from the monitor and while white noise was played in the background.

The difference in the pattern of results with different sound-presentation conditions indicates that recalibration is not obtained by modifying sensory signals always in the same sensory modality. Instead, the two types of transfer can be attributed to modifications of perceptual latency in either the auditory or the visual modality. We argue that the difference can be attributed to a shift in trust given to the sensory estimate. Ideally, relative trust should be inversely related to the likelihood of bias in the sensory estimate. However this likelihood cannot be known directly from the signal, as this does not contain information about the likelihood of bias. Thus trust must be based on prior knowledge. In many ways this concept is analogous to the one used for describing cue integration: in the cue combination framework the weighting of sensory signals is inversely proportional to the variance of the probability density function associated with a sensory signal (e.g., e.g. Ernst & Banks, 2002; Landy, Maloney, Johnston, & Young, 1995), whereas in the recalibration framework discussed here the relative trust given to a sensory signal is inversely related to the likelihood of bias of a sensory estimate.

What determines the likelihood of bias? In the literature there are not many models that provide a quantitative account of perceptual recalibration. One exception is Ghahramani, Wolpert, and Jordan (1997) who suggest that the amount of recalibration of a sensory estimate should be proportional to the relative variance of the signals. However, it seems intuitive that the variance of a sensory signal (i.e., its inverse reliability) does not necessarily determine the likelihood of bias, so that this model does not seem satisfactory. To illustrate this, consider that every type of technical sensor has a different "variance-to-likelihood of bias" ratio. For example, the position estimate derived from an Inertial Measurement Unit (IMU) usually has little variance, but it has a considerable drift and it becomes biased over time, which means it has a low accuracy. On the other hand, the position estimate derived from the Global Positioning System (GPS) has large variance but is almost never biased (i.e., it has a high accuracy). As can be seen from comparing these two sensors, the variance of the signal and the amount of bias are generally not predicting each other. Thus, it is non-optimal to correct for the bias based on the variance of the estimate as suggested by Gharamani et al. (1997) as the likelihood of bias depends primarily on other factors. For example, whether or not a system has previously been biased should be a determining factor in determining the likelihood of bias. That is, prior knowledge about the previous accuracy should be one of the factors in determining the trust in a sensory estimate. To stay within the example of estimating positions using IMU and GPS data, the bias in the IMU estimate is often corrected for by using the GPS estimate, although the variance of the GPS estimate is significantly higher than of the IMU estimate. This is because from past measurements we have prior knowledge that even a high variance GPS estimate is unlikely to contain a large bias whereas it is very likely for the IMU estimate to be biased.

An optimal way of combining two estimates, which is used in technical systems, is a Kalman filter. This filter merges the advantages of the IMU and GPS estimates in the example above by recalibrating for the bias in the position information derived from the IMU using the GPS estimate, while still obtaining a low variance position estimate. Burge, Ernst, and Banks (2008) recently used a model based on a Kalman filter to describe human visuomotor recalibration. In this study they did not measure which of the two systems recalibrated—the visual or the motor system—but they confirmed another property of the Kalman filter, namely that the variance of the signals is a determining factor for the rate of recalibration. That is, although the variance of the signals should not determine the amount of recalibration in each signal, it determines the rate at which recalibration occurs. In contrast, in the present experiments we investigated which signal is recalibrated and our data seem consistent

with the idea that signal variance is not a determining factor: auditory reliability does not change significantly whether or not the auditory signals are presented via headphones. Still it makes a large difference which of the two signals will recalibrate depending on the presentation conditions.

In summary, if our theory is correct and recalibration is based on the likelihood of bias, we can deduce from the results of Experiment 1 (in which we found that with co-located signals it is the latency of visual stimuli that recalibrates) that under normal presentation conditions for what concerns perception it must be the visual and not the auditory modality that is more likely to contain a temporal bias (see also Arnold et al., 2005).

There are other factors that might affect the trust in a sensory estimate (i.e., the likelihood of bias). If there were more than two signals contributing to the estimation process, one of which is being biased, the majority of signals would provide congruent information whereas one is off. Such an asymmetry could be a clear indicator for which of the estimates is more likely to contain a bias. Another indicator for one signal being the odd one out might be that the signals do not behave as normally expected when interacting with the world. For example, when wearing headphones the sound moves with the head motion and so does not represent a fixed external stimulus as most auditory signals naturally do. This is the situation we have in Experiment 2: here the auditory signals are presented via headphones and so they are perceived at a different location from the other sensory stimuli, attached to the head and its motion (there were small head movements throughout the experiment, as we did not use head or chin rests). This makes it reasonable for the perceptual system to assume that under the conditions of Experiment 2 it is the auditory estimates that are more likely to be biased, so they are trusted less. As a consequence, given these conditions it is the auditory estimates that are temporally recalibrated toward the visual, which are here used as the standard.

In all experiments the magnitude of the recalibration effect was roughly 10% of the asynchrony during exposure. Since the effect of recalibration was measured as a contrast (ΔPSS, ΔRT) between the two exposure conditions (visual leading vs. audio leading), the magnitude of recalibration within one session is circa 15 ms. It is unclear whether prolonged exposure to AV asynchrony would cause the recalibration effect to increase or whether there is a limit on the magnitude of recalibration imposed by the mechanism that modifies perceptual latency. Others, however, have shown that this limit is well above 15 ms (Fujisaki et al., 2004). At this point we do not know how the perceptual system adjusts perceptual latency, but multiple mechanisms are conceivable. Three of those are illustrated in Figure 8, which shows a schematic version of an hypothetical neural response to a stimulus obtained by convolving the signal with an impulse response function: (a) the time required to process a stimulus could
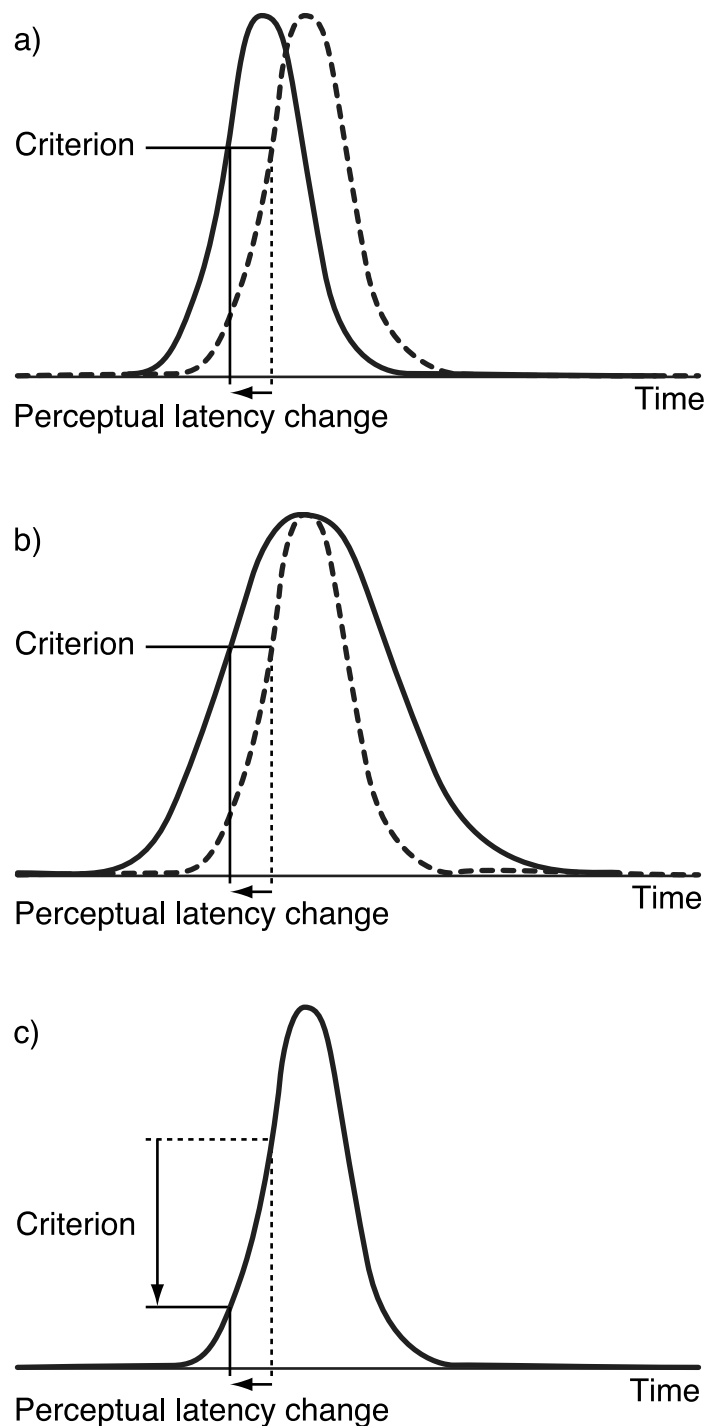

Figure 8. Schematic illustration of a signal convolved with an impulse response function (IRF) indicating possible mechanisms leading to the same perceptual latency change: (a) decrease of processing time, (b) change in the shape of the response function, (c) change in the criterion.

be varied by speeding up or slowing down processing; (b) alternatively, by adapting the processing characteristics the shape of the response function could vary, which would affect the perceptual latency if that is based on a

	Stimulus pair	S1	S2	S3	S4	S5	S6	S7	Mean	SEM
PSS audio leading	AV	8.7	−53.7	−35.2	66.8	116.6	29.4	−51.0	11.7	26.2
	AT	−61.5	−40.1	−31.0	−65.8	−23.6	−13.4	−72.9	−44.0	9.4
	VT	−44.3	−58.6	−74.8	32.5	−4.0	−100.8	−26.2	−39.5	18.3
PSS visual leading	AV	−31.9	−62.2	−43.6	53.4	76.4	−2.4	−86.6	−13.8	24.5
	AT	−62.4	−37.3	−53.8	−77.2	7.2	−10.4	−90.3	−46.3	14.4
	VT	−56.2	−48.3	−105.4	26.7	−32.0	−102.9	−40.9	−51.3	18.4
ΔPSS	AV	40.6	8.5	8.3	13.4	40.1	31.7	35.6	25.5	6.0
	AT	0.9	−2.8	22.8	11.4	−30.8	−3.0	17.4	2.3	7.2
	VT	12.0	−10.3	30.6	5.8	28.0	2.2	14.7	11.9	5.9

Table A1. Experiment 1.

criterion for detection; and (c) by simply adjusting the criterion to detect a stimulus the perceived latency may differ (see Sternberg & Knoll, 1973). This is probably the simplest mechanism to realize and thus the most likely one. Some results indicate that early components of neural processing can be speeded up by attention (Vibell, Klinge, Zampini, Spence, & Nobre, 2007) and concurrent sensory input (van Wassenhove, Grant, & Poeppel, 2005; see Ghazanfar & Schroeder, 2006, for a review). This speedup is compatible with all three proposed mechanisms for perceptual latency change. Similar modifications could happen also after prolonged exposure to asynchronous multimodal stimuli as the ones utilized here. These changes, however, might last beyond the end of the exposure, so that the change in perceptual latency would slowly fade away. The perceptual system, with the attribution of relative trust to the two estimates, determines in what proportion the adjustments should happen.

Appendix A

Values of PSS after audio leading and visual leading exposure obtained in the two experiments with TOJ task. Negative values indicate the case where the visual stimulus preceded the auditory (AV), the visual stimulus preceded the tactile (VT), and the tactile stimulus preceded the auditory (AT; Tables A1 and A2).

Acknowledgments

Footnote

[1]There may also be a trust in the mapping. The relative trust between sensory estimates and the mapping should determine a weight for the estimates to be adjusted or the mapping to be updated (Ernst & Di Luca, under review).

	Stimulus pair	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean	SEM
PSS audio leading	AV	−69.1	141.9	−104.0	−12.7	−7.3	−7.2	23.0	−65.1	−41.3	−15.8	23.6
	AT	−50.1	−156.5	−77.4	−55.1	−85.1	41.2	−70.2	−43.7	−58.7	−61.7	17.1
	VT	−71.0	90.9	−56.2	22.0	−10.6	−81.9	24.2	−73.7	−54.3	−23.4	19.7
PSS visual leading	AV	−54.6	54.5	−134.2	12.8	−16.5	−16.3	−27.8	−107.3	−100.4	−43.3	20.5
	AT	−64.8	−232.9	−84.9	−55.1	−95.7	−0.3	−93.3	−57.9	−70.8	−84.0	20.9
	VT	−86.5	83.4	−36.0	24.9	−2.7	−63.0	5.4	−58.5	−64.8	−22.0	18.1
ΔPSS	AV	−14.4	87.4	30.2	−25.5	9.1	9.2	50.9	42.2	59.1	27.6	12.1
	AT	14.7	76.4	7.6	0.0	10.6	41.5	23.1	14.2	12.0	22.2	7.8
	VT	15.5	7.6	−20.2	−2.9	−8.0	−18.8	18.8	−15.2	10.5	−1.4	5.0

Table A2. Experiment 2.

References

Allison, T., Matsumiya, Y., Goff, G. D., & Goff, W. R. (1977). The scalp topography of human visual evoked potentials. *Electroencephalography and Clinical Neurophysiology, 42,* 185–197.

Arnold, D., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research, 45,* 1275–1284. [PubMed]

Backus, B. T., & Haijiang, Q. (2007). Competition between newly recruited and pre-existing visual cues during the construction of visual appearance. *Vision Research, 47,* 919–924. [PubMed] [Article]

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology, 50,* 147–155. [PubMed]

Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision, 6*(5):2, 554–564, http://journalofvision.org/6/5/2/, doi:10.1167/6.5.2. [PubMed] [Article]

Burge, J., Ernst, M. O., & Banks, M. S. (2008). The statistical determinants of adaptation rate in human reaching. *Journal of Vision, 8*(4):20, 1–19, http://journalofvision.org/8/4/20/, doi:10.1167/8.4.20. [PubMed] [Article]

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

Chong, T., & Mattingley, J. B. (2000). Preserved crossmodal attentional links in the absence of conscious vision: Evidence from patients with primary visual cortex lesions. *Journal of Cognitive Neuroscience, 12,* 38.

Driver, J., & Spence, C. (1998). Crossmodal links in spatial attention. *Philosophical Transactions of the Royal Society of London B, 353,* 1319–1331.

Driver, J., & Spence, C. (2004). Crossmodal links in endogenous spatial attention. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 179–220). Oxford, UK: Oxford University Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. [PubMed]

Ernst, M. O., & Di Luca, M. (under review). Multisensory perception: From integration to remapping. In J. Trommershäuser, M. S. Landy, & K. Körding (Eds.), *Sensory cue integration*. Oxford, UK: Oxford University Press.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63,* 719–725. [PubMed] [Article]

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147,* 332–343. [PubMed]

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7,* 773–778. [PubMed]

Ghahramani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In P. G. Morasso & V. Sanguineti (Eds.), *Self-organization, computational maps and motor control* (pp. 117–147). Amsterdam: Elsevier Press.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cognitive Science, 10,* 278–285. [PubMed]

Hanson, J. V. M., Heron, J., & Whitaker, D. (2008). Recalibration of perceived time across sensory modalities. *Experimental Brain Research, 185,* 347–352. [PubMed]

Harrar, V., & Harris, L. R. (2005). Simultaneity constancy: Detecting events with touch and vision. *Experimental Brain Research, 166,* 465–473. [PubMed]

Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research, 186,* 517–524. [PubMed]

Heron, J., Whitaker, D., McGraw, P. V., & Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision, 7*(13):5, 1–8, http://journalofvision.org/7/13/5/, doi:10.1167/7.13.5. [PubMed] [Article]

Jonides, J. (1981). Towards a model of the mind's eye's movement. *Canadian Journal of Psychology, 34,* 103–112. [PubMed]

Kato, M., & Kashino, M. (2001). Audio-visual link in auditory spatial discrimination. *Acoustic Science & Technology, 22,* 380–382.

Keetels, M., & Vroomen, J. (2007). No effect of auditory-visual spatial disparity on temporal recalibration. *Experimental Brain Research, 182,* 559–565. [PubMed] [Article]

Keetels, M., & Vroomen, J. (2008). Temporal recalibration to tactile-visual asynchronous stimuli. *Neuroscience Letters, 430,* 130–134. [PubMed]

King, A. J. (2005). Multisensory integration: Strategies for synchronization. *Current Biology, 15,* R339–R341. [PubMed]

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination. In defense of weak fusion. *Vision Research, 35,* 389–412. [PubMed]

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17,* 154–163. [PubMed]

Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Science, 106,* 9123–9124. [PubMed]

Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience Letters, 413,* 72–76. [PubMed]

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research, 25,* 499–507. [PubMed]

Schneider, K. A., & Bavelier, D. (2003). Components of visual prior entry. *Cognitive Psychology, 47,* 333–366. [PubMed]

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14,* 147–152. [PubMed]

Shore, D. I., & Spence, C. (2005). Prior entry. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 89–95). USA: Elsevier.

Shore, D. I., Spence, C., & Klein, R. M. (2001). Visual prior entry. *Psychology Science, 12,* 205–212. [PubMed]

Spence, C. (2001). Crossmodal attentional capture: A controversy resolved? In C. Folk & B. Gibson (Eds.), *Attention, distraction and action: Multiple perspectives on attentional capture* (pp. 231–262). Amsterdam: Elsevier.

Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics, 65,* 318–328. [PubMed]

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1005–1030. [PubMed]

Spence, C., & McDonald, J. (2004). The cross-modal consequences of the exogenous spatial orienting of attention. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 3–25). Cambridge, MA: The MIT Press.

Spence, C., & McGlone, F. P. (2001). Reflexive spatial orienting of tactile attention. *Experimental Brain Research, 141,* 323–330. [PubMed]

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General, 130,* 799–832. [PubMed]

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

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 629–685). New York, NY: Academic Press.

Takahashi, K., Saiki, J., & Watanabe, K. (2008). Realignment of temporal simultaneity between vision and touch. *Neuroreport, 19,* 319–322. [PubMed]

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention.* New York: Macmillan.

van Wassenhove, V., Grant, K.W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, 102,* 1181–1186. [PubMed]

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research, 181,* 173–181. [PubMed]

Vibell, J., Klinge, C., Zampini, M., Spence, C., & Nobre, A. C. (2007). Temporal order is coded temporally in the brain: Early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *Journal of Cognitive Neuroscience, 19,* 109–120. [PubMed]

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

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research, 158,* 252–258. [PubMed]

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

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception & Psychophysics, 63,* 1293–1313. [PubMed]

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics, 67,* 531–544. [PubMed]

Zampini, M., Shore, D. I., & Spence, C. (2003). Audio-visual temporal order judgments. *Experimental Brain Research, 152,* 198–210. [PubMed]