
9 Perception of Synchrony between the Senses

Mirjam Keetels and Jean Vroomen

CONTENTS

9.1	Introduction	147
9.2	Measuring Intersensory Synchrony: Temporal Order Judgment Task and Simultaneity Judgment Task	148
9.3	Point of Subjective Simultaneity.....	150
9.3.1	Attention Affecting PSS: Prior Entry.....	151
9.4	Sensitivity for Intersensory Asynchrony	152
9.4.1	Spatial Disparity Affects JND.....	153
9.4.2	“Stimulus Complexity” Affects JND	154
9.4.3	Stimulus Rate Affects JND	155
9.4.4	Predictability Affects JND	155
9.4.5	Does Intersensory Pairing Affect JND?.....	156
9.5	How the Brain Deals with Lags between the Senses	156
9.5.1	Window of Temporal Integration	156
9.5.2	Compensation for External Factors	158
9.5.3	Temporal Recalibration	161
9.5.4	Temporal Ventriloquism	164
9.6	Temporal Synchrony: Automatic or Not?	167
9.7	Neural Substrates of Temporal Synchrony	169
9.8	Conclusions	170
	References.....	171

9.1 INTRODUCTION

Most of our real-world perceptual experiences are specified by synchronous redundant and/or complementary multisensory perceptual attributes. As an example, a talker can be heard and seen at the same time, and as a result, we typically have access to multiple features across the different senses (i.e., lip movements, facial expression, pitch, speed, and temporal structure of the speech sound). This is highly advantageous because it increases perceptual reliability and saliency and, as a result, it might enhance learning, discrimination, or the speed of a reaction to the stimulus (Sumby and Pollack 1954; Summerfield 1987). However, the multisensory nature of perception also raises the question about how the different sense organs cooperate so as to form a coherent representation of the world. In recent years, this has been the focus of much behavioral and neuroscientific research (Calvert et al. 2004). The most commonly held view among researchers in multisensory perception is what has been referred to as the “assumption of unity.” It states that, as information from different modalities share more (amodal) properties, the more likely the brain will treat them as originating from a common object or source (see, e.g., Bedford 1989; Bertelson 1999; Radeau 1994; Stein and Meredith 1993; Welch 1999; Welch and Warren 1980). Without a doubt, the most important amodal

property is temporal coincidence (e.g., Radeau 1994). From this perspective, one expects intersensory interactions to occur if, and only if, information from the different sense organs arrives at around the same time in the brain; otherwise, two separate events are perceived rather than a single multimodal one.

The perception of time and, in particular, synchrony between the senses is not straightforward because there is no dedicated sense organ that registers time in an absolute scale. Moreover, to perceive synchrony, the brain has to deal with differences in *physical* (outside the body) and *neural* (inside the body) transmission times. Sounds, for example, travel through air much slower than visual information does (i.e., 300,000,000 m/s for vision vs. 330 m/s for audition), whereas no physical transmission time through air is involved for tactile stimulation as it is presented directly at the body surface. The *neural* processing time also differs between the senses, and it is typically slower for visual than it is for auditory stimuli (approximately 50 vs. 10 ms, respectively), whereas for touch, the brain may have to take into account where the stimulation originated from as the traveling time from the toes to the brain is longer than from the nose (the typical conduction velocity is 55 m/s, which results in a ~30 ms difference between toe and nose when this distance is 1.60 m; Macefield et al. 1989). Because of these differences, one might expect that for audiovisual events, only those occurring at the so-called “horizon of simultaneity” (Pöppel 1985; Poppel et al. 1990)—a distance of approximately 10 to 15 m from the observer—will result in the approximate synchronous arrival of auditory and visual information at the primary sensory cortices. Sounds will arrive before visual stimuli if the audiovisual event is within 15 m from the observer, whereas vision will arrive before sounds for events farther away. Although surprisingly, despite these naturally occurring lags, observers perceive intersensory synchrony for most multisensory events in the external world, and not only for those at 15 m.

In recent years, a substantial amount of research has been devoted to understanding how the brain handles these timing differences (Calvert et al. 2004; King 2005; Levitin et al. 2000; Spence and Driver 2004; Spence and Squire 2003). Here, we review several key issues about intersensory timing. We start with a short overview of how intersensory timing is generally measured, and then discuss several factors that affect the point of subjective simultaneity and sensitivity. In the sections that follow, we address several ways in which the brain might deal with naturally occurring lags between the senses.

9.2 MEASURING INTERSENSORY SYNCHRONY: TEMPORAL ORDER JUDGMENT TASK AND SIMULTANEITY JUDGMENT TASK

Before examining some of the basic findings, we first devote a few words on how intersensory synchrony is usually measured. There are two classic tasks that have been used most of the time in the literature. In both tasks, observers are asked to judge—in a direct way—the relative timing of two stimuli from different modalities: the temporal order judgment (TOJ) task and the simultaneity judgment (SJ) task. In the TOJ task, stimuli are presented in different modalities at various stimulus onset asynchronies (SOA; Dixon and Spitz 1980; Hirsh and Sherrick 1961; Sternberg and Knoll 1973), and observers may judge which stimulus came first or which came second. In an audiovisual TOJ task, participants may thus respond with “sound-first” or “light-first.” If the percentage of “sound-first” responses is plotted as a function of the SOA, one usually obtains an S-shaped logistic psychometric curve. From this curve, one can derive two measures: the 50% crossover point, and the steepness of the curve at the 50% point. The 50% crossover point is the SOA at which observers were—presumably—maximally unsure about temporal order. In general, this is called the “point of subjective simultaneity” (PSS) and it is assumed that at this SOA, the information from the different modalities is perceived as being maximally simultaneous. The second measure—the steepness at the crossover point—reflects the observers’ sensitivity to temporal asynchronies. The steepness can also be expressed in terms of the just noticeable difference (JND; half the difference in SOA between the 25% and 75% point), and it represents the smallest interval observers can reliably

Please check
if the changes
made are OK.

notice. A steep psychometric curve thus implies a small JND, and sensitivity is thus good as observers are able to detect small asynchronies (see Figure 9.1).

The second task that has been used often is the SJ task. Here, stimuli are also presented at various SOAs, but rather than judging which stimulus came first, observers now judge whether the stimuli were presented simultaneously or not. In the SJ task, one usually obtains a bell-shaped Gaussian curve if the percentage of “simultaneous” responses is plotted as a function of the SOA. For the audiovisual case, the raw data are usually not mirror-symmetric, but skewed toward more “simultaneous” responses on the “light-first” side of the axis. Once a curve is fitted on the raw data, one can, as in the TOJ task, derive the PSS and the JND: the peak of the bell shape corresponds to the PSS, and the width of the bell shape corresponds to the JND.

The TOJ and SJ tasks have, in general, been used more or less interchangeably. Although comparative studies have found differences in performance measures derived from both tasks. Possibly, it reflects that judgments about simultaneity and temporal order are based on different sources of information (Hirsh and Fraisse 1964; Mitrani et al. 1986; Schneider and Bavelier 2003; Zampini et al. 2003a). As an example, van Eijk et al. (2008) examined task effects on the PSS. They presented observers a sound and light, or a bouncing ball and an impact sound at various SOAs, and had them perform three tasks: an audiovisual TOJ task (“sound-first” or “light-first” responses required), an SJ task with two response categories (SJ2; “synchronous” or “asynchronous” responses required), and an SJ task with three response categories (SJ3; “sound-first,” “synchronous,” or “light-first”

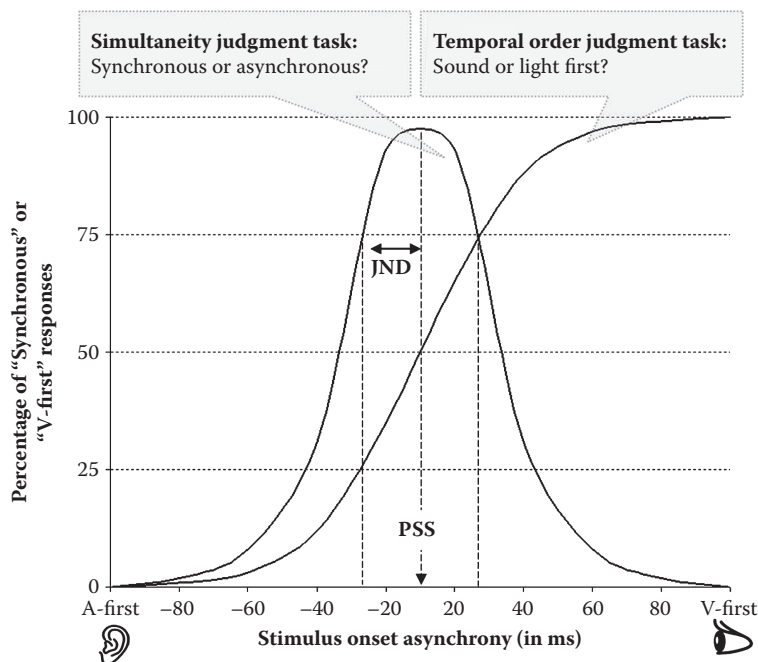


FIGURE 9.1 S-shaped curve typically obtained for a TOJ task and bell-shaped curve typically obtained in a simultaneity task (SJ). Stimuli from different modalities are presented at varying SOAs, ranging from clear auditory-first (A-first) to clear vision-first (V-first). In a TOJ task, the participant’s task is to judge which stimulus comes first, sound or light, whereas in SJ task, subjects judge whether stimuli are synchronous or not. PSS represents interval at which information from different modalities is perceived as being maximally simultaneous (~0 ms). In SJ task, this is the point at which the most synchronous responses are given; in TOJ task, it is the point at which 50% of responses is vision-first and 50% is auditory-first. JND represents the smallest interval observers can reliably notice (~27 ms). In SJ task, this is the average interval (of A-first and V-first) at which participant responds with 75% synchronous responses. In TOJ task, it is the difference in SOA at 25% and 75% point divided by two.

responses required). Results from both stimulus types showed that the individual PSS values for the two SJ tasks correlated well, but there was no correlation between the TOJ and SJ tasks. This made the authors conclude, arguably, that the SJ task should be preferred over the TOJ task if one wants to measure perception of audiovisual synchrony.

Although in our view, there is no straightforward solution about how to measure the PSS or JND for intersensory timing because the tasks are subject to different kinds of response biases (see Schneider and Bavelier 2003; Van Eijk et al. 2008; Vatakis et al. 2007, 2008b for discussion). In the TOJ task, in which only temporal order responses can be given (“sound-first” or “light-first”), observers may be inclined to adopt the assumption that stimuli are *never* simultaneous, which thus may result in rather low JNDs. On the other hand, in the SJ task, observers may be inclined to assume that stimuli actually belong together because the “synchronous” response category is available. Depending on criterion settings, this may result in many “synchronous” responses, and thus, a wide bell-shaped curve which will lead to the invalid conclusion that sensitivity is poor.

In practice, both the SJ and TOJ task will have their limits. The SJ2 task suffers heavily from the fact that observers have to adopt a criterion about what counts as “simultaneous/nonsimultaneous.” And in the SJ3 task, the participant has to dissociate sound-first stimuli from synchronous ones, and light-first stimuli from synchronous ones. Hence, in the SJ3 task there are two criteria: a “sound-first/simultaneous” criterion, and a “light-first/simultaneous” criterion. If observers change, for whatever reason, their criterion (or criteria) along the experiment or between experimental manipulations, it changes the width of the curve and the corresponding JND. If sensitivity is the critical measure, one should thus be careful using the SJ task because JNDs depend heavily on these criterion settings.

A different critique can be applied to the TOJ task. Here, the assumption is made that observers respond at about 50% for each of the two response alternatives when maximally unsure about temporal order. Although in practice, participants may adopt a different strategy and respond, for example, “sound-first” (and others may, for arbitrary reasons, respond “light-first”) whenever unsure about temporal order. Such a response bias will shift the derived 50% point toward one side of the continuum or the other, and the 50% point will then not be a good measure of the PSS, the point at which simultaneity is supposed to be maximal. If performance of an individual observer on an SJ task is compared with a TOJ task, it should thus not come as too big of a surprise that the PSS and JND derived from both tasks do not converge.

9.3 POINT OF SUBJECTIVE SIMULTANEITY

The naïve reader might think that stimuli from different modalities are perceived as being maximally simultaneous if they are presented the way nature does it, that is, synchronous, so at 0 ms SOA. Although surprisingly, most of the time, this is not the case. For audiovisual stimuli, the PSS is usually shifted toward a visual-lead stimulus, so perceived simultaneity is maximal if vision comes slightly before sounds (e.g., Kayser et al. 2008; Lewald and Guski 2003; Lewkowicz 1996; Slutsky and Recanzone 2001; Zampini et al. 2003a, 2005b, 2005c). This bias was found in a classic study by Dixon and Spitz (1980). Here, participants monitored continuous videos consisting of an audiovisual speech stream or an object event consisting of a hammer hitting a peg. The videos started off in synchrony and were then gradually desynchronized at a constant rate of 51 ms/s up to a maximum asynchrony of 500 ms. Observers were instructed to respond as soon as they noticed the asynchrony. They were better at detecting the audiovisual asynchrony if the sound preceded the video rather than if the video preceded the sound (131 vs. 258 ms thresholds for speech, and 75 vs. 188 ms thresholds for the hammer, respectively). PSS values also pointed in the same direction as simultaneity was maximal when the video preceded the audio by 120 ms for speech, and by 103 ms for the hammer. Many other studies have reported this vision-first PSS (Dinnerstein and Zlotogura 1968; Hirsh and Fraisse 1964; Jaskowski et al. 1990; Keetels and Vroomen 2005; Spence et al. 2003; Vatakis and Spence 2006a; Zampini et al. 2003a), although some also reported opposite results (Bald et al. 1942; Rutschmann and Link 1964; Teatini et al. 1976; Vroomen et al. 2004). There have

Please verify
if the changes
made are OK.

Please check
units used here.

been many speculations about the underlying reason for this overall visual–lead asymmetry, the main one being that observers are tuned toward the natural situation in which lights arrive before sounds on the sense organs (King and Palmer 1985). There will then be a preference for vision to have a head start over sound so as to be perceived as simultaneous.

Although besides this possibility, there are many other reasons why the PSS can differ quite substantially from 0 ms SOA. To point out just a few: the PSS depends, among others, on stimulus intensity (more intense stimuli are processed faster or come to consciousness more quickly (Jaskowski 1999; Neumann and Niepel 2004; Roefs 1963; Sanford 1971; Smith 1933), stimulus duration (Boenke et al. 2009), the nature of the response that participants have to make (e.g., “Which stimulus came first?” vs. “Which stimulus came second?”; see Frey 1990; Shore et al. 2001), individual differences (Boenke et al. 2009; Mollon and Perkins 1996; Stone et al. 2001), and the modality to which attention is directed (Mattes and Ulrich 1998; Schneider and Bavelier 2003; Shore et al. 2001, 2005; Stelmach and Herdman 1991; Zampini et al. 2005c). We do not intend to list all the factors known thus far, but we only pick out the one that has been particularly important in theorizing about perception in general, that is, the role of attention.

9.3.1 ATTENTION AFFECTING PSS: PRIOR ENTRY

A vexing issue in experimental psychology is the idea that attention speeds up sensory processing. Titchener (1908) termed it the “law of prior entry,” implying that attended objects come to consciousness more quickly than unattended ones. Although many of the old studies on prior entry suffered from the fact that they might simply reflect response biases (see Schneider and Bavelier 2003; Shore et al. 2001; Spence et al. 2001; Zampini et al. 2005c for discussions on the role of response bias in prior entry). As an example, observers may, whenever unsure, just respond that the attended stimulus was presented first without really having that impression. This strategy would reflect a change in decision criterion rather than a low-level sensory interaction between attention and the attended target stimulus. To disentangle response biases from truly perceptual effects, Spence et al. (2001) performed a series of important TOJ experiments in which visual–tactile, visual–visual, or tactile–tactile stimulus pairs were presented from the left or right of fixation. The focus of attention was directed toward either the visual or tactile modality by varying the probability of each stimulus modality (e.g., in the attend–touch condition, there were 50% tactile–tactile pairs, 0% visual–visual, and 50% critical tactile–visual pairs). Participants had to indicate whether the left or right stimulus was presented first. The idea tested was that attention to one sensory modality would speed up perception of stimuli in that modality, thus resulting in a change of the PSS (see also Mattes and Ulrich 1998; Schneider and Bavelier 2003; Shore et al. 2001, 2005; Stelmach and Herdman 1991; Zampini et al. 2005c). Their results indeed supported this notion: when attention was directed to touch, visual stimuli had to lead by much greater intervals (155 ms) than when attention was directed to vision (22 ms) for them to be perceived as simultaneous. Additional experiments demonstrated that attending to one side (left or right) also speeded perception of stimuli presented at that side. Therefore, both spatial attention and attention to modality were effective in shifting the PSS, presumably because they speeded up perceptual processes. To minimize the contribution of any simple response bias on the PSS, Spence et al. (2001) performed these experiments in which attention was manipulated in a dimension (modality or side) that was orthogonal to that of responding (side or modality, respectively). Thus, while attending to vision or touch, participants had to judge which side came first; and while attending to the left or right, participants judged which modality came first. The authors reported similar shifts of the PSS in these different tasks, thus favoring a perceptual basis for prior entry.

Besides such behavioral data, there is also extensive electrophysiological support for the idea that attention affects perceptual processing. Very briefly, in the electroencephalogram (EEG) one can measure the event-related response (ERP) of stimuli that were either attended or unattended. Naïvely speaking, if attention speeds up stimulus processing, one would expect ERPs of attended

stimuli to be faster than unattended ones. In a seminal study by Hillyard and Munte (1984), participants were presented a stream of brief flashes and tones on the left or right of fixation. The participant's task was to attend either the auditory or visual modality, and to respond to infrequent targets in that modality at an attended location (e.g., respond to a slightly longer tone on the left). The attended modality was constant during the experiment (but varied between subjects), and the relevant location was specified at the beginning of each block of trials. The authors found enhanced negativity in the ERP for stimuli at attended locations if compared to nonattended locations. The negativity started at about 150 ms poststimulus for visual stimuli and at about 100 ms for auditory stimuli. Evidence for a cross-modal link in spatial attention was also found, as the enhancement (although smaller) was also found for stimuli at the attended location in the unattended modality (see also Spence and Driver 1996; Spence et al. 2000 for behavioral results). Since then, analogous results have been found by many others. For example, Eimer and Schröger (1998) found similar results using a different design in which the side of the attended location varied from trial to trial. Again, their results demonstrated enhanced negativities (between 160 and 280 ms after stimulus onset) for attended locations as compared to unattended locations, and the effect was again bigger for the relevant rather than irrelevant modality.

Although going back to the original question, the critical issue for prior entry is whether these ERP effects reflect that attended stimuli are processed *faster*. In most EEG studies, attention affects the amplitude of the ERP rather than speed (for a review, see Eimer and Driver 2001). The problem is that there are many other interpretations for an amplitude modulation rather than increased processing speed (e.g., less smearing of the EEG signal over trials if attended). A shift in the latencies of the ERP would have been easier to interpret in terms of increased processing speed, but the problem is that even if a latency shift in the ERP is obtained, it is usually small if compared to the behavioral data. As an example, in an ERP study by Vibell et al. (2007), attention was directed toward the visual or tactile modality in a visual–tactile TOJ task. Results showed that the peak latency of the visual evoked potentials (P1 and N1) was earlier when attention was directed to vision (P1 = 147 ms, and N1 = 198 ms) rather than when directed to touch (P1 = 151 ms, and N1 = 201 ms). This shift in the P1 may be taken as evidence that attention indeed speeds up perception in the attended modality, but it should also be noted that the 4-ms shift in the ERP is in a quite different order of magnitude than the 38 ms shift of the PSS in the behavioral data, or the 133 ms shift reported by Spence et al. (2001) in a similar study. In conclusion, there is both behavioral and electrophysiological support for the idea that attention speeds up perceptual processing, but the underlying neural mechanisms remain, for the time being, elusive.

9.4 SENSITIVITY FOR INTERSENSORY ASYNCHRONY

Besides the point at which simultaneity is perceived to be maximal (the PSS), the second measure that one can derive from the TOJ and SJ task—but which is unfortunately not always reported—is the observers' sensitivity to timing differences, the JND. The sensitivity to intersensory timing differences is not only of interest for theoretical reasons, but it is also of practical importance, for example, in video broadcasting or multimedia Internet where standards are required for allowable audio or video delays (Finger and Davis 2001; Mortlock et al. 1997; Rihs 1995). One of the classic studies on sensitivity for intersensory synchrony was done by Hirsh and Sherrick (1961). They presented audio–visual, visual–tactile, and audio–tactile stimuli in a TOJ task and reported JNDs to be approximately 20 ms regardless of the modalities used. Although more recent studies have found substantially bigger JNDs and larger differences between the sensory modalities. For simple cross-modal stimuli such as auditory beeps and visual flashes, JNDs have been reported in the order of approximately 25 to 50 ms (Keetels and Vroomen 2005; Zampini et al. 2003a, 2005b), but for audio–tactile pairs, Zampini et al. (2005a) obtained JNDs of about 80 ms, and for visual–tactile pairs, JNDs have been found in the order of 35 to 65 ms (Keetels and Vroomen 2008b; Spence et al. 2001). More importantly, JNDs are not constant, but have been shown to depend on various

other factors like the spatial separation between the components of the stimuli, stimulus complexity, whether it is speech or not, and—more controversial—the semantic congruency. Some of these factors will be described below.

9.4.1 SPATIAL DISPARITY AFFECTS JND

A factor that has been shown to affect sensitivity for intersensory timing is the spatial separation between the components of a stimulus pair. Typically, sensitivity for temporal order improves if the components of the cross-modal stimuli are spatially separated (i.e., lower JNDs; Bertelson and Aschersleben 2003; Spence et al. 2003; Zampini et al. 2003a, 2003b, 2005b). Bertelson and Aschersleben, for example, reported audiovisual JNDs to be lower when a beep and a flash were presented from different locations rather than from a common and central location. Zampini et al. (2003b) qualified these findings and observed that sensitivity in an audiovisual TOJ task improved if the sounds and lights were presented from different locations, but only so if presented at the left and right from the median (at 24°). No effect of separation was found for vertically separated stimuli. This made Zampini et al. conclude that the critical factor for the TOJ improvement was that the individual components of an audiovisual stimulus were presented in different hemifields. Keetels and Vroomen (2005), though, examined this notion and varied the (horizontal) size of the spatial disparity. Their results showed that JNDs also improved when spatial disparity was large rather than small, even if stimuli did not cross hemifields. Audiovisual JNDs thus depend on both the relative position from which stimuli are presented and on whether hemifields are crossed or not. Spence et al. (2001) further demonstrated that sensitivity improves for spatially separated visual–tactile stimulus pairs, although no such effect was found for audio–tactile pairs (Zampini et al. 2005a). In blind people, on the other hand, audio–tactile temporal sensitivity was found to be affected by spatial separation (Occelli et al. 2008) and similar spatial modulation effects were demonstrated in rear space (Kitagawa 2005).

What is the underlying reason that sensitivity to temporal differences improves if the sources are spatially separated? Or, why does the brain fail to notice temporal intervals when stimuli comes from a single location? Two accounts have been proposed (Spence et al. 2003). First, it has been suggested that *intersensory pairing* impairs sensitivity for temporal order. The idea underlying “intersensory pairing” is that the brain has a list of criteria on which it decides whether information from different modalities belong together or not. Commonality in time is, without a doubt, a very important criterion, but there may be others like commonality in space, association based on co-occurrence, or semantic congruency. Stimuli from the same location may, for this reason, be more likely paired into a single multimodal event if compared to stimuli presented far apart (see Radeau 1994). Any such tendency to pair stimuli could then make the relative temporal order of the components lost, thereby worsening the temporal sensitivity in TOJ or SJ tasks.

Although in contrast with this notion, many cross-modal effects occur despite spatial discordance, and there are reasons to argue that spatial congruency may *not* be an important criterion for intersensory pairing (Bertelson 1994; Colin et al. 2001; Jones and Munhall 1997; Keetels et al. 2007; Keetels and Vroomen 2007, 2008a; Stein et al. 1996; Teder-Salejarvi et al. 2005; Vroomen and Keetels 2006). But why, then, does sensitivity for temporal order improve with spatially separated stimuli if not because intersensory pairing is impeded? A second reason why JNDs may improve is that of *spatial redundancy*. Whenever multisensory information is presented from different locations, observers actually have extra spatial information on which to base their response. That is, observers may initially not know which modality had been presented first, but still know on which side the first stimulus appeared, and they may then infer which modality had been presented first. As an example, in an audiovisual TOJ task, an observer may have noticed that the first stimulus came from the left (possibly because attention was captured by the first stimulus toward that side). They may also remember that the light was presented on the right. By inference, then, the sound must have been presented first. Sensitivity for temporal order for spatially separated stimuli then improves because there are extra spatial cues that are not present for colocated stimuli.

Not found in the Reference list; please provide the details for this citation.

Not found in the Reference list; please provide the details for this citation.

9.4.2 “STIMULUS COMPLEXITY” AFFECTS JND

Many studies exploring temporal sensitivity have used relatively simple stimuli such as flashes and beeps that have a single and rather sharp transient onset. However, in real-world situations, the brain has to deal with much more complex stimuli that often have complicated variations in temporal structure over time (e.g., seeing and hearing someone speaking; or seeing, hearing, and touching the keys on a computer keyboard). How does the brain notice timing differences between these more complicated and dynamic stimuli? Theoretically, one might expect that more complex stimuli also provide a richer base on which to judge temporal order. Audiovisual speech would be the example “par excellence” because it is rich in content and fluctuating over time. Although in fact, several studies have found the opposite, and in particular for audiovisual speech, the “temporal window” for which the auditory and visual streams are perceived as synchronous is rather wide (Conrey and Pisoni 2006; Dixon and Spitz 1980; Jones and Jarick 2006; Stekelenburg and Vroomen 2007; a series of studies by Vatakis et al.; van Wassenhove et al. 2007). For example, in a study by van Wassenhove et al. (2007), observers judged in an SJ task whether congruent audiovisual speech stimuli and incongruent McGurk-like speech stimuli* (McGurk and MacDonald 1976) were synchronous or not. The authors found a temporal window of 203 ms for the congruent pairs (ranging from –76 ms sound-first to +127 ms vision-first, with PSS at 26 ms vision-first) and a 159 ms window for the incongruent pairs (ranging from –40 to +119 ms, with PSS at 40 ms vision-first). These windows are rather wide if compared to the much smaller windows found for simple flashes and beeps (mostly below 50 ms; Hirsh and Sherrick 1961; Keetels and Vroomen 2005; Zampini et al. 2003a, 2005b). The relatively wide temporal window for complex stimuli has also been demonstrated by indirect tests. For example, the McGurk effect was found to diminish if the auditory and visual information streams are out of sync, but this only occurred at rather long intervals (comparable with the ones found in SJ tasks; Grant et al. 2004; Massaro et al. 1996; McGrath and Summerfield 1985; Munhall et al. 1996; Pandey et al. 1986; Tanaka et al. 2009b; van Wassenhove et al. 2007).

There have been several recent attempts to compare sensitivity for intersensory timing in audiovisual speech with other audiovisual events such as music (guitar and piano) and object actions (e.g., smashing a television set with a hammer, or hitting a soda can with a block of wood; Vatakis and Spence 2006a, 2006b). Observers made TOJs about which stream (auditory or visual) appeared first. Overall, results showed better temporal sensitivity for audiovisual stimuli of “lower complexity” in comparison with stimuli having continuously varying properties (i.e., syllables vs. words and/or sentences). Similar findings were reported by Stekelenburg and Vroomen (2007), who compared JNDs of audiovisual speech (pronunciation of the syllable /bi/) with that of natural nonspeech events (a video of a handclap) in a TOJ task. Again, JNDs were much better for the nonspeech events (64 ms) than for speech (105 ms).

On the basis of these findings, some have concluded that “speech is special” (van Wassenhove et al. 2007; Vatakis et al. 2008a) or that when “stimulus complexity” increases, sensitivity for temporal order deteriorates (Vatakis and Spence 2006a). Although in our view, these proposals do not really clarify the issue because the notion of “speech is special” and “stimulus complexity” are both ill-defined, and most likely, these concepts are confounded with other stimulus factors that can be described more clearly. As an example, it is known that the rate at which stimuli are presented affects audiovisual JNDs for intersensory timing (Benjamins et al. 2008; Fujisaki and Nishida 2005). Sensitivity may also be affected by whether there is anticipatory information that predicts the onset of an audiovisual event (Stekelenburg and Vroomen 2007; Van Eijk 2008; Vroomen and Stekelenburg 2009), and by whether there is a sharp transition that can serve as a temporal anchor

* In the McGurk illusion (McGurk and MacDonald 1976), it is shown that the perception of nonambiguous speech tokens can be modified by the simultaneous presentation of visually incongruent articulatory gestures. Typically, when presented with an auditory syllable /ba/ dubbed onto a face articulating /ga/, participants report hearing /da/. The occurrence of this so-called McGurk effect has been taken as a particularly powerful demonstration of the use of visual information in speech perception.

(Fujisaki and Nishida 2005). Each of these stimulus characteristics—and likely many others—need to be controlled if one wants to compare across stimuli in a nonarbitrary way. Below, we address some of these factors.

9.4.3 STIMULUS RATE AFFECTS JND

It has been demonstrated that perception of intersensory synchrony breaks down if stimuli are presented with a temporal frequency of above ~4Hz. This is very slow if compared to unimodal visual or auditory sensitivity for temporal coherence. Fujisaki and Nishida (2005) examined this using audiovisual stimuli consisting of a luminance-modulated Gaussian blob and an amplitude-modulated white noise presented at various rates. They demonstrated that synchrony–asynchrony discrimination for temporally dense random pulse trains became nearly impossible at temporal frequencies larger than 4 Hz, even when the audiovisual interval was large enough for discrimination of single pulses (the discrimination thresholds were 75, 81, and 119 ms for single pulses, 2 and 4 Hz repetitive stimuli, respectively). This 4-Hz boundary was also reported by Benjamins et al. (2008). They explored the temporal limit of audiovisual integration using a visual stimulus that alternated in color (red or green) and a sound that alternated in frequency (high or low). Observers had to indicate which sound (high or low) accompanied the red disk. Their results demonstrated that at rates of 4.2 Hz and higher, observers were no longer able to match the visual and auditory stimuli across modalities (proportion correct matches dropped from 0.9 at 1.9 Hz to 0.5 at a 4.2 Hz). Further experiments also demonstrated that manipulating other temporal stimulus characteristics such as the stimulus offsets and/or audiovisual SOAs did not change the 4-Hz threshold. Here, it should be mentioned that the 4-Hz rate is also the approximate rate with which syllables are spoken in continuous speech, and temporal order in audiovisual speech might thus be difficult simply because stimulus presentation is too fast, and not because speech is special*.

Fujisaki et al. (2005) was changed to Fujisaki and Nishida (2005); please verify if this is correct.

9.4.4 PREDICTABILITY AFFECTS JND

Another factor that may play a role in intersensory synchrony judgments, but one that has not yet been studied extensively, is the extent to which (one of the components of) a multisensory event can be predicted. As an example, for many natural events—such as the clapping of hands—vision provides predictive information about when a sound is to occur, as there is visual anticipatory information about sound onset. Stimuli with predictive information allow observers to make a clear prediction about when a sound is to occur, and this might improve sensitivity for temporal order. A study by van Eijk et al. (2008, Chapter 4) is of relevance here. They explored the effect of visual predictive information (or, the way the authors called it, “apparent causality”) on perceived audiovisual synchrony. Visual predictive information was either present or absent by showing all or part of a Newton’s cradle toy (i.e., a ball that appears to fall from a suspended position on the left of the display, strikes the leftmost of four contiguous balls, and then launches the rightmost ball into an arc motion away from the other balls). The collision of the balls was accompanied by a sound that varied around the time of the impact. The predictability of the sound was varied by showing either the left side of the display (motion followed by a collision and sound so that visual motion predicted sound occurrence) or the right side of the display (a sound followed by visual motion; so no predictable information about sound onset). In line with the argument made here, the authors reported

* It has also been reported that the presentation rate may shift the PSS. In a study by Arrighi et al. (2006), participants were presented a video of hands drumming on a conga at various rates (1, 2, and 4 Hz). Observers were asked to judge whether the auditory and visual streams appeared to be synchronous or not (an SJ task). Results showed that the auditory delay for maximum simultaneity (the PSS) varied inversely with drumming tempo from about 80 ms at 1 Hz, and 60 ms at 2 Hz, to 40 ms at 4 Hz. Video sequences of random drumming motion and of a disk moving along the motion profile matching the hands of to the drummer produced similar results, with higher tempos requiring less auditory delay.

better temporal sensitivity if visual predictive information about sound onset was available (the left display) rather than if it was absent (the right display).

9.4.5 DOES INTERSENSORY PAIRING AFFECT JND?

A more controversial issue in the literature on intersensory timing is the extent to which information from different modalities is treated by the brain as belonging to the same event. Some have headed it under the already mentioned notion of “intersensory pairing,” others under the “unity assumption” (Welch and Warren 1980). The idea is that observers find it difficult to judge temporal order if the information streams naturally belong together, for reasons other than temporal coincidence, because there is then more intersensory integration; in which case, temporal order is lost. Several studies have examined this issue but with varying outcomes. In a study by Vatakis and Spence (2007), participants judged the temporal order of audiovisual speech stimuli that varied in gender and phonemic congruency. Face and voice congruency could vary in gender (a female face articulating /pi/ with a sound of either a female or male /pi/), or phonemic content (a face saying /ba/ with a voice saying /ba/ or /da/). In support of the unity assumption, results showed that for both the gender and phonemic congruency manipulation, sensitivity for temporal order improved if the auditory and visual streams were incongruent rather than congruent. In a recent study, Vatakis et al. (2008a) qualified these findings and reported that this effect may be specific for human speech. In this study, the effect of congruency was examined using matching or mismatching call types of monkeys (“cooing” vs. “grunt” or threat calls). For audiovisual speech, the sensitivity of temporal order was again better for the incongruent rather than congruent trials, but there was no congruency effect for the monkey calls. In another study, Vatakis and Spence (2008) also found no congruency effect for audiovisual music and object events that either matched (e.g., the sight of a note being played on a piano together with the corresponding sound, or the video of a bouncing ball with a corresponding sound) or mismatched. At this stage, it therefore appears that the “unity assumption” may only apply to audiovisual speech. It leaves one wonder, though, whether this effect is best explained in terms of the “special” nature of audiovisual speech, or whether other factors are at play (e.g., the high level of exposure to speech stimuli in daily life, the possibly more attention-grabbing nature of speech stimuli, or the specific low-level acoustic stimulus features of speech; Vatakis et al. 2008a).

Please check if the changes made here are OK.

Possibly change to “This leaves one question though, whether this...” instead?

9.5 HOW THE BRAIN DEALS WITH LAGS BETWEEN THE SENSES

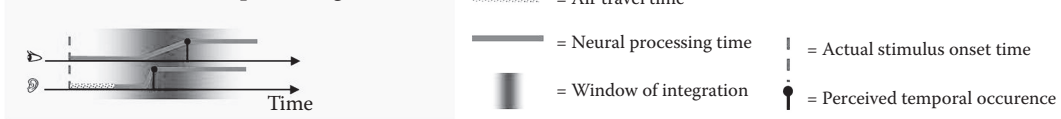
In any multisensory environment, the brain has to deal with lags in arrival and processing time between the different senses. Surprisingly though, despite these lags, temporal coherence is usually maintained, and only in exceptional circumstances such as the thunder, which is heard *after* the lightning, a single multisensory event is perceived as being separated. This raises the question of how temporal coherence is maintained. In our view, at least four options are available: (1) the brain might be insensitive for small lags, or it could just ignore them (a window of temporal integration); (2) the brain might be “intelligent” and bring deeply rooted knowledge about the external world into play that allows it to compensate for various external factors; (3) the brain might be flexible and shift its criterion about synchrony in an adaptive fashion (recalibration); or (4) the brain might actively shift the time at which one information stream is perceived to occur toward the other (temporal ventriloquism). Below, we discuss each of these notions. It should be noted beforehand that none of these options mutually excludes the other.

9.5.1 WINDOW OF TEMPORAL INTEGRATION

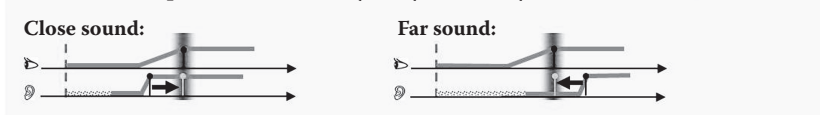
The first notion, that the brain is rather insensitive to lags, comes close to the idea that there is a “window of temporal integration.” Any information that falls within this hypothetical window is potentially assigned to the same external event and streams within the window are then treated as to

have occurred simultaneously (see Figure 9.2, panel 1). Many have alluded to this concept, but what is less satisfying about it is that it is basically a description rather than an explanation. To make this point clear, some have reported that the temporal window for audiovisual speech can be quite large because it can range from approximately 40 ms audio-first to 240 ms vision-first. However, sensitivity for intersensory asynchronies (JND) is usually much smaller than the size of this window. For example, Munhall et al. (1996) demonstrated that exact temporal coincidence between the auditory and visual parts of audiovisual speech stimuli is not a very strict constraint on the McGurk effect (McGurk and MacDonald 1976). Their results demonstrated that the McGurk effect was biggest when vowels were synchronized (see also McGrath and Summerfield 1985), but the effect survived even if audition lagged vision by 180 ms (see also Soto-Faraco and Alsius 2007, 2009; these studies

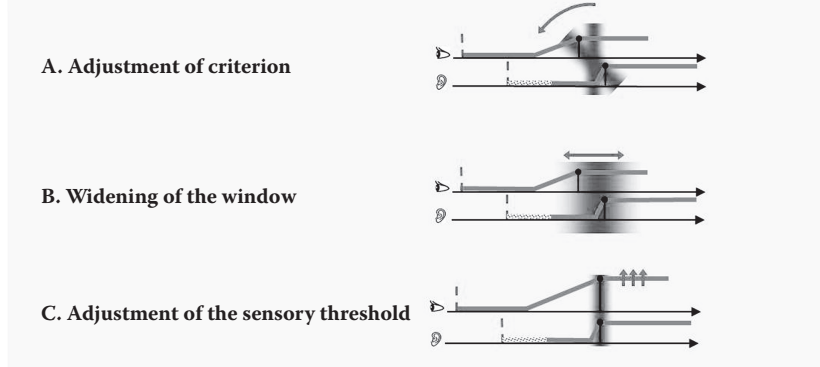
1) A wide window of temporal integration



2) The brain compensates for auditory delays caused by sound distance



3) Adaptation to intersensory asynchrony via:



4) Temporal ventriloquism: The perceived visual onset time is shifted towards audition



FIGURE 9.2 Synchrony can be perceived despite lags. How is this accomplished? Four possible mechanisms are depicted for audiovisual stimuli like a flash and beep. Similar mechanisms might apply for other stimuli and other modality pairings. Time is represented on *x*-axis, and accumulation of sensory evidence on *y*-axis. A stimulus is time-stamped once it surpasses a sensory threshold. Stimuli in audition and vision are perceived as being synchronous if they occur within a certain time window. (1) The brain might be insensitive for naturally occurring lags because window of temporal integration is rather wide. (2) The brain might compensate for predictable variability—here, sound distance—by adjusting perceived occurrence of a sound in accordance with sound travel time. (3) Temporal recalibration. Three different mechanisms might underlie adaptation to asynchrony: (a) a shift in criterion about synchrony for adapted stimuli or modalities, (b) a widening of temporal window for adapted stimuli or modalities, and (c) a change in threshold of sensory detection (when did the stimulus occur?) within one of adapted modalities. (4) Temporal ventriloquism: a visual event is actively shifted toward an auditory event.

show that participants can still perceive a McGurk effect when they can quite reliably perform TOJs). Outside the speech domain, similar findings have been reported. In a study by Shimojo et al. (2001), the role of temporal synchrony was examined using the *streaming–bouncing* illusion (i.e., two identical visual targets that move across each other and are normally perceived as a streaming motion are typically perceived to bounce when a brief sound is presented at the moment that the visual targets coincide; Sekuler et al. 1997). The phenomenon is dependent on the timing of the sound relative to the coincidence of the moving objects. Although it has been demonstrated that a brief sound induced the visual bouncing percept most effectively when it was presented about 50 ms before the moving objects coincide, their data furthermore showed a rather large temporal window of integration because intervals ranging from 250 ms before visual coincidence to 150 ms after coincidence still induced the bouncing percept (see also Bertelson and Aschersleben 1998, for the effect of temporal asynchrony on spatial ventriloquism; or Shams et al. 2002, for the illusory-flash effect). All these intersensory effects thus occur at asynchronies that are much larger than JNDs normally reported when directly exploring the effect of asynchrony using TOJ or SJ tasks (van Wassenhove et al. 2007). One might argue that despite the fact that observers do notice small delays between the senses, the brain can still ignore it if it is of help for other purposes, such as understanding speech (Soto-Faraco and Alsius 2007, 2009). But the question then becomes, why is there more than one window; that is, one for understanding, the other for noticing timing differences.

Besides the width of the temporal window varying with the purpose of the task, it has also been found to vary for different kinds of stimuli. As already mentioned, the temporal window is much smaller for clicks and flashes than it is for audiovisual speech. However, why would the size be different for different stimuli? Does the brain have a separate window for each stimulus and each purpose? If so, we are left with explaining how and why it varies. Some have taken the concept of a window quite literally, and have argued that “speech is special” because the window for audiovisual speech is wide (van Wassenhove et al. 2007; Vatakis et al. 2008a). Although we would rather refrain from such speculations, and consider it more useful to examine what the critical features are that determine when perception of simultaneity becomes easy (a small window) or difficult (a large window). The size of the window is thus, in our view, the factor that needs to be explained rather than that it is the explanation itself.

9.5.2 COMPENSATION FOR EXTERNAL FACTORS

The second possibility—the intelligent brain that compensates for various delays—is a controversial issue that has received support mainly from studies that examined whether observers take distance into account when judging audiovisual synchrony (see Figure 9.2, panel 2). The relatively slow transduction time of sounds through air causes natural differences in arrival time between sounds and lights. It implies that the farther away an audiovisual event, the more the sound will lag the visual stimulus; although such a lag might be compensated for by the brain if distance were known. The brain might then treat a lagging sound as being synchronous to a light, provided that the audiovisual event occurred at the right distance. Some have indeed reported that the brain does just that as judgments about audiovisual synchrony were found to depend on perceived distance (Alais and Carlile 2005; Engel and Dougherty 1971; Heron et al. 2007; Kopinska and Harris 2004). Although others have failed to demonstrate compensation for distance (Arnold et al. 2005; Lewald and Guski 2004).

Sugita and Suzuki (2003) explored compensation for distance with an audiovisual TOJ task. The visual stimuli were delivered by light-emitting diodes (LEDs) at distances ranging from 1 to 50 m in free-field circumstances (and were compensated for by intensity, although not size). Of importance, the sounds were delivered through headphones, and no attempt was made to equate the distance of the sound with that of the light. Note that this, in essence, undermines the whole idea that the brain compensates for lags of audiovisual events out in space. Nevertheless, PSS values were found to shift with visual stimulus distance. When the visual stimulus was 1 m away, the PSS was at about a

Please verify
if the changes
made here are
OK.

~5 ms sound delay, and the delay increased when the LEDs were farther away. The increment was consistent with the velocity of sounds up to a viewing distance of about 10 m, after which it leveled off. This led the authors to conclude that lags between auditory and visual inputs are perceived as synchronous not because the brain has a wide temporal window for audiovisual integration, but because the brain actively changes the temporal location of the window depending on the distance of the source.

Alais and Carlile (2005) came to similar conclusions, but with different stimuli. In their study, auditory stimuli were presented over a loudspeaker and auditory distance was simulated by varying the direct-to-reverberant energy ratio as a depth cue for sounds (Bronkhorst 1995; Bronkhorst and Houtgast 1999). The near sounds simulated a depth of 5 m and had substantial amounts of direct energy with a sharp transient onset; the far sounds simulated a depth of 40 m and did not have a transient. The visual stimulus was a Gaussian blob on a computer screen in front of the observer without variations in distance. Note that, again, no attempt was made to equate auditory and visual distance, thus again undermining the underlying notion. The effect of apparent auditory distance on temporal alignment with the blob on the screen was measured in a TOJ task. The authors found compensation for depth, thus the PSS in the audiovisual TOJ task shifted with the apparent distance of the sound in accordance with the speed of sounds through air up to 40 m. Although on closer inspection of their data, it is clear that the shift in the PSS was mainly caused by the fact that sensitivity for intersensory synchrony became increasingly worse for more distant sounds. Judging from their figures, sensitivity for nearby sounds at 5 m was in the normal range, but for the most distant sound, sensitivity was extremely poor as it never reached plateau, and even at a sound delay of 200 ms, 25% of the responses was still “auditory-first” (see also Arnold et al. 2005; Lewald and Guski 2004). This suggests that observers, while performing the audiovisual TOJ task, could not use the onset of the far sound as a cue for temporal order, possibly because it lacks a sharp transient and that they had to rely on other cues instead. Besides controversial stimuli and data, there are others who simply failed to observe compensation for distance (Arnold et al. 2005; Heron et al. 2007; Lewald and Guski 2004; Stone et al. 2001). For example, Stone et al. (2001) used an audiovisual SJ task and varied stimulus–observer distances from 0.5 m in the near condition to 3.5 m in the far condition. This resulted in a 3-m difference that would theoretically correspond to an 11 ms difference in the PSS if sound–travel time would not be compensated (sound velocity of 330 m/s corresponds to 3.5 m/11 ms). For three out of five subjects, the PSS values were indeed shifted in that direction, which led the authors to conclude that distance was not compensated. Against this conclusion, it should be said that the SJ tasks depend heavily on criterion settings, that “three-out-of-five” is not persuasively above chance, and that the range of distances was rather restricted.

Please verify if the data here was captured correctly.

Please verify if the changes made are OK.

Less open to these kinds of criticisms is a study by Lewald and Guski (2004). They used a rather wide range of distances (1, 5, 10, 20, and 50 m), and their audiovisual stimuli (a sequence of five beeps/flashes) were delivered by colocated speakers/LEDs placed in the open field. Note that in this case, there were no violations in the “naturalness” of the audiovisual stimuli and that they were physically colocated. Using this setup, the authors did *not* observe compensation for distance. Rather, their results showed that when the physical observer–stimulus distance increased, the PSS shifted precisely with the variation in sound transmission time through air. For audiovisual stimuli that are far away, sounds thus had to be presented earlier than for nearby stimuli to be perceived as simultaneous, and there was no sign that the brain would compensate for sound–traveling time. The authors also suggested that the discrepancy between their findings and those who did find compensation for distance lies in the fact that the latter *simulated* distance rather than using the natural situation.

Please check if the changes made here are OK.

Similar conclusions were also reached by Arnold et al. (2005), who examined whether the stream/bounce illusion (Sekuler et al. 1997) varies with distance. The authors examined whether the optimal time to produce a “bounce” percept varied with the distance of the display, which ranged from ~1 to ~15 m. The visual stimuli were presented on a computer monitor—keeping retinal properties constant—and the sounds were presented either over loudspeakers at these distances or over

headphones. The optimal time to induce a bounce percept shifted with the distance of the sound if they were presented over loudspeakers, but there was no shift if the sound was presented over headphones. Similar effects of timing shifts with viewing distance after loudspeaker, but not headphone, presentation were obtained in an audiovisual TOJ task in which observers judged whether a sound came before or after two disks collided. This led the authors to conclude that there is *no* compensation for distance if distance is real and presented over speakers rather than simulated and presented over headphones.

This conclusion might well be correct, but it raises the question of how to account for the findings by Kopinska and Harris (2004). These authors reported *complete* compensation for distance despite using colocated sounds and lights produced at natural distances. In their study, the audiovisual stimulus was a bright disk that flashed once on a computer monitor and it was accompanied by a tone burst presented from the computer's inbuilt speaker. Participants were seated at various distances from the screen (1, 4, 8, 16, 24, and 32 m) and made TOJs about the flash and the sound. The authors also selectively slowed down visual processing by presenting the visual stimulus at 20° of eccentricity rather than in the fovea, or by having observers wear darkened glasses. As an additional control, they used simple reaction time tasks and found that all these variations—distance, eccentricity, and dark glasses—had predictable effects on auditory or visual speeded reaction. However, audiovisual simultaneity was *not* affected by distance, eccentricity, or darkened glasses. Thus, there was no shift in the PSS despite the fact that the change in distance, illumination, and retinal location affected simple reaction times. This made the authors conclude that observers recover the external world by taking into account all kinds of predictable variations, most importantly distance, alluding to similar phenomena such as size or color constancy.

There are some studies that varied audiovisual distance in a natural way, but came to diametrically opposing conclusions: Lewald and Guski (2004) and Arnold et al. (2005) found no compensation for distance, whereas Kopinska and Harris (2004) reported complete compensation. What's the critical difference between them? Our conjecture is that they differ in two critical aspects, that is, (1) whether distance was randomized on a trial-by-trial basis or blocked, and (2) whether sensitivity for temporal order was good or poor. In the study by Lewald and Guski, the distance of the stimuli was varied on a trial-by-trial basis as they used a setup of five different speakers/LEDs. In Kopinska and Harris's study, though, the distance between the observer and the screen was blocked over trials because otherwise subjects would have to be shifted back and forth after each trial. If the distance is blocked, then either adaptation to the additional sound lag may occur (i.e., recalibration), or subjects may equate response probabilities to the particular distance that they are seated. Either way, the effect of distance on the PSS will diminish if trials are blocked, and no shift in the PSS will then be observed, leading to the "wrong" conclusion that distance is compensated. This line of reasoning corresponds with a recent study by Heron et al. (2007). In their study, participants performed a TOJ task in which audiovisual stimuli (a white disk and a click) were presented at varying distances (0, 5, 10, 20, 30, and 40 m). Evidence for compensation was only found after a period of adaptation (1 min + 5 top-up adaptation stimuli between trials) to the naturally occurring audiovisual asynchrony associated with a particular viewing distance. No perceptual compensation for distance-induced auditory delays could be demonstrated whenever there was no adaptation period (although we should notice that in the present study, observer distance was always blocked).

The second potentially relevant difference between studies that do or do not demonstrate compensation is the difficulty of the stimuli. Lewald and Guski (2004) used a sequence of five pulses/sounds, whereas Kopinska and Harris (2004) presented a single sound/flash. In our experience, a sequence of pulses/flashes drastically improves accuracy for temporal order if compared to a single pulse/flash because there are many more cues in the signal. In the study by Arnold et al. (2005), judgments about temporal order could also be relatively accurate because the two colliding disks provided anticipatory information about when to expect the sound. Most likely, observers in the study of Kopinska and Harris were inaccurate because their single sound/flash stimuli without anticipatory information were difficult (unfortunately, none of the studies reported JNDs). In effect,

"darkening glasses" changed to "darkened glasses." Please verify if this is correct.

Possibly use the term "comprehend" instead of "recover" here?

The term "at present" was removed here; the studies cited are from 2004–2005.

this amounts to adding noise to the psychometric function, which then effectively masks the effect of distance on temporal order. It might easily lead one to conclude “falsely” that there is compensation for distance.

9.5.3 TEMPORAL RECALIBRATION

The third possibility of how the brain might deal with lags between the senses entails that the brain is flexible in adopting what it counts as synchronous (see Figure 9.2, panel 3). This phenomenon is also known as “temporal recalibration.” Recalibration is a well-known phenomenon in the spatial domain, but it has only recently been demonstrated in the temporal domain (Fujisaki et al. 2004; Vroomen et al. 2004). As for the spatial case, more than a century ago, von Helmholtz (1867) had already shown that the visual–motor system was remarkably flexible as it adapts to shifts of the visual field induced by wedge prisms. If prism-wearing subjects had to pick up a visually displaced object, they would quickly adapt to the new sensor–motor arrangement and even after only a few trials, small visual displacements might get unnoticed. Recalibration was the term used to explain this phenomenon. In essence, recalibration is thought to be driven by a tendency of the brain to minimize discrepancies between the senses about objects or events that normally belong together. For the prism case, it is the position of where the hand is seen and felt. Nowadays, it is also known that the least reliable source is adjusted toward the more reliable one (Ernst and Banks 2002; Ernst et al. 2000; Ernst and Bulthoff 2004).

The first evidence of recalibration in the temporal domain came from two studies with very similar designs: an exposure–test paradigm. Both Fujisaki et al. (2004) and Vroomen et al. (2004) first exposed observers to a train of sounds and light flashes with a constant but small intersensory interval, and then tested them by using an audiovisual TOJ or SJ task. The idea was that observers would adapt to small audiovisual lags in such a way that the adapted lag is eventually perceived as synchronous. Therefore, after a light-first exposure, light-first trials would be perceived as synchronous, and after a sound-first exposure, a sound-first stimulus would be perceived as synchronous (see Figure 9.3). Both studies indeed observed that the PSS was shifted in the direction of the exposure lag. For example, Vroomen and Keetels exposed subjects for ~3 min to a sequence of sound bursts/light flashes with audiovisual lags of either ± 100 or ± 200 ms (sound-first or light-first). During the test, the PSS was shifted, on average, by 27 and 18 ms (PSS difference between sound-first and light-first) for the SJ and TOJ tasks, respectively. Fujisaki et al. used slightly bigger lags (± 235 ms sound-first or light-first) and found somewhat bigger shifts in the PSS (59 ms shifts of the PSS in SJ and 51 ms in TOJ), but data were, in essence, comparable. Many others have reported similar effects (Asakawa et al. 2009; Di Luca et al. 2007; Hanson et al. 2008; Keetels and Vroomen 2007, 2008b; Navarra et al. 2005, 2007, 2009; Stetson et al. 2006; Sugano et al. 2010; Sugita and Suzuki 2003; Takahashi et al. 2008; Tanaka et al. 2009a; Yamamoto et al. 2008).

The mechanism underlying temporal recalibration, though, remains elusive at this point. One option is that there is a shift in the criterion for simultaneity in the adapted modalities (Figure 9.2, panel 3a). After exposure to light-first pairings, participants may thus change their criterion for audiovisual simultaneity in such a way that light-first stimuli are taken to be simultaneous. On this view, other modality-pairings (e.g., vision–touch) would be unaffected and the change in criterion should then not affect unimodal processing of visual and auditory stimuli presented in isolation. Another strong prediction is that stimuli that were once synchronous, before adaptation, can become asynchronous after adaptation. The most dramatic case of this phenomenon can be found in motor–visual adaptation. In a study by Eagleman and Holcombe (2002), participants were asked to repeatedly tap their finger on a key, and after each key tap, a delayed flash was presented. If the visual flash occurred at an unexpectedly short delay after the tap (or synchronous), it was actually perceived as occurring *before* the tap, an experience that runs against the law of causality.

It may also be the case that one modality (vision, audition, or touch) is “shifted” toward the other, possibly because the sensory threshold for stimulus detection in one of the adapted modalities is

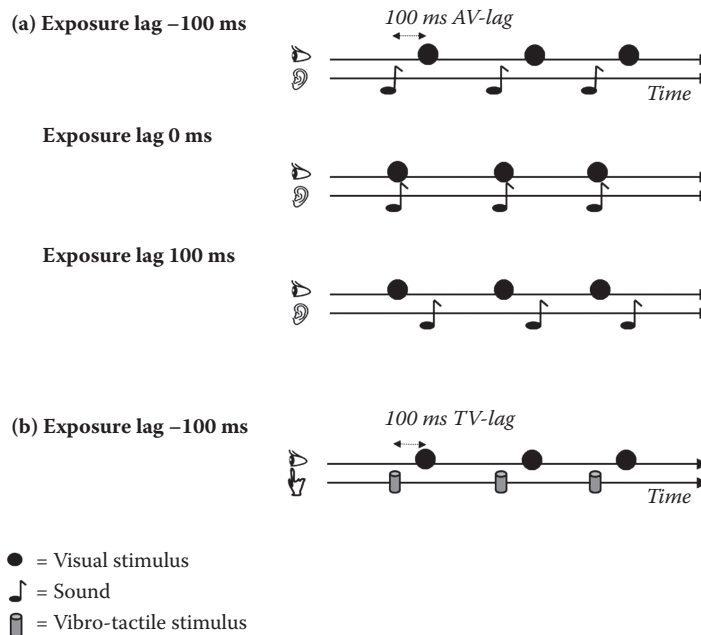


FIGURE 9.3 Schematic illustration of exposure conditions typically used in a temporal recalibration paradigm. During exposure, participants are exposed to a train of auditory–visual (AV) or tactile–visual (TV) stimulus pairs (panels a and b, respectively) with a lag of –100, 0, or +100 ms. To explore possible shifts in perceived simultaneity or sensitivity to asynchrony, typically a TOJ or SJ task is performed in a subsequent test phase. (From Fujisaki et al., *Nat. Neurosci.*, 7, 773–8, 2004; Vroomen et al., *Cogn. Brain Res.*, 22, 32–5, 2004; Keetels, M., Vroomen, J., *Percept. Psychophys.*, 70, 765–71, 2008; Keetels, J., Vroomen, M., *Neurosci. Lett.*, 430, 130–4, 2008. With permission.)

changed (see Figure 9.2, panel 3b). For example, as an attempt to perceive simultaneity during light-first exposure, participants might delay processing time in the visual modality by adopting a more stringent criterion for sensory detection of visual stimuli. After exposure to light-first audiovisual pairings, one might then expect slower processing times of visual stimuli in general, and other modality pairings that involve the visual modality, say vision–touch, would then also be affected.

Two strategies have been undertaken to explore the mechanism underlying temporal recalibration. The first is to examine whether temporal recalibration generalizes to other stimuli within the adapted modalities, the second is to examine whether temporal recalibration affects different modality pairings than the ones adapted. Fujisaki et al. (2004) have already demonstrated that the effect of adaptation in temporal misalignment was effective even when the visual test stimulus was very different from the exposure situation. The authors exposed observers to asynchronous tone–flash stimulus pairs and later tested them on the “stream/bounce” illusion (Sekuler et al. 1997). Fujisaki et al. reported that the optimal delay for obtaining a bounce percept in the stream/bounce illusion was shifted in the same direction as the adapted lag. Furthermore, after exposure to a “wall-display,” in which tones were timed with a ball bouncing off the inner walls of a square, similar shifts in the PSS on the bounce percept were found (a ~45 ms difference when comparing the PSS of the –235 ms sound-first exposure with the +235 ms vision-first exposure). Audiovisual temporal recalibration thus generalized well to other visual stimuli.

Navarra et al. (2005) and Vatakis et al. (2008b) also tested generalization for audiovisual temporal recalibration using stimuli from different domains (speech/nonspeech). Their observers had to monitor a continuous speech stream for target words that were presented either in synchrony with the video of a speaker, or with the audio stream lagging 300 ms behind. During the monitoring

task, participants performed a TOJ (Navarra et al. 2005; Vatakis et al. 2007) or SJ task (Vatakis et al. 2008b) on simple flashes and white noise bursts that were overlaid on the video. Their results showed that *sensitivity*, rather than a shift in the PSS, became worse if subjects were exposed to desynchronized rather than synchronized audiovisual speech. Similar effects (larger JNDs) were found with music stimuli. This led the authors to conclude that the “window of temporal integration” was widened (see Figure 9.2, panel 3c) because of asynchronous exposure (see also Navarra et al. 2007 for effects on JND after adaptation to asynchronous audio–tactile stimuli). The authors argued that this effect on the JND may reflect an initial stage of recalibration in which a more lenient criterion is adopted for simultaneity. With prolonged exposure, subjects may then shift the PSS. An alternative explanation—also considered by the authors, but rejected—might be that subjects became confused by the nonmatching exposure stimuli, which as a result may also affect the JND rather than the PSS because it adds noise to the distribution.

The second way to study the underlying mechanisms of temporal recalibration is to examine whether temporal recalibration generalizes to different modality pairings. Hanson et al. (2008) explored whether a “supramodal” mechanism might be responsible for the recalibration of multi-sensory timing. They examined whether adaptation to audiovisual, audio–tactile, and tactile–visual asynchronies (10 ms flashes, noise bursts, and taps on the left index finger) generalized across modalities. The data showed that a brief period of repeated exposure to ± 90 ms asynchrony in any of these pairings resulted in shifts of about 70 ms of the PSS on subsequent TOJ tasks, and that the size and nature of the shifts were very similar across all three pairings. This made them conclude that there is a “general mechanism.” Opposite conclusions though, were reached by Harrar and Harris (2005). They exposed participants for 5 min to audiovisual pairs with a fixed time lag (250 ms light-first), but did not obtain shifts in the PSSs for touch–light pairs. In an extension of this topic (Harrar and Harris 2008), observers were exposed for 5 min to ~ 100 ms lags of light-first stimuli for the audiovisual case, and touch-first stimuli for the auditory–tactile and visual–tactile case. Participants were tested on each of these pairs before and after exposure. Shifts of the PSS in the predicted direction were only found in the audiovisual exposure–test stimuli, but not for the other cases. Di Luca et al. (2007) also exposed participants to asynchronous audiovisual pairs (~ 200 ms lags of sound-first and light-first) and measured the PSS for audiovisual, audio–tactile, and visual–tactile test stimuli. Besides obtaining a shift in the PSS for audiovisual pairs, the effect was found to generalize to audio–tactile, but not to visual–tactile test pairs. This pattern made the authors conclude that adaptation resulted in a phenomenal shift of the auditory event (Di Luca et al. 2007).

Navarra et al. (2009) also recently reported that the auditory rather than visual modality is more flexible. Participants were exposed to synchronous or asynchronous audiovisual stimuli (224 ms vision-first, or 84 ms auditory-first for 5 min of exposure) after which they performed a speeded reaction time task on unimodal visual or auditory stimuli. In contrast with the idea that visual stimuli get adjusted in time to the relatively more accurate auditory stimuli (Hirsh and Sherrick 1961; Shipley 1964; Welch 1999; Welch and Warren 1980), their results seemed to show the opposite, namely, that auditory rather than visual stimuli were shifted in time. The authors reported that simple reaction times to sounds became approximately 20 ms faster after vision-first exposure and about 20 ms slower after auditory-first exposure, whereas simple reaction times for visual stimuli remained unchanged. They explained this finding by alluding to the idea that visual information can serve as the temporal anchor because it is a more exact estimate of the time of occurrence of a distal event rather than auditory information because light travel time does not depend on distance. Further research is needed, however, to examine whether a change in simple reaction times is truly reflective of a change in the timing of that event, as there is quite some evidence showing that the two do not always go hand-in-hand (e.g., reaction times are more affected by variations in intensity than TOJs; Jaskowski and Verleger 2000; Neumann and Niepel 2004).

To summarize, until now, there is no clear explanation for the mechanism underlying temporal recalibration as there is some discrepancy in the data regarding generalization across modalities. It seems safe to conclude that the audiovisual exposure–test situation is the most reliable one to obtain

a shift in the PSS. Arguably, audiovisual pairs are more flexible because the brain has to correct for timing differences between auditory and visual stimuli because of naturally occurring delays caused by distance. Tactile stimuli might be more rigid in time because visual–tactile and audio–tactile events always occur at the body surface, so less compensation for latency differences might be required here. As already mentioned above, a widening of the JND, rather than a shift in the PSS, has also been observed and it might possibly reflect an initial stage of recalibration in which a more lenient criterion about simultaneity is adopted. The reliability of each modality on its own is also likely to play role. For visual stimuli, it is known that they are less reliable in time than auditory or tactile stimuli (Fain 2003), and as a consequence they may be more malleable (Ernst and Banks 2002; Ernst et al. 2000; Ernst and Bulthoff 2004), but there is also evidence that the auditory modality is, in fact, shifted.

9.5.4 TEMPORAL VENTRILOQUISM

The fourth possibility of how the brain might deal with lags between the senses, and how they may get unnoticed, is that the perceived timing of a stimulus in one modality is actively shifted toward the other (see Figure 9.2, panel 4). This phenomenon is also known as “temporal ventriloquism,” and it is named in analogy with the spatial ventriloquist effect. For spatial ventriloquism, it was already known for a long time that listeners who heard a sound while seeing a spatially displayed flash had the (false) impression that the sound originated from the flash. This phenomenon was named the “ventriloquist illusion” because it was considered a stripped-down version of what the ventriloquist was doing when performing on stage. The temporal ventriloquist effect is analogous to the spatial variant, except that here, sound attracts vision in the time dimension rather than vision attracting sound in the spatial dimension. There are, by now, many demonstrations of this phenomenon, and we describe several in subsequent paragraphs. They all show that small lags between sound and vision go unnoticed because the perceived timing of visual events is flexible and is attracted toward events in other modalities.

Scheier et al. (1999) were one of the first to demonstrate temporal ventriloquism using a visual TOJ task (see Figure 9.4). Observers were presented with two lights at various SOAs, one above and one below a fixation point, and their task was to judge which light came first (the upper or the lower). To induce temporal ventriloquism, Scheier et al. added two sounds that could either be presented before the first and after the second light (condition AVVA), or the sounds could be presented in between the two lights (condition VAAV). Note that they used a *visual* TOJ task, and that sounds were task-irrelevant. The results showed that observers were more sensitive (i.e., smaller intervals were still perceived correctly) in the AVVA condition compared to the VAAV condition (visual JNDs were approximately 24 and 39 ms, respectively). Presumably, the two sounds attracted the temporal occurrence of the two lights, and thus, effectively pulled the lights farther apart in the AVVA condition, and closer together in the VAAV condition. In single-sound conditions AVV and VVA, sensitivity was not different from a visual-only baseline, indicating that the effects were not because of the initial sound acting as a warning signal, or some cognitive factor related to the observer’s awareness of the sounds.

Morein-Zamir et al. (2003) replicated these effects and further explored the sound–light intervals at which the effect occurred. Sound–light intervals of ~100 to ~600 ms were tested, and it was shown that the second sound was mainly responsible for the temporal ventriloquist effect up to a sound–light interval of 200 ms, whereas the interval of the first sound had little effect.

The results were also consistent with earlier findings of Fendrich and Corballis (2001) who used a paradigm in which participants judged when a flash occurred by reporting the clock position of a rotating marker. The repeating flash was seen earlier when it was preceded by a click and later when the click lagged the visual stimulus. Another demonstration of temporal ventriloquism using a different paradigm came from a study by Vroomen and de Gelder (2004b). Here, temporal ventriloquism was demonstrated using the flash-lag effect (FLE). In the typical FLE (Mackay 1958;

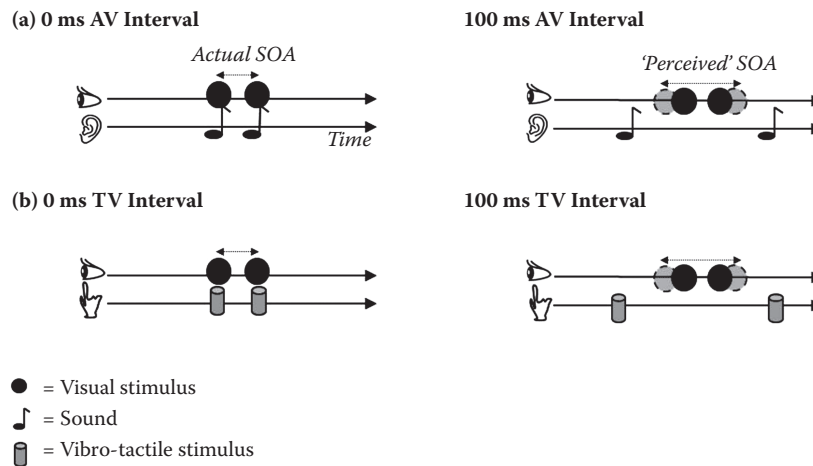


FIGURE 9.4 A schematic illustration of conditions typically used to demonstrate auditory–visual temporal ventriloquism (panel a) and tactile–visual temporal ventriloquism (panel b). The first capturing stimulus (i.e., either a sound or a vibro–tactile stimulus) precedes the first light by 100 ms, whereas the second capturing stimulus trails the second light by 100 ms. Baseline condition consists of presentation of two capturing stimuli simultaneous with light onsets. Temporal ventriloquism is typically shown by improved visual TOJ sensitivity when capture stimuli are presented with a 100-ms interval. (From Scheier et al., *Invest. Ophthalmol. Vis. Sci.*, 40, 4169, 1999; Morein-Zamir et al., *Cogn. Brain Res.*, 17, 154–63, 2003; Vroomen, J., Keetels, M., *J. Exp. Psychol. Hum. Percept. Perform.*, 32, 1063–71, 2006; Keetels et al., *Exp. Brain Res.*, 180, 449–56, 2007; Keetels, M., Vroomen, J., *Percept. Psychophys.*, 70, 765–71, 2008, Keetels, M., Vroomen, J., *Neurosci. Lett.*, 430, 130–4, 2008. With permission.)

Nijhawan 1994, 1997, 2002), a flash appears to lag behind a moving visual stimulus even though the stimuli are presented at the same physical location. To induce temporal ventriloquism, Vroomen and de Gelder added a single click presented slightly before, at, or after the flash (intervals of 0, 33, 66, and 100 ms). The results showed that the sound attracted the temporal onset of the flash and shifted it in the order of ~5%. A sound ~100 ms before the flash thus made the flash appear ~5 ms earlier, and a sound 100 ms after the flash made the flash appear ~5 ms later. A sound, including the synchronous one, also improved sensitivity on the visual task because JNDs on the visual task were better if a sound was present rather than absent.

Yet another recent manifestation of temporal ventriloquism used an *apparent visual motion paradigm*. Visual apparent motion occurs when a stimulus is flashed in one location and is followed by another identical stimulus flashed in another location (Korte 1915). Typically, an illusory movement is observed that starts at the lead stimulus and is directed toward the second lagging stimulus (the strength of the illusion depends on the exposure time of the stimuli, and the temporal and spatial separation between them). Getzmann (2007) explored the effects of irrelevant sounds on this motion illusion. In their study, two temporally separated visual stimuli (SOAs ranged from 0 to 350 ms) were presented and participants classified their impression of motion using a categorization system. The results demonstrated that sounds intervening between the visual stimuli facilitated the impression of apparent motion relative to no sounds, whereas sounds presented before the first and after the second visual stimulus reduced motion perception (see Bruns and Getzmann 2008 for similar results). The idea was that because exposure time and spatial separation were both held constant in this study, the impression of apparent motion was systematically affected by the perceived length of the interstimulus interval. The effect was explained in terms of temporal ventriloquism, as sounds attracted the illusory onset of visual stimuli.

Freeman and Driver (2008) investigated whether the timing of a static sound could influence spatiotemporal processing of visual apparent motion. Apparent motion was induced by visual stimuli

alternating between opposite hemifields. The perceived direction typically depends on the relative timing interval between the left–right and right–left flashes (e.g., rightwards motion dominating when left–right interflash intervals are shortest; von Grunau 1986). In their study, the interflash intervals were always 500 ms (ambiguous motion), but sounds could slightly lead the left flash and lag the right flash by 83 ms or vice versa. Because of temporal ventriloquism, this variation made visual apparent motion depend on the timing of the sound stimuli (e.g., more rightward responses if a sound preceded the left flash, and lagged the right flash, and more leftward responses if a sound preceded the right flash, and lagged the left flash).

The temporal ventriloquist effect has also been used as a diagnostic tool to examine whether commonality in space is a constraint on intersensory pairing. Vroomen and Keetels (2006) adopted the visual TOJ task of Scheier et al. (1999) and replicated that sounds improved sensitivity in the AVVA version of the visual TOJ task. Importantly, the temporal ventriloquist effect was unaffected by whether sounds and lights were colocated or not. For example, the authors varied whether the sounds came from a central location or a lateral one, whether the sounds were static or moving, and whether the sounds and lights came from the same or different sides of fixation at either small or large spatial disparities. All these variations had no effect on the temporal ventriloquist effect, despite that discordant sounds were shown to attract reflexive spatial attention and to interfere with speeded visual discrimination. These results made the author conclude that intersensory interactions in general do not require spatial correspondence between the components of the cross-modal stimuli (see also Keetels et al. 2007).

In another study (Keetels and Vroomen 2008a), it was explored whether touch affects vision on the time dimension as audition does (visual–tactile ventriloquism), and whether spatial disparity between the vibrator and lights modifies this effect. Given that tactile stimuli are spatially better defined than tones because of their somatotopic rather than tonotopic initial coding, this study provided a strong test case for the notion that spatial co-occurrence between the senses is required for intersensory temporal integration. The results demonstrated that tactile–visual stimuli behaved like audiovisual stimuli, in that temporally misaligned tactile stimuli captured the onsets of the lights and spatial discordance between the stimuli did not harm this phenomenon.

Besides exploring whether spatial disparity affects temporal ventriloquism, the effect of synesthetic congruency between modalities was also recently explored (Keetels and Vroomen in press; Parise and Spence 2008). Parise and Spence (2008) suggested that pitch size synesthetic congruency (i.e., a natural association between the relative pitch of a sound and the relative size of a visual stimulus) might affect temporal ventriloquism. In their study, participants made visual TOJs about small-sized and large-sized visual stimuli whereas high-pitched or low-pitched tones were presented before the first and after the second light. The results showed that, at large sound–light intervals, sensitivity for visual temporal order was better for synesthetically congruent than incongruent pairs. In a more recent study, Keetels and Vroomen (in press) reexamined this effect and showed that this congruency effect could not be attributed to temporal ventriloquism, as it disappeared at short sound–light intervals if compared to a synchronous AV baseline condition that excludes response biases. In addition, synesthetic congruency did not affect temporal ventriloquism even if participants were made explicitly aware of congruency before testing, challenging the view that synesthetic congruency affects temporal ventriloquism.

Stekelenburg and Vroomen (2005) also investigated the time course and the electrophysiological correlates of the audiovisual temporal ventriloquist effect using ERPs in the FLE. Their results demonstrated that the amplitude of the visual N1 was systematically affected by the temporal interval between the visual target flash and the task-irrelevant sound in the FLE paradigm (Mackay 1958; Nijhawan 1994, 1997, 2002). If a sound was presented in synchrony with the flash, the N1 amplitude was larger than when the sound lagged the visual stimulus, and it was smaller when the sound lead the flash. No latency shifts, however, were found. Yet, based on the latency of the cross-modal effect (N1 at 190 ms) and its localization in the occipitoparietal cortex, this study confirmed the sensory nature of temporal ventriloquism. An explanation for the absence of a temporal shift of

the ERP components may lie in the small size of the temporal ventriloquist effect found (3 ms). Such a small temporal difference may not be reliably reflected in the ERPs because it reaches the lower limit of the temporal resolution of the sampled EEG.

In most of the studies examining temporal ventriloquism (visual TOJ, FLE, reporting clock position or motion direction), the timing of the visual stimulus is the task-relevant dimension. Although recently, Vroomen and Keetels (2009) explored whether a temporally offset sound could improve the *identification* of a visual stimulus whereas temporal order is not involved. In this study, it was examined whether four-dot masking was affected by temporal ventriloquism. In the four-dot masking paradigm, visual target identification is impaired when a briefly presented target is followed by a mask that consists of four dots that surround but do not touch the visual target (Enns 2004; Enns and DiLollo 1997, 2000). The idea tested was that a sound presented slightly before the target and slightly after the mask might lengthen the perceived interval between target and mask. By lengthening the perceived target–mask interval, there is more time for the target to consolidate, and in turn target identification should be easier. Results were in line with this hypothesis as a small release from four-dot masking was reported (1% improvement, which corresponds to an increase of the target–mask ISI of 4.4 ms) if two sounds were presented at approximately 100-ms intervals before the target and after the mask, rather than if only a single sound was presented before the target or a silent condition.

To summarize, there are by now many demonstrations that vision is flexible on the time dimension. In general, the perceived timing of a visual event is attracted toward other events in audition and touch, provided that the lag between them is less than ~200 ms. The deeper reason why there is this mutual attraction is still untested. Although in our view, it serves to reduce natural lags between the senses so that they become unnoticed, thus maintaining coherence between the senses.

If so, one can ask what the relationship is between temporal ventriloquism and temporal recalibration. Despite that temporal ventriloquism occurs immediately when a temporal asynchrony is presented, whereas temporal recalibration manifests itself as an aftereffect, both effects are explained as perceptual solutions to maintain intersensory synchrony. The question can then be asked whether the same mechanism underlies the two phenomena. At first sight, one might argue that the magnitude of the temporal ventriloquist effect seems smaller than the temporal recalibration effects (temporal ventriloquism: Morein-Zamir et al. 2003, ~15 ms JND improvement; Scheier et al. 1999, 15 ms JND improvement; Vroomen and Keetels 2006, ~6 ms JND improvement; temporal recalibration: Fujisaki et al. 2004, ~30 ms PSS shifts for 225 ms adaptation lags; Hanson et al. 2008, ~35 ms PSS shifts for 90 ms adaptation lags; Navarra et al. 2009, ~20 ms shifts in reaction times; although relatively small effects were found by Vroomen et al. 2004, ~8 ms PSS shifts for 100 ms adaptation lags). However, these magnitudes cannot be compared directly because the temporal ventriloquist effect refers to an improvement in JNDs, whereas the temporal recalibration effect is typically a shift of the PSS. Moreover, in studies measuring temporal recalibration, there is usually much more exposure to temporal asynchronies than in studies measuring temporal ventriloquism. Therefore, it remains up to future studies to examine whether the mechanisms that are involved in temporal ventriloquism and temporal recalibration are the same.

9.6 TEMPORAL SYNCHRONY: AUTOMATIC OR NOT?

An important property about the perception of intersensory synchrony is to know whether it is perceived in an automatic fashion or not. As is often the case, there are two opposing views on this issue. Some have reported that the detection of temporal alignment is a slow, serial, and attention-demanding process, whereas others have argued that it is fast and only requires a minimal amount of attention that is needed to perceive the visual stimulus, but once this criterion is met, audiovisual or visual–tactile integration comes for free.

An important signature of automatic processing is that the stimulus in question is salient and “pops out.” If so, the stimulus is easy to find among distracters. What about intersensory synchrony:

does it “pop out”? In a study by van de Par and Kohlrausch (2004), this question was addressed by presenting observers a visual display of a number of independently moving circles moving up and down along a Gaussian profile. Along with the motion display, a concurrent sound was presented in which amplitude was modulated coherently with one of the circles. The participant’s task was to identify the coherently moving visual circle as quickly as possible. The authors found that response times increased approximately linearly with the numbers of distracters (~500 ms/distracter), indicating a slow serial search process rather than pop-out.

Fujisaki et al. (2006) came to similar conclusions. They examined search functions for a visual target that changed in synchrony with an auditory stimulus. The visual display consisted of two, four, or eight luminance-modulated Gaussian blobs presented at 5, 10, 20, and 40 Hz that were accompanied by a white noise sound whose amplitude was modulated in synch with one of the visual stimuli. Other displays contained clockwise/counterclockwise rotations of windmills synchronized with a sound whose frequency was modulated up or down at a rate of 10 Hz. The observers’ task was to indicate which visual stimulus was luminance-modulated in synch with the sound. Search functions for both displays were slow (~1 s/distractor in target-present displays), and increased linearly with the number of visual distracters. In a control experiment, it was also shown that synchrony discrimination was unaffected by the presence of distractors if attention was directed at the visual target. Fujisaki et al. therefore concluded that perception of audiovisual synchrony is a slow and serial process based on a comparison of salient temporal features that need to be individuated from within-modal signal streams.

Others, though, came to quite opposing conclusions and found that intersensory synchrony can be detected in an automatic fashion. Most notably, van der Burg et al. (2008b) reported an interesting study in which they showed that a simple auditory pip can drastically reduce search times for a color-changing object that is synchronized with the pip. The authors presented a horizontal or vertical target line among a large array of oblique lines. Each of the lines (target and distracters) changed color from green-to-red or red-to-green in a random fashion. If a pip sound was synchronized with a color change, visual attention was automatically drawn to the location of the line that changed color. When the sound was synchronized with the color change of the target, search times improved drastically and the number of irrelevant distracters had virtually no effect on search times (a nearly flat slope indicating pop-out). The authors concluded that the temporal information of the auditory signal was integrated with the visual signal generating a relatively salient emergent feature that automatically draw spatial attention (see also van der Burg et al. 2008a). Similar effects were also demonstrated for tactile stimuli instead of auditory pips (Olivers and van der Burg 2008; van der Burg et al. 2009).

Kanai et al. (2007) also explored temporal correspondences in visually ambiguous displays. They presented multiple disks flashing sequentially at one of eight locations in a circle, thus inducing the percept of a disk revolving around fixation. A sound was presented at one particular location in every cycle, and participants had to indicate the disk that was temporally aligned with the sound. The disk seen as being synchronized with the sound was perceived as brighter with a sharper onset and offset (Vroomen and de Gelder 2000). Moreover, it fluctuated over time and its position changed every 5 to 10 s. Kanai et al. explored whether this flexibility was dependent on attention by having observers perform a concurrent task in which they had to count the number of X’s in a letter stream. The results demonstrated that the transitions disappeared whenever attention was distracted from the stimulus. On the other hand, if attention was directed to one particular visual event—either by making it “pop-out” by using a different color, by presenting a cue next to the target dot, or by overtly cueing it—the perceived timing of the sound was attracted toward that event. These results thus suggest that perception of intersensory synchrony is flexible, and is not completely immune to attention.

These opposing views on the role of attention can be reconciled on the assumption that perception of synchrony depends on a matching process of *salient* temporal features (Fujisaki et al. 2006; Fujisaki and Nishida 2007). Saliency may be lost when stimuli are presented at fast rates (typically above 4 Hz), when perceptually grouped into other streams, or if they lack a sharp transition

(Keetels et al. 2007; Sanabria et al. 2004; Vroomen and de Gelder 2004a; Watanabe and Shimojo 2001). In line with this notion, studies reporting that audiovisual synchrony detection is slow, either presented stimuli at fast rates (>4 Hz up to 80/s) or the stimuli lacked a sharp onset/offset (e.g., van de Par, using a Gaussian amplitude modulation). Others reporting automatic detection of auditory–visual synchrony used much slower rates (1.11 Hz; van der Burg et al. 2008b) and sharp transitions (a pip).

9.7 NEURAL SUBSTRATES OF TEMPORAL SYNCHRONY

Although temporal correspondence is frequently considered one of the most important constraints on cross-modal integration (e.g., Bedford 1989; Bertelson 1999; Radeau 1994; Stein and Meredith 1993; Welch 1999; Welch and Warren 1980), the neural correlates for the ability to detect and use temporal synchrony remain largely unknown. Most likely, however, a whole network is involved. Seminal studies examining the neural substrates of intersensory temporal correspondence were done in animals. The finding that the firing rate of a subsample of cells in the superior colliculi (SC) increases dramatically and more than what can be expected by summing the unimodal impulses when auditory (tones) and visual stimuli (flashes) occur in close temporal and spatial proximity (Meredith et al. 1987; Stein et al. 1993) is well-known. More recently, Calvert et al. (2001) used functional magnetic resonance imaging (fMRI) on human subjects for studying brain areas that demonstrate facilitation and suppression effects in the blood oxygenation level–dependent (BOLD) signal for temporally aligned and temporally misaligned audiovisual stimuli. Their stimulus consisted of a reversing checkerboard pattern of alternating black and white squares with sounds presented either simultaneously with the onset of a reversal (synchronous condition) or a randomly phase-shifted asynchronous condition. The results showed an involvement of the SC as its response was superadditive for temporally matched stimuli and depressed for the temporally mismatched ones. Other cross-modal interactions were also identified in a network of cortical brain areas that included several frontal cortical sites; the right inferior frontal gyrus, multiple sites within the right lateral sulcus, and the ventromedial frontal gyrus. Furthermore, response enhancement and depression was observed in the insula bilaterally, right superior parietal lobule, right inferior parietal sulcus, left superior occipital gyrus, and left superior temporal sulcus (STS).

Please check
if the changes
made here are
OK.

Bushara et al. (2001) examined the effect of temporal asynchrony in a positron emission tomography study. Here, observers had to decide whether a colored circle was presented simultaneously with a tone or not. The stimulus pairs could either be auditory-first (AV) or vision-first (VA) at three levels of SOAs that varied in difficulty. A control condition (C) was included in which the auditory and visual stimuli were presented simultaneously, and in which participants performed a visual color discrimination task whenever a sound was present. The brain areas involved in auditory–visual synchrony detection were identified by subtracting the activity of the control condition from that in the asynchronous conditions (AV-C and VA-C). Results revealed a network of heteromodal brain areas that included the right anterior insula, the right ventrolateral prefrontal cortex, right inferior parietal lobe, and left cerebellar hemisphere. The activity in the areas that correlated positively with decreasing asynchrony revealed a cluster within the right insula, suggesting that this region is most important for the detection of auditory–visual synchrony. Given that interactions were also found between the insula, the posterior thalamus and the SC, it was suggested that intersensory temporal processing is mediated via subcortical tecto-thalamo-insula pathways.

In a positron emission tomography study by Macaluso et al. (2004), subjects were looking at a video monitor that showed a face mouthing words. In different blocks of trials, the audiovisual signals were either presented synchronously or asynchronously (the auditory stimulus was leading by a clearly noticeable 240 ms). In addition, the visual and auditory sources were either presented at the same location or in opposite hemifields. Results showed that activity in ventral occipital areas and left STS increased during synchronous audiovisual speech, regardless of the relative location of the auditory and visual input.

Please check
if the changes
made here are
OK.

More recently, in an fMRI study, Dhamala et al. (2007) examined the networks that are involved in the perception of physically synchronous versus asynchronous audiovisual events. Two timing parameters were varied: the SOA between sound and light (−200 to +200 ms) and the stimulation rate (0.5–3.5 Hz). In the behavioral task, observers had to report whether stimuli were perceived as simultaneous, sound-first, light-first, or “Can’t tell,” resulting in the classification of three distinct perceptual states, that is, the perception of synchrony, asynchrony, and “no clear perception.” The fMRI data showed that each of these stages involved activation in different brain networks. Perception of asynchrony activated the primary sensory, prefrontal, and inferior parietal cortices, whereas perception of synchrony disengaged the inferior parietal cortex and further recruited the SC.

An fMRI study by Noesselt et al. (2007) also explored the effect of temporal correspondence between auditory and visual streams. The stimuli were arranged such that auditory and visual streams were temporally corresponding or not, using irregular and arrhythmic temporal patterns that either matched between audition and vision or mismatched substantially whereas maintaining the same overall temporal statistics. For the coincident audiovisual streams, there was an increase in the BOLD response in multisensory STS contralateral to the visual stream. The contralateral primary visual and auditory cortex were also found to be affected by the synchrony–asynchrony manipulations, and a connectivity analysis indicated enhanced influence from mSTS on primary sensory areas during temporal correspondence.

In an EEG paradigm, Senkowski et al. (2007) examined the neural mechanisms underlying intersensory synchrony by measuring oscillatory gamma-band responses (GBRs; 30–80 Hz). Oscillatory GBRs have been linked to feature integration mechanisms and to multisensory processing. The authors reasoned that GBRs might also be sensitive to the temporal alignment of intersensory stimulus components. The temporal synchrony of auditory and visual components of a multisensory signal was varied (tones and horizontal gratings with SOAs ranging from −125 to +125 ms). The GBRs to the auditory and visual components of multisensory stimuli were extracted for five subranges of asynchrony and compared with GBRs to unisensory control stimuli. The results revealed that multisensory interactions were strongest in the early GBRs when the sound and light stimuli were presented with the closest synchrony. These effects were most evident over medial–frontal brain areas after 30 to 80 ms and over occipital areas after 60 to 120 ms, indicating that temporal synchrony may have an effect on early intersensory interactions in the human cortex.

Overall, it should be noted that there is a lot of variation in the outcomes of studies that have examined the neural basis of intersensory temporal synchrony. At present, the issue is far from resolved and more research has to be performed to unravel the exact neural substrates underlying it. The overall finding is that the SC and mSTS are repeatedly reported in intersensory synchrony detection studies, which at least suggests a prominent role for these structures in the processing of intersensory stimuli based on their temporal correspondence. For the time being, however, it is unknown how these areas would affect the perception of intersensory synchrony if they were damaged or temporarily blocked by, for example, transcranial magnetic stimulation.

9.8 CONCLUSIONS

In recent years, a substantial amount of research has been devoted to understanding how the brain handles lags between the senses. The most important conclusion we draw is that intersensory timing is *flexible* and *adaptive*. The flexibility is clearly demonstrated by studies showing one or another variant of temporal ventriloquism. In that case, small lags go unnoticed because the brain actively shifts one information stream (usually vision) toward the other, possibly to maintain temporal coherence. The adaptive part rests on studies of temporal recalibration demonstrating that observers are flexible in adopting what counts as synchronous. The extent to which temporal recalibration generalizes to other stimuli and domains, however, remains to be further explored. The idea that the

brain compensates for predictable variability between the senses—most notably distance—is, in our view, not well-founded. We are more enthusiastic about the notion that intersensory synchrony is perceived mostly in an automatic fashion, provided that the individual components of the stimuli are sufficiently salient. The neural mechanisms that underlie this ability are of clear importance for future research.

REFERENCES

- Alais, D., and S. Carlile. 2005. Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences of the United States of America* 102(6):2244–7.
- Arnold, D.H., A. Johnston, and S. Nishida. 2005. Timing sight and sound. *Vision Research* 45(10):1275–84.
- Arrighi, R., D. Alais, and D. Burr. 2006. Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision* 6(3):260–8.
- Asakawa, K., A. Tanaka, and H. Imai. 2009. Temporal Recalibration in Audio-Visual Speech Integration Using a Simultaneity Judgment Task and the McGurk Identification Task. Paper presented at the *31st Annual Meeting of the Cognitive Science Society*.
- Bald, L., F.K. Berrien, J.B. Price, and R.O. Sprague. 1942. Errors in perceiving the temporal order of auditory and visual stimuli. *Journal of Applied Psychology* 26:283–388.
- Bedford, F.L. 1989. Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology: Human Perception and Performance* 15(2):232–48.
- Benjamins, J.S., M.J. van der Smagt, and F.A. Verstraten. 2008. Matching auditory and visual signals: Is sensory modality just another feature? *Perception* 37(6):848–58.
- Bertelson, P. 1994. The cognitive architecture behind auditory-visual interaction in scene analysis and speech identification. *Cahiers de Psychologie Cognitive* 13(1):69–75.
- Bertelson, P. 1999. Ventriloquism: A case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann and J. Musseler (eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events*, 347–63. North-Holland: Elsevier.
- Bertelson, P., and G. Aschersleben. 1998. Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review* 5(3):482–89.
- Bertelson, P., and G. Aschersleben. 2003. Temporal ventriloquism: Crossmodal interaction on the time dimension: 1. Evidence from auditory–visual temporal order judgment. *International Journal of Psychophysiology* 50(1–2):147–55.
- Boenke, L.T., M. Deliano, and F.W. Ohl. 2009. Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. *Experimental Brain Research* 198(2–3):233–44.
- Bronkhorst, A.W. 1995. Localization of real and virtual sound sources. *Journal of the Acoustical Society of America* 98(5):2542–53.
- Bronkhorst, A.W., and T. Houtgast. 1999. Auditory distance perception in rooms. *Nature* 397:517–20.
- Bruns, P., and S. Getzmann. 2008. Audiovisual influences on the perception of visual apparent motion: Exploring the effect of a single sound. *Acta Psychologica* 129(2):273–83.
- Bushara, K.O., J. Grafman, and M. Hallett. 2001. Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience* 21(1):300–4.
- Calvert, G., P.C. Hansen, S.D. Iversen, and M.J. Brammer. 2001. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage* 14(2):427–38.
- Calvert, G., C. Spence, and B. Stein. 2004. *The Handbook of Multisensory Processes*. Cambridge, MA: The MIT Press.
- Colin, C., M. Radeau, P. Deltenre, and J. Morais. 2001. Rules of intersensory integration in spatial scene analysis and speechreading. *Psychologica Belgica* 41(3):131–44.
- Conrey, B., and D.B. Pisoni. 2006. Auditory–visual speech perception and synchrony detection for speech and nonspeech signals. *Journal of the Acoustical Society of America* 119(6):4065–73.
- Dhamala, M., C.G. Assisi, V.K. Jirsa, F.L. Steinberg, and J.A. Kelso. 2007. Multisensory integration for timing engages different brain networks. *NeuroImage* 34(2):764–73.
- Di Luca, M., T. Machulla, and M.O. Ernst. 2007. Perceived Timing Across Modalities. Paper presented at the *International Intersensory Research Symposium 2007: Perception and Action*.
- Dinnerstein, A.J., and P. Zlotogura. 1968. Intermodal perception of temporal order and motor skills: Effects of age. *Perceptual and Motor Skills* 26(3):987–1000.
- Dixon, N.F., and L. Spitz. 1980. The detection of auditory visual desynchrony. *Perception* 9(6):719–21.

Provide conference date and location.

Provide conference date and location.

- Eagleman, D.M., and A.O. Holcombe. 2002. Causality and the perception of time. *Trends in Cognitive Sciences* 6(8);323–5.
- Eimer, M., and J. Driver. 2001. Crossmodal links in endogenous and exogenous spatial attention: Evidence from event-related brain potential studies. *Neuroscience and Biobehavioral Reviews* 25(6);497–511.
- Eimer, M., and E. Schroger. 1998. ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology* 35(3);313–27.
- Engel, G.R., and W.G. Dougherty. 1971. Visual–auditory distance constancy. *Nature* 234(5327);308.
- Enns, J.T. 2004. Object substitution and its relation to other forms of visual masking. *Vision Research* 44(12);1321–31.
- Enns, J.T., and V. DiLollo. 1997. Object substitution: A new form of masking in unattended visual locations. *Psychological Science* 8;135–9.
- Enns, J.T., and V. DiLollo. 2000. What's new in visual masking? *Trends in Cognitive Sciences* 4(9);345–52.
- Ernst, M.O., and M.S. Banks. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870);429–33.
- Ernst, M.O., and H.H. Bulthoff. 2004. Merging the senses into a robust percept. *Trends in Cognitive Sciences* 8(4);162–9.
- Ernst, M.O., M.S. Banks, and H.H. Bulthoff. 2000. Touch can change visual slant perception. *Nature Neuroscience* 3(1);69–73.
- Fain, G.L. 2003. *Sensory Transduction*. Sunderland, MA: Sinauer Associates.
- Fendrich, R., and P.M. Corballis. 2001. The temporal cross-capture of audition and vision. *Perception & Psychophysics* 63(4);719–25.
- Finger, R., and A.W. Davis. 2001. Measuring Video Quality in Videoconferencing Systems. Technical Report SN187-D. Los Gatos, CA: Pixel Instrument Corporation.
- Freeman, E., and J. Driver. 2008. Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology* 18(16);1262–6.
- Frey, R.D. 1990. Selective attention, event perception and the criterion of acceptability principle: Evidence supporting and rejecting the doctrine of prior entry. *Human Movement Science* 9;481–530.
- Fujisaki, W., and S. Nishida. 2005. Temporal frequency characteristics of synchrony–asynchrony discrimination of audio-visual signals. *Experimental Brain Research* 166(3–4);455–64.
- Fujisaki, W., and S. Nishida. 2007. Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision Research* 47(8);1075–93.
- Fujisaki, W., S. Shimojo, M. Kashino, and S. Nishida. 2004. Recalibration of audiovisual simultaneity. *Nature Neuroscience* 7(7);773–8.
- Fujisaki, W., A. Koene, D. Arnold, A. Johnston, and S. Nishida. 2006. Visual search for a target changing in synchrony with an auditory signal. *Proceedings of Biological Science* 273(1588);865–74.
- Getzmann, S. 2007. The effect of brief auditory stimuli on visual apparent motion. *Perception* 36(7);1089–103.
- Grant, K.W., V. van Wassenhove, and D. Poeppel. 2004. Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. *Speech Communication* 44;43–53.
- Hanson, J.V., J. Heron, and D. Whitaker. 2008. Recalibration of perceived time across sensory modalities. *Experimental Brain Research* 185(2);347–52.
- Harrar, V., and L.R. Harris. 2005. Simultaneity constancy: Detecting events with touch and vision. *Experimental Brain Research* 166(3–4);465–73.
- Harrar, V., and L.R. Harris. 2008. The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research* 186(4);517–24.
- Heron, J., D. Whitaker, P.V. McGraw, and K.V. Horoshenkov. 2007. Adaptation minimizes distance-related audiovisual delays. *Journal of Vision* 7(13);51–8.
- Hillyard, S.A., and T.F. Munte. 1984. Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics* 36(2);185–98.
- Hirsh, I.J., and P. Fraisse. 1964. Simultaneous character and succession of heterogeneous stimuli. *L'Année Psychologique* 64;1–19.
- Hirsh, I.J., and C.E. Sherrick. 1961. Perceived order in different sense modalities. *Journal of Experimental Psychology* 62(5);423–32.
- Jaskowski, P. 1999. Reaction time and temporal-order judgment as measures of perceptual latency: The problem of dissociations. In G. Aschersleben, T. Bachmann and J. Müsseler (eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events* (pp. 265–82). North-Holland: Elsevier Science B.V.
- Jaskowski, P., and R. Verleger. 2000. Attentional bias toward low-intensity stimuli: An explanation for the intensity dissociation between reaction time and temporal order judgment? *Consciousness and Cognition* 9(3);435–56.

- Jaskowski, P., F. Jaroszyk, and D. Hojan-Jeziarska. 1990. Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research*, 52(1);35–8.
- Jones, J.A., and M. Jarick. 2006. Multisensory integration of speech signals: The relationship between space and time. *Experimental Brain Research* 174(3);588–94.
- Jones, J.A., and K.G. Munhall. 1997. The effects of separating auditory and visual sources on the audiovisual integration of speech. *Canadian Acoustics* 25(4);13–9.
- Kanai, R., B.R. Sheth, F.A. Verstraten, and S. Shimojo. 2007. Dynamic perceptual changes in audiovisual simultaneity. *PLoS ONE* 2(12);e1253.
- Kayser, C., C.I. Petkov, and N.K. Logothetis. 2008. Visual modulation of neurons in auditory cortex. *Cerebral Cortex* 18(7);1560–74.
- Keetels, M., and J. Vroomen. 2005. The role of spatial disparity and hemifields in audio-visual temporal order judgements. *Experimental Brain Research* 167;635–40.
- Keetels, M., and J. Vroomen. 2007. No effect of auditory-visual spatial disparity on temporal recalibration. *Experimental Brain Research* 182(4);559–65.
- Keetels, M., and J. Vroomen. 2008a. Tactile–visual temporal ventriloquism: No effect of spatial disparity. *Perception & Psychophysics* 70(5);765–71.
- Keetels, M., and J. Vroomen. 2008b. Temporal recalibration to tactile–visual asynchronous stimuli. *Neuroscience Letters* 430(2);130–4.
- Keetels, M., and J. Vroomen, in press. No effect of synesthetic congruency on temporal ventriloquism. *Attention, Perception, & Psychophysics*.
- Keetels, M., J. Stekelenburg, and J. Vroomen. 2007. Auditory grouping occurs prior to intersensory pairing: Evidence from temporal ventriloquism. *Experimental Brain Research* 180(3);449–56.
- King, A.J. 2005. Multisensory integration: Strategies for synchronization. *Current Biology* 15(9);R339–41.
- King, A.J., and A.R. Palmer. 1985. Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research* 60(3);492–500.
- Kopinska, A., and L.R. Harris. 2004. Simultaneity constancy. *Perception* 33(9);1049–60.
- Korte, A. 1915. Kinematoskopische untersuchungen. *Zeitschrift für Psychologie mit Zeitschrift für Angewandte Psychologie* 72;194–296.
- Levitin, D., K. MacLean, M. Mathews, and L. Chu. 2000. The perception of cross-modal simultaneity. *International Journal of Computing and Anticipatory Systems*, 323–9.
- Lewald, J., and R. Guski. 2003. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research* 16(3);468–78.
- Lewald, J., and R. Guski. 2004. Auditory–visual temporal integration as a function of distance: No compensation for sound-transmission time in human perception. *Neuroscience Letters* 357(2);119–22.
- Lewkowicz, D.J. 1996. Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology. Human Perception and Performance* 22(5);1094–106.
- Macaluso, E., N. George, R. Dolan, C. Spence, and J. Driver. 2004. Spatial and temporal factors during processing of audiovisual speech: A PET study. *NeuroImage* 21(2);725–32.
- Macefield, G., S.C. Gandevia, and D. Burke. 1989. Conduction velocities of muscle and cutaneous afferents in the upper and lower limbs of human subjects. *Brain* 112(6);1519–32.
- Mackay, D.M. 1958. Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature* 181(4607);507–8.
- Massaro, D.W., M.M. Cohen, and P.M. Smeele. 1996. Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America* 100(3);1777–86.
- Mattes, S., and R. Ulrich. 1998. Directed attention prolongs the perceived duration of a brief stimulus. *Perception & Psychophysics* 60(8);1305–17.
- McGrath, M., and Q. Summerfield. 1985. Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America* 77(2);678–85.
- McGurk, H., and J. MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264(5588);746–8.
- Meredith, M.A., J.W. Nemitz, and B.E. Stein. 1987. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience* 7(10);3215–29.
- Mitrani, L., S. Shekardjiski, and N. Yakimoff. 1986. Mechanisms and asymmetries in visual perception of simultaneity and temporal order. *Biological Cybernetics* 54(3);159–65.
- Mollon, J.D., and A.J. Perkins. 1996. Errors of judgement at Greenwich in 1796. *Nature* 380(6570);101–2.
- Morein-Zamir, S., S. Soto-Faraco, and A. Kingstone. 2003. Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research* 17(1);154–63.
- Mortlock, A.N., D. Machin, S. McConnell, and P. Sheppard. 1997. Virtual conferencing. *BT Technology Journal* 15;120–9.

Provide updated details for Keetels and Vroomen (date, volume number and page range used) if available.

Missing volume number for Levitin et al. 2000.

- Munhall, K.G., P. Gribble, L. Sacco, and M. Ward. 1996. Temporal constraints on the McGurk effect. *Perception & Psychophysics* 58(3);351–62.
- Navarra, J., A. Vatakis, M. Zampini et al. 2005. Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research* 25(2);499–507.
- Navarra, J., S. Soto-Faraco, and C. Spence. 2007. Adaptation to audiotactile asynchrony. *Neuroscience Letters* 413(1);72–6.
- Navarra, J., J. Hartcher-O'Brien, E. Piazza, and C. Spence. 2009. Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences of the United States of America* 106(23);9169–73.
- Neumann, O., and M. Niepel. 2004. Timing of “perception” and perception of “time.” In C. Kaernbach, E. Schröger, and H. Müller (eds.), *Psychophysics Beyond Sensation: Laws and Invariants of Human Cognition* (pp. 245–70): Lawrence Erlbaum Associates, Inc.
- Nijhawan, R. 1994. Motion extrapolation in catching. *Nature* 370(6487);256–7.
- Nijhawan, R. 1997. Visual decomposition of colour through motion extrapolation. *Nature* 386(6620);66–9.
- Nijhawan, R. 2002. Neural delays, visual motion and the flash-lag effect. *Trends in Cognitive Science* 6(9);387.
- Noesselt, T., J.W. Rieger, M.A. Schoenfeld et al. 2007. Audiovisual temporal correspondence modulates human multisensory human superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience* 27(42);11431–41.
- Olivers, C.N., and E. van der Burg. 2008. Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research* 1242;191–9.
- Pandey, P.C., H. Kunov, and S.M. Abel. 1986. Disruptive effects of auditory signal delay on speech perception with lipreading. *Journal of Auditory Research* 26(1);27–41.
- Parise, C., and C. Spence. 2008. Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters* 442(3);257–61.
- Pöppel, E. 1985. *Grenzen des bewusstseins, Stuttgart: Deutsche Verlags-Anstalt*, translated as *Mindworks: Time and Conscious Experience*. New York: Harcourt Brace Jovanovich. 1988.
- Poppel, E., K. Schill, and N. von Steinbüchel. 1990. Sensory integration within temporally neutral systems states: A hypothesis. *Naturwissenschaften* 77(2);89–91.
- Radeau, M. 1994. Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie Cognitive* 13(1);3–51.
- Rihs, S. 1995. The Influence of Audio on Perceived Picture Quality and Subjective Audio-Visual Delay Tolerance. Paper presented at the *MOSAIC Workshop: Advanced methods for the evaluation of television picture quality, Eindhoven, 18–19 September*.
- Roefs, J.A.J. 1963. Perception lag as a function of stimulus luminance. *Vision Research* 3;81–91.
- Rutschmann, J., and R. Link. 1964. Perception of temporal order of stimuli differing in sense mode and simple reaction time. *Perceptual and Motor Skills* 18;345–52.
- Sanabria, D., S. Soto-Faraco, and C. Spence. 2004. Exploring the role of visual perceptual grouping on the audiovisual integration of motion. *Neuroreport* 15(18);2745–9.
- Sanford, A.J. 1971. Effects of changes in the intensity of white noise on simultaneity judgements and simple reaction time. *Quarterly Journal of Experimental Psychology* 23;296–303.
- Scheier, C.R., R. Nijhawan, and S. Shimojo. 1999. Sound alters visual temporal resolution. *Investigative Ophthalmology & Visual Science* 40;4169.
- Schneider, K.A., and D. Bavelier. 2003. Components of visual prior entry. *Cognitive Psychology* 47(4);333–66.
- Sekuler, R., A.B. Sekuler, and R. Lau. 1997. Sound alters visual motion perception. *Nature* 385;308–08.
- Senkowski, D., D. Talsma, M. Grigutsch, C.S. Herrmann, and M.G. Woldorff. 2007. Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45(3);561–71.
- Shams, L., Y. Kamitani, and S. Shimojo. 2002. Visual illusion induced by sound. *Cognitive Brain Research* 14(1);147–52.
- Shimojo, S., C. Scheier, R. Nijhawan et al. 2001. Beyond perceptual modality: Auditory effects on visual perception. *Acoustical Science & Technology* 22(2);61–67.
- Shipley, T. 1964. Auditory flutter-driving of visual flicker. *Science* 145;1328–30.
- Shore, D.I., C. Spence, and R.M. Klein. 2001. Visual prior entry. *Psychological Science* 12(3);205–12.
- Shore, D.I., C. Spence, and R.M. Klein. 2005. Prior entry. In L. Itti, G. Rees and J. Tsotsos (eds.), *Neurobiology of Attention* (pp. 89–95). North Holland: Elsevier.

- Slutsky, D.A., and G.H. Recanzone. 2001. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12(1);7–10.
- Smith, W.F. 1933. The relative quickness of visual and auditory perception. *Journal of Experimental Psychology* 16;239–257.
- Soto-Faraco, S., and A. Alsius. 2007. Conscious access to the unisensory components of a cross-modal illusion. *Neuroreport* 18(4);347–50.
- Soto-Faraco, S., and A. Alsius. 2009. Deconstructing the McGurk–MacDonald illusion. *Journal of Experimental Psychology. Human Perception and Performance* 35(2);580–7.
- Spence, C., and J. Driver. 1996. Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology. Human Perception and Performance* 22(4);1005–30.
- Spence, C., and J. Driver. 2004. *Crossmodal Space and Crossmodal Attention*. Oxford: Oxford University Press.
- Spence, C., F. Pavani, and J. Driver. 2000. Crossmodal links between vision and touch in covert endogenous spatial attention. *Journal of Experimental Psychology. Human Perception and Performance* 26(4);1298–319.
- Spence, C., and S. Squire. 2003. Multisensory integration: Maintaining the perception of synchrony. *Current Biology* 13(13);R519–21.
- Spence, C., D.I. Shore, and R.M. Klein. 2001. Multisensory prior entry. *Journal of Experimental Psychology. General* 130(4);799–832.
- Spence, C., R. Baddeley, M. Zampini, R. James, and D.I. Shore. 2003. Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics* 65(2);318–28.
- Stein, B.E., and M.A. Meredith. 1993. *The Merging of the Senses*. Cambridge MA: The MIT Press.
- Stein, B.E., M.A. Meredith, and M.T. Wallace. 1993. The visually responsive neuron and beyond: Multisensory integration in cat and monkey. *Progress in Brain Research* 95;79–90.
- Stein, B.E., N. London, L.K. Wilkinson, and D.D. Price. 1996. Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience* 8(6);497–506.
- Stekelenburg, J.J., and J. Vroomen. 2005. An event-related potential investigation of the time-course of temporal ventriloquism. *Neuroreport* 16;641–44.
- Stekelenburg, J.J., and J. Vroomen. 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience* 19(12);1964–73.
- Stelmach, L.B., and C.M. Herdman. 1991. Directed attention and perception of temporal order. *Journal of Experimental Psychology. Human Perception and Performance* 17(2);539–50.
- Sternberg, S., and R.L. Knoll. 1973. The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (ed.), *Attention and Performance* (vol. IV, pp. 629–85). New York: Academic Press.
- Stetson, C., X. Cui, P.R. Montague, and D.M. Eagleman. 2006. Motor–sensory recalibration leads to an illusory reversal of action and sensation. *Neuron* 51(5);651–9.
- Stone, J.V., N.M. Hunkin, J. Porrill, et al. 2001. When is now? Perception of simultaneity. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 268(1462);31–8.
- Sugano, Y., M. Keetels, and J. Vroomen. 2010. Adaptation to motor–visual and motor–auditory temporal lags transfer across modalities. *Experimental Brain Research* 201(3);393–9.
- Sugita, Y., and Y. Suzuki. 2003. Audiovisual perception: Implicit estimation of sound-arrival time. *Nature* 421(6926);911.
- Sumby, W.H., and I. Pollack. 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* 26;212–15.
- Summerfield, Q. 1987. A comprehensive account of audio-visual speech perception. In B. Dodd and R. Campbell (eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 3–51). London: Lawrence Erlbaum Associates.
- Takahashi, K., J. Saiki, and K. Watanabe. 2008. Realignment of temporal simultaneity between vision and touch. *Neuroreport* 19(3);319–22.
- Tanaka, A., S. Sakamoto, K. Tsumura, and S. Suzuki. 2009a. Visual speech improves the intelligibility of time-expanded auditory speech. *Neuroreport* 20;473–7.
- Tanaka, A., S. Sakamoto, K. Tsumura, and Y. Suzuki. 2009b. Visual speech improves the intelligibility of time-expanded auditory speech. *Neuroreport* 20(5);473–7.
- Teatini, G., M. Ferne, F. Verzella, and J.P. Berruecos. 1976. Perception of temporal order: Visual and auditory stimuli. *Giornale Italiano di Psicologia* 3;157–64.
- Teder-Salejari, W.A., F. Di Russo, J.J. McDonald, and S.A. Hillyard. 2005. Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience* 17(9);1396–409.

- Titchener, E.B. 1908. *Lectures on the Elementary Psychology of Feeling and Attention*. New York: Macmillan.
- van de Par, S., and A. Kohlrausch. 2004. Visual and auditory object selection based on temporal correlations between auditory and visual cues. Paper presented at the *18th International Congress on Acoustics, Kyoto, Japan*.
- van der Burg, E., C.N. Olivers, A.W. Bronkhorst, and J. Theeuwes. 2008a. Audiovisual events capture attention: Evidence from temporal order judgments. *Journal of Vision* 8(5);2 1–10.
- van der Burg, E., C.N. Olivers, A.W. Bronkhorst, and J. Theeuwes. 2008b. Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology. Human Perception and Performance* 34(5);1053–65.
- van der Burg, E., C.N. Olivers, A.W. Bronkhorst, and J. Theeuwes. 2009. Poke and pop: Tactile–visual synchrony increases visual saliency. *Neuroscience Letters* 450(1);60–4.
- Van Eijk, R.L. 2008. *Audio-Visual Synchrony Perception*.
- Van Eijk, R.L., A. Kohlrausch, J.F. Juola, and S. van de Par. 2008. Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics* 70(6);955–68.
- van Wassenhove, V., K.W. Grant, and D. Poeppel. 2007. Temporal window of integration in auditory–visual speech perception. *Neuropsychologia* 45;598–601.
- Vatakis, A., and C. Spence. 2006a. Audiovisual synchrony perception for music, speech, and object actions. *Brain Research* 1111(1);134–42.
- Vatakis, A., and C. Spence. 2006b. Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience Letters* 393(1);40–4.
- Vatakis, A., and C. Spence. 2007. Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & Psychophysics* 69(5);744–56.
- Vatakis, A., and C. Spence. 2008. Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica* 127(1);12–23.
- Vatakis, A., J. Navarra, S. Soto-Faraco, and C. Spence. 2007. Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research* 181(1);173–81.
- Vatakis, A., A.A. Ghazanfar, and C. Spence. 2008a. Facilitation of multisensory integration by the “unity effect” reveals that speech is special. *Journal of Vision* 8(9);14 1–11.
- Vatakis, A., J. Navarra, S. Soto-Faraco, and C. Spence. 2008b. Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Experimental Brain Research* 185(3);521–9.
- Vibell, J., C. Klinge, M. Zampini, C. Spence, and A.C. Nobre. 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(1);109–20.
- von Grunau, M.W. 1986. A motion aftereffect for long-range stroboscopic apparent motion. *Perception & Psychophysics* 40(1);31–8.
- Von Helmholtz, H. 1867. *Handbuch der Physiologischen Optik*. Leipzig: Leopold Voss.
- Vroomen, J., and B. de Gelder. 2000. Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology. Human Perception and Performance* 26(5);1583–90.
- Vroomen, J., and B. de Gelder. 2004a. Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. In G.A. Calvert, C. Spence, and B.E. Stein (eds.). *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press.
- Vroomen, J., and B. de Gelder. 2004b. Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology. Human Perception and Performance* 30(3);513–8.
- Vroomen, J., and M. Keetels. 2006. The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology. Human Perception and Performance* 32(4);1063–71.
- Vroomen, J., and M. Keetels. 2009. Sounds change four-dot masking. *Acta Psychologica* 130(1);58–63.
- Vroomen, J., and J.J. Stekelenburg. 2009. Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*.
- Vroomen, J., M. Keetels, B. de Gelder, and P. Bertelson. 2004. Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research* 22(1);32–5.
- Watanabe, K., and S. Shimojo. 2001. When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science* 12(2);109–16.
- Welch, R.B. 1999. Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann and J. Müsseler (eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events* (pp. 371–87). Amsterdam: Elsevier.

Book title?
Please add publisher name and location if so.

Please indicate volume and page range used for Vroomen and Stekelenburg 2009.

- Welch, R.B., and D.H. Warren. 1980. Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin* 88(3);638–67.
- Yamamoto, S., M. Miyazaki, T. Iwano, and S. Kitazawa. 2008. Bayesian calibration of simultaneity in audio-visual temporal order judgment. Paper presented at the *9th International Multisensory Research Forum, Hamburg, Germany*.
- Zampini, M., D.I. Shore, and C. Spence. 2003a. Audiovisual temporal order judgments. *Experimental Brain Research* 152(2);198–210.
- Zampini, M., D.I. Shore, and C. Spence. 2003b. Multisensory temporal order judgments: The role of hemispheric redundancy. *International Journal of Psychophysiology* 50(1–2);165–80.
- Zampini, M., T. Brown, D.I. Shore et al. 2005a. Audiotactile temporal order judgments. *Acta Psychologica* 118(3);277–91.
- Zampini, M., S. Guest, D.I. Shore, and C. Spence. 2005b. Audio-visual simultaneity judgments. *Perception & Psychophysics* 67(3);531–44.
- Zampini, M., D.I. Shore, and C. Spence. 2005c. Audiovisual prior entry. *Neuroscience Letters* 381(3);217–22.

Provide conference date.

