

15

Mechanisms of simultaneity constancy

LAURENCE HARRIS, VANESSA HARRAR, PHILIP JAEKL,
AND AGNIESZKA KOPINSKA

Summary

There is a delay before sensory information arising from a given event reaches the central nervous system. This delay may be different for information carried by different senses. It will also vary depending on how far the event is from the observer and stimulus properties such as intensity. However, it seems that at least some of these processing time differences can be compensated for by a mechanism that resynchronizes asynchronous signals and enables us to perceive simultaneity correctly. This chapter explores how effectively simultaneity constancy can be achieved, both intramodally within the visual and tactile systems and cross-modally between combinations of auditory, visual, and tactile stimuli. We propose and provide support for a three-stage model of simultaneity constancy in which (1) signals within temporal and spatial windows are identified as corresponding to a single event, (2) a crude resynchronization is applied based on simple rules corresponding to the average processing speed differences between the individual sensory systems, and (3) fine-tuning adjustments are applied based on previous experience with particular combinations of stimuli.

15.1 Introduction

Although time is essential for the perception of the outside world, there is no energy that carries duration information, and consequently there can be no sensory system for time. Time needs to be constructed by the brain, and because this process itself takes time, it follows that the perception of when an event occurs must necessarily lag behind the occurrence of the event itself. In fact, Libet (2004) purports to have been able to measure this delay and sets it at about half a second. Several theories have been proposed as to how the perceived timing of events might be reconstructed during this gap, but it is not the aim of this chapter to review these often largely philosophical theories. Here we address a specific aspect of the reconstruction process; how stimuli are judged as being simultaneous. Although the connection between the perception of the absolute time of events (“the now”) and actual time is arbitrary, the relative timing of various stimuli is not. Are we able to correct for errors that arise from the fact that various stimuli (both within and across sensory modalities) take different amounts of time to be processed and therefore to be

perceived accurately as simultaneous? This chapter reviews the circumstances under which simultaneity is correctly perceived and proposes a model for achieving accurate perception of simultaneity. Although the chemical senses have been demonstrated as having exquisite temporal sensitivity (e.g., differences of 100 msec between the time of arrival of odors to the two nostrils can be detected, von Békésy 1967) this chapter will concern itself specifically with the senses of touch, vision, and hearing.

Several challenges must be overcome to determine which, of all possible stimuli, correspond to a given event and thus are likely to have been simultaneous. The different senses collect data concerning an object or event in different temporal and spatial reference frames, generally about different attributes, and with different degrees of resolution and reliability. Sometimes information picked up by different senses is redundant – for example, information about the time of a handclap can be relayed by auditory, visual, tactile, and proprioceptive systems – which raises the issue of how a single data point (time of clap) might be extracted from such a diverse range of signals.

Determining whether different sensory events occurred at the same time requires the perceptual systems to take into account the different processing times of each sensory signal. Timing differences can arise from both intrinsic and extrinsic factors. Extrinsic factors include the time it takes the energy from the event to reach the neural sensors. Intrinsic factors include the time for the transduction process (King & Palmer 1985) and the neural transmission time for the information to pass from the transducers to the central nervous system (von Békésy 1963; Macefield et al. 1989). Variations in the stimuli themselves, such as their intensity and eccentricity (Wilson & Anstis 1969; Nickalls 1996), also contribute to the variability in processing times that need to be taken into account. Further, personal differences, for example in attention, can change the processing speeds of stimuli (Spence et al. 2001). Considering these factors, it is therefore easy to see how the neural correlates of any two initially simultaneous stimuli can be asynchronous even when the stimuli relate to the same event.

Reconstructing the actual time of an event or the relative timing of its components involves making some allowance for these variable delays in order to identify the simultaneous multimodal components that relate to a given event. We and others have demonstrated that in some situations these variations can indeed be taken into account and stimuli from various modalities can be veridically perceived as synchronous (Engel & Dougherty 1971; Sugita & Suzuki 2003; Kopinska & Harris 2004; Alais & Carlile 2005). This chapter explores how effectively these time variations are dealt with, both intermodally and intramodally, and suggests a mechanism for how this might be done. When the mechanism operates perfectly, the relative timing of sensory events is correctly perceived.

15.2 Simultaneity constancy

The ability to perceive simultaneous events correctly despite variations in the timing of the sensory representations of the component stimuli is known as simultaneity constancy (Kopinska & Harris 2004). Simultaneity constancy is in line with other perceptual

constancies. For example, size constancy in which the perceived size of an object is maintained despite variations in the retinal information (Gregory 1963; McKee & Smallman 1998; see Walsh & Kulikowski 1998 for a review).

15.3 Assessing the challenge

Before we consider how well incidental temporal delays between sensory signals are compensated by the brain, we need to assess the magnitude of the temporal asynchronies that need to be compensated. Some delays can be objectively measured, such as the time it takes for sound to reach the observer, but delays introduced internally by differential processing times are more difficult to assess. Theoretically, objective measures can be obtained from delays in evoked potentials to unimodal stimuli (Jeffreys & Axford 1972; Celesia 1976; Lesevre 1982; Liegeois-Chauvel et al. 1991). But signals arrive at different parts of the cortex at a range of times so the measurable response might not correspond to the signals involved in temporal judgments.

Reaction times to stimuli can provide an indirect measure of processing time. When used with care they provide a relatively straightforward method for assessing the effects of stimulus characteristics on neural processing time (Luce 1986). Reaction times are comprised of several components: the time it takes energy to reach the receptors, the transduction and conduction times, the central processing time, and the time it takes to plan and execute the motor response. There is no logical reason to think that the motor component would be affected by manipulation of stimulus properties such as intensity. Therefore, we assume that the difference in reaction times to various individual stimuli represents the total time difference that needs to be compensated for to resynchronize stimuli. A reaction time difference is an estimate of the staggered delay with which the stimuli arrive and meet the detection criterion; a simultaneity constancy mechanism needs to correct for this delay because it was not present in the originally synchronous stimuli.

Similar to evoke potentials, however, the reaction time for detecting simple stimuli may not reflect the processing times used for tasks other than detection (e.g., Jaśkowski 1999; Jaśkowski & Verleger 2000). Further dissociations between reaction time and temporal order judgments (TOJs) can occur if the neural responses to the stimuli are of different magnitudes. Because reaction times rely on a stimulus meeting criterion before a response is made, it is important that the magnitude of the internal representations are approximately equal (Sternberg & Knoll 1973): “Unless two responses are identical in size and shape, differing by a time translation only, there is not uniquely defined latency difference” (p. 649). It is therefore important for experiments using this method to equate stimuli for intensity as much as possible.

15.4 Methods used to assess perceived simultaneity

There are two methods generally used to assess the perception of simultaneity among auditory, visual, and tactile stimuli:

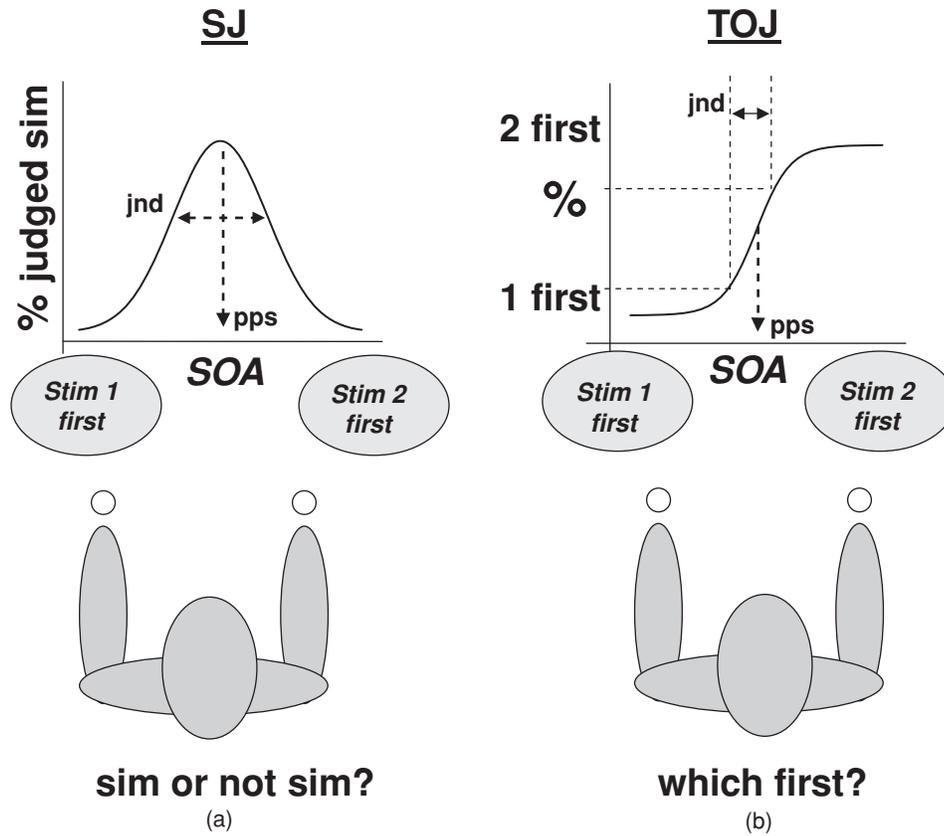
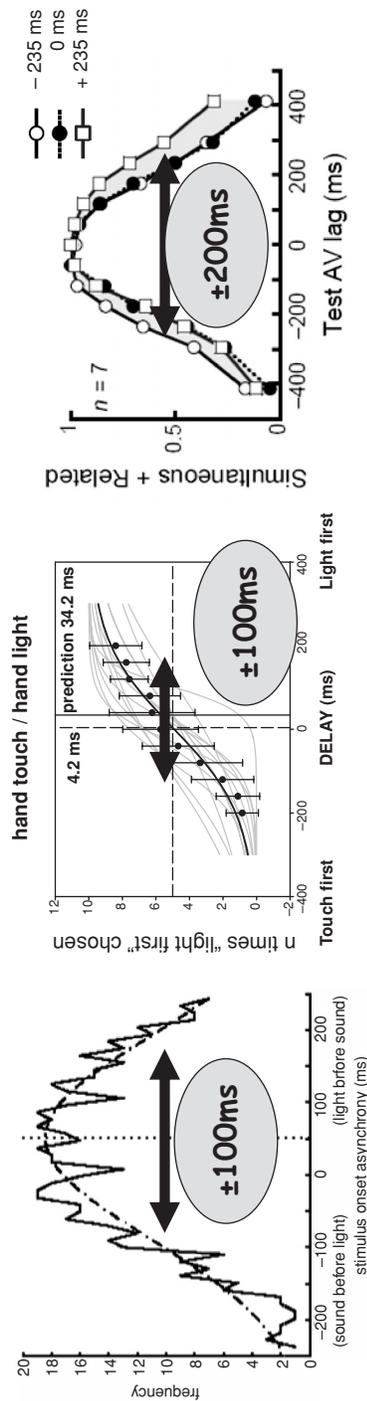


Fig. 15.1 Two methods to assess simultaneity. (A). Simultaneity judgments (SJs). The SJ data graph shows the percentage of times a given SOA was chosen as appearing simultaneous. The curve is well approximated by a Gaussian where the point of subjective simultaneity (PSS) is given by the peak of the curve and the standard deviation (STD $\pm 34\%$) is equivalent to the just noticeable difference (JND). (B). Temporal order judgments (TOJs). The TOJ data graph shows the percentage of times that stimulus 2 (e.g., a light) was chosen as appearing before stimulus 1 (e.g., a sound). This function is well approximated by a cumulative Gaussian. The point of subjective simultaneity (PSS) is given by the 50% point where the subject is equally likely to say either stimulus came first and the standard deviation (STD) is equivalent to the just noticeable difference (JND) defined as $\pm 34\%$ from 50%. Note that both these functions assume symmetry around the PSS, which may not be a valid assumption under all circumstances.

1. *Simultaneity judgments (SJ)*: A forced-choice decision is made between whether two stimuli are “simultaneous” or “successive.” Generally these decisions are reported as a frequency distribution of the “number of times subjects reported simultaneous” that tends to be normally distributed when plotted as a function of the stimulus onset asynchrony (SOA) between the two stimuli (Fig. 15.1A). The peak of this curve indicates the SOA at which subjects are most likely to say “simultaneous”: The point of subjective simultaneity (PSS). It is convenient to define the just noticeable difference (JND) as one standard deviation ($+34\%$) from the PSS. JNDs so defined typically indicate that combinations of auditory, visual, and tactile stimuli need to be separated by



(a) from Stone et al. (2001)

(b) from Harris and Harrar. (In press)

(c) from Fujisaki et al. (2004)

Fig. 15.2 The width of the temporal integration window. Several studies using different methods suggest that the window within which temporal integration occurs is between 100 and 200 ms. (A) Auditory visual simultaneity judgments. Data reproduced by permission from Stone, R. V. et al. (2001) "When is now?" Perception and simultaneity. *Proc Roy Soc Lond B* 268: 31–38. (B) Visual tactile temporal order judgments. Data reproduced with permission from Harrar, V. & Harris, L. R. (2005). Simultaneity constancy: detecting events with touch and vision. *Exp Brain Res* 166: 465–473; (C) Auditory visual simultaneity judgments before and after exposure to an adaptation regime designed to shift the PSS. Note the asymmetric changes from the baseline (filled circles) with the dominant effect on JNDs. Data reproduced with permission from Fujisaki et al. (2004). "Recalibration of audiovisual simultaneity." *Nat Neurosci* 7: 773–778.

about 150–200 ms to be reliably perceived as asynchronous (see, e.g., Figs. 15.2A and C, taken from Stone et al. 2001, and Fujisaki et al. 2004, respectively).

In interpreting SJ data, care has to be taken to distinguish between when subjects are asked (a) whether the members of a single stimulus pair were simultaneous or not (e.g., Zampini, Guest et al. 2005) or (b) which of two successively presented pairs appeared most simultaneous. The former is highly subjective, psychophysically uncontrolled, and subject to criterion shifts in the JND. The latter design requires that one of the sets be “truly simultaneous.” It therefore cannot be used to determine the PSS because it requires the PSS to be presented as the reference for each forced choice. It can, however, allow one to determine relatively criterion-free estimates of the JNDs.

2. *Temporal order judgments (TOJ)*: A forced choice is made as to which of two stimuli came on first. A psychometric curve is fitted to the proportion of times a subject perceived one of the stimuli as occurring first, plotted as a function of the SOA (Fig. 15.1B). Such functions can be well described by a cumulative Gaussian with the 50% point defining the PSS and the standard deviation defining the JND. JNDs as such are typically around ± 50 ms (Spence et al. 2001; and see Fig. 15.2B, taken from Harrar & Harris 2005).

Care has to be taken to consider biases – for example, response bias such as a tendency to respond with the right hand, or a bias toward responding “light” as it is generally the most salient cue. Biases can be reduced by alternating blocks in which subjects report which one came first with blocks in which they report which one came second.

15.5 Comparison of SJs and TOJs

SJ and TOJ measurements provide estimates of the PSS and associated JNDs. An advantage of measuring SJs is that direct judgments of perceived simultaneity are obtained that are relatively free of cognitive bias (Schneider & Bavelier 2003; Zampini, Shore et al. 2005). A disadvantage of SJs is that judgments of whether two things appear simultaneous or not are highly subjective and vulnerable to random fluctuations of criterion and bias. The participant’s criteria for perceived simultaneity may not be consistent throughout or between experimental sessions. An advantage of TOJs is that they provide statistically reliable psychometric data relatively immune to subjective criteria biases because of the forced choice between two independent alternatives; a disadvantage is that the decision is based on the remembered (i.e., retrospective) temporal sequence and is therefore vulnerable to postperceptual biases.

At first glance it would seem that the PSS and JND obtained by SJs and TOJs would be the same. But in fact the values obtained using SJs and TOJs are not necessarily the same. It is possible to perceive two stimuli as asynchronous but to not know which one came first. Allan (1975) suggested that SJs and TOJs are processed at different stages and that successiveness is needed before correct temporal order can be perceived.

SJs and TOJs can be affected differently by certain properties of the stimulus pairs (e.g., Shore et al. 2002). For example, spatial redundancy: When multimodal stimuli are separated both by modality and space, multimodal TOJs are facilitated but SJs are impeded (Spence

et al. 2003). SJs and unimodal TOJs are easiest when stimuli are in close spatial proximity probably because apparent motion can be used to cue temporal order and asynchrony and because detecting apparent motion is easier when the stimuli are close together.

Thus the choice of measure should be carefully tailored to the particular experimental question and stimuli used. Some problems plague both techniques, however. For example, JNDs are always vulnerable to the phenomenon of temporal ventriloquism whether assessed by SJs or TOJs. As a parallel to spatial ventriloquism, in which the perceived location of a sound is altered by the presence of a related light, temporal ventriloquism is when the perceived *time* of a light (or other stimulus) is affected by the presence of a related sound (or other stimulus) (Aschersleben & Bertelson 2003; Bertelson & Aschersleben 2003; Morein-Zamir et al. 2003; Jaekl & Harris 2007). If stimuli are drawn together in time in this way, they will tend to be perceived as simultaneous over a larger range than they might otherwise be – causing the JND to appear larger. Temporal ventriloquism as such only affects the JND. However, if the effects of temporal ventriloquism are asymmetric (Morein-Zamir et al. 2003) then the PSS could also be shifted closer to true simultaneity (see also Spence & Squire 2003). It is therefore not possible to measure the accuracy (JNDs) by which the relative timing of the two components is assessed by the intervention of separating them by a small amount of time because the act of separating them can create a different percept in which the components' perceived location in time is distorted. Only if the factors determining temporal ventriloquism are kept constant can two sets of responses be compared. We will now review some of the experiments using SJs and TOJs to investigate the perceived relative time of various stimulus pairings.

15.6 Simultaneity constancy in the auditory/visual system

Although light travels enormously faster than sound, the complexity of the transduction process for light takes about 45 ms longer than sound (vision about 60 ms: Jeffreys & Axford 1972; Lesevre 1982; auditory about 15 ms: Celestia 1976; Liegeois-Chauvel et al. 1991). The speed of travel of sound adds a delay of 3 ms/m so that at 15 meters the faster transduction speed and the slower travel time roughly cancel out and auditory and visual information reach the brain at approximately the same time. This theoretical concept has been called “the horizon of simultaneity” by philosophers such as Pöppel (1988) and Dennett (1991). When TOJs for sound/light pairs at different distances were made to measure this “horizon” it was interesting to notice that there was, in fact, nothing special about auditory/visual pairs presented at this distance. The relative timing of pairs over a whole range of distances up to 32 m was correctly perceived (Engel & Dougherty 1971; Kopinska & Harris 2004). A simultaneity constancy mechanism thus seems to be active over this range of distances. Using headphones while watching visual targets at different distances, Sugita and Suzuki (2003) confirmed that the time difference needed to match the auditory and visual targets increased with distance up to about 20 m. Their data are reproduced in Fig. 15.3. Because Sugita and Suzuki's subjects wore headphones, the increasing delay of the sound required for it to appear simultaneous with a light at increasing distance corresponds to the brain allowing for the time it would have taken the sound to reach the head from the location of

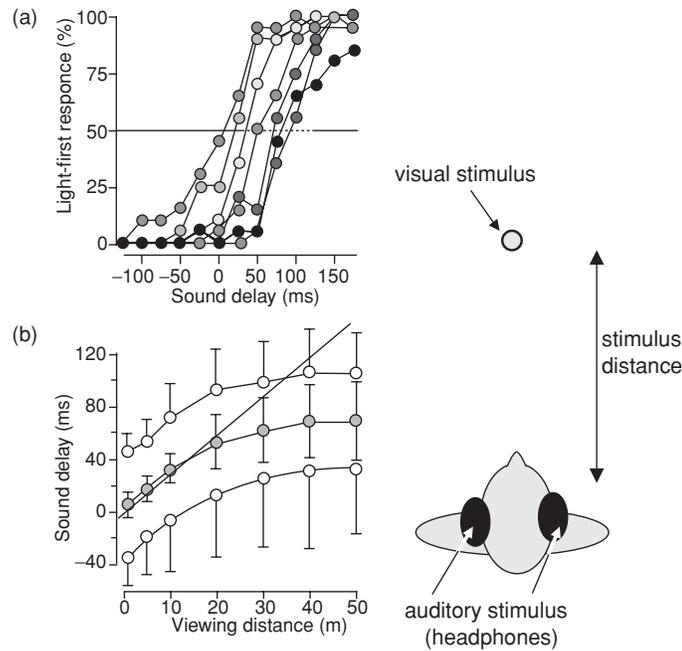


Fig. 15.3 Data taken from Sugita and Suzuki (2003) showing TOJs obtained for sound presented through headphones and lights presented at different distances (top panel). The lower panel shows the PSS obtained from each curve plotted as a function of distance (filled circles). There is an increase in PSS with distance up to about 15 m even though the sounds were presented in headphones. This increase corresponds to the brain allowing for the time it would have taken sound to reach the head from the light source if the sound had been externalized. Reproduced with permission from Sugita and Suzuki (2003) “Audiovisual perception: Implicit estimation of sound-arrival time.” *Nature* 421: 911.

the light. This is a powerful demonstration of simultaneity constancy driven by the visual cues to distance.

Kopinska and Harris (2004) found that many factors that introduce differences in the timing of the light and sound stimuli (eccentricity, distance, and intensity) were all taken into account up to 32 m so that the simultaneity of many stimulus pairs was correctly perceived up to this distance (Fig. 15.4).

Under circumstances where distance information is not easily available, such as in a dark anechoic chamber (Lewald & Guski 2004) or when causality was involved (Arnold et al. 2005), simultaneity constancy was not always found. Lewald and Guski’s data are reproduced in Fig. 15.5. An example involving causality is when two disks move on a collision course; they can either be seen to pass through each other or to bounce off each other (Arnold et al. 2005). The bounce interpretation is perceived more often if a sound accompanies the “collision.” The sound is most effective in doing this if it is presented not at the instant of collision but just prior to it – suggesting that the sound and the visual event need to match in “brain time” rather than in real time for optimum performance.

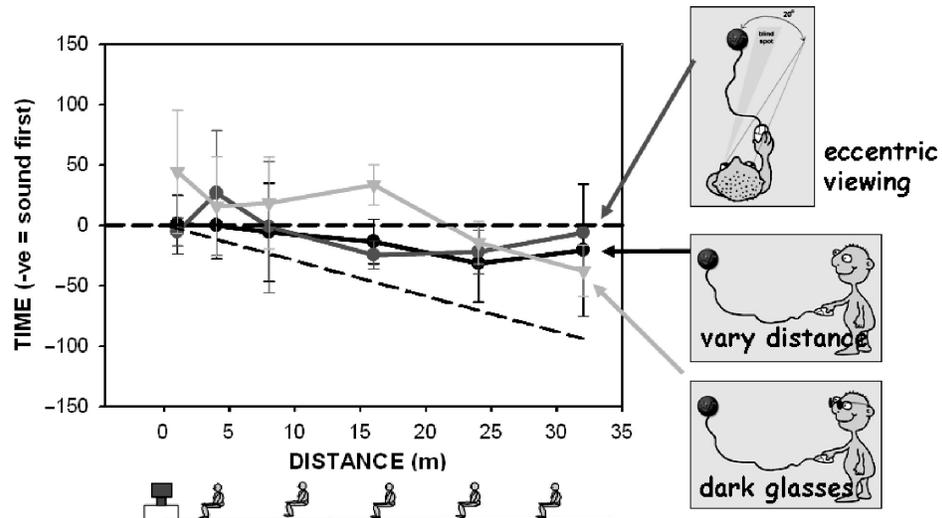


Fig. 15.4 Data redrawn from Kopinska and Harris (2004) showing the PSSs obtained from TOJs between lights and sounds presented at distances of up to 32 m. The PSSs are not significantly different from zero and are not significantly shifted by manipulations that dramatically alter the speed with which light is processed (eccentric viewing, or wearing dark glasses).

Interestingly, Arnold et al. (2005) found that the most effective stimulus onset asynchrony (SOA) for perceiving the bounce needed to be increased with the distance to the observer in accordance with the longer time taken for the sound to reach the observer. Thus it appears that the simultaneity mechanism is not engaged in circumstances involving the perception of causality.

Alais and Carlile (2005) showed that robust cues to distance are needed for auditory/visual resynchronization. Of course, if distance cues are not available, such as in an anechoic chamber, or in a field, then it would be next to impossible to correctly compensate for the delayed auditory stimulus. Dixon and Spitz (1980) determined the amount of temporal delay for which the soundtrack of a movie could be desynchronized and still be perceived as synchronous. Their results showed a large tolerance within which the desynchronization went unnoticed. However distance cues in movies, so important for simultaneity constancy to be achieved, almost always represent a conflict because the distance from the viewer to the screen is rarely the distance at which the filmmaker wishes the audience to perceive the action being portrayed.

15.7 Simultaneity constancy in the visual/tactile system

Sensory timing differences also need to be taken into account for the veridical perception of simultaneous visual and tactile stimuli. Tactile stimuli, like auditory stimuli, are transduced

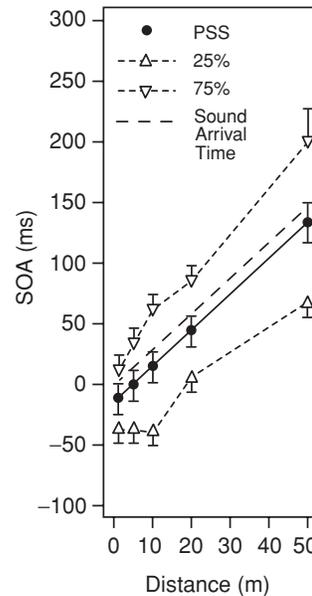


Fig. 15.5 Data taken from Lewald and Guski (2004) showing that when distance is hard to determine (in this case in a dark anechoic room) PSS increases with distance (for lights and sounds presented at different distances) and does not show simultaneity constancy. Note that in this example, when the stimuli are presented at a distance of zero, the sound needs to be turned on slightly before the light to be heard as simultaneous, rather than 45 ms after to allow for its faster processing time, suggesting that some compensation has taken place. Reproduced with permission from Lewald, J., & Guski, R. (2003) "Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli." *Cogn Brain Res* 16: 468–478.

faster than visual stimuli (King & Palmer 1985; Pöppel et al. 1990). For sounds, we noted that the longer time of transmission from the source to the end organ tended to offset this time advantage. For touches, the distance from the point of tactile stimulation to the brain determines the delay, which depends on the length of the nerves to the various parts of the body (Macefield et al. 1989).

To determine if simultaneity constancy was applied to visual and tactile timing differences Harrar and Harris (2005) measured TOJs between lights and touch stimulation on the hand. This point of tactile contact was chosen because stimulation on the hand is processed approximately 34 ms faster than visual stimuli, as measured by reaction time differences. The PSS values for lights and touches on the hand were not significantly different from zero, implying that simultaneity constancy was achieved.

When a touch and a visual stimulus are synchronous but occur on different parts of the body they are more likely to have separate causes or correspond to separate events. Indeed, spatial congruency may be required, and in some cases has been found to be essential, for stimuli to be bound into a single event and simultaneity constancy activated (Spence et al.

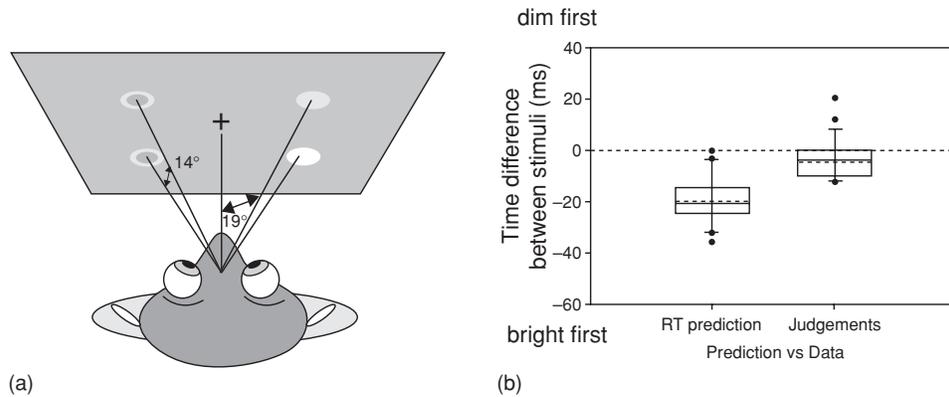


Fig. 15.6 Subjects were shown dim and bright lights (left panel) and asked which one came first. Reaction times to the dim lights were about 20 ms than for the bright ones as shown in the bar chart on the right, but the PSSs obtained from TOJs were not significantly different from zero.

2001; Spence et al. 2003; Spence & Squire 2003; Soto-Faraco et al. 2004; Harrar & Harris 2005; Zampini, Brown et al. 2005; Zampini, Guest et al. 2005).

15.8 Simultaneity constancy within a modality

The simultaneity detecting mechanism does not only need to deal with timing differences arising between senses but also within a sense. A single event usually has different attributes, for example, color, movement, and shape, that are processed with different speeds even by a single sense, for example, vision.

Therefore there is also a need for an intramodal simultaneity constancy mechanism within vision. In the experiment illustrated in Fig. 15.6, Kopinska and Harris used targets with different luminances. Reaction times suggested that these patches were processed at different speeds, but TOJs showed that the timing difference was accounted for and simultaneity constancy was achieved, that is, the PSS was not significantly different from true simultaneity. These results provide evidence that simultaneity constancy is achieved between separately processed visual signals and demonstrate the adjustment of temporal information occurs also within a single modality.

Reaction times to touches are longer for body parts that are further from the center of the head (Bergenheim et al. 1996; Harrar & Harris 2005), and the perceived temporal order of two touches on the body is predictable not from which body part was actually touched first but from the relative distances of the two touches from the head (Bergenheim et al. 1996; Craig & Baihua 1990; Shore et al. 2002; Harrar & Harris 2005). During active touches, when the foot is touched by the hand for example, rather than having the two sensations at different times, as they would if simultaneity constancy were not engaged, one of the two touches is inhibited (von Békésy 1963). Thus the touch system seems to have found other

ways of dealing with timing differences. Bergenheim et al. (1996) suggested that the touch system has “a degree of precision that is functional for the individual.”

15.9 Flexibility

Fujisaki et al. (2004) showed that the PSS for a given comparison is not fixed. The PSS of a light/sound pair can be shifted following a period of repeated exposure to a light/sound pair separated by a particular delay (Fujisaki et al. 2004; see also Vroomen et al. 2004). These results may be interpreted as indicating that the system is able to learn a new interstimulus time difference as corresponding to simultaneity. Adaptability of this kind suggests that the mechanism is not hardwired but is influenced by learning.

Fujisaki et al. (2004) adapted subjects to several different temporal intervals and found that responses varied depending on the previously experienced SOA. However, in some conditions (see Fujisaki et al.’s Figs. 15.2B and 15.2D) subjects’ PSSs were all in the direction of “sound first” even after adapting to “light first” stimuli. Further, many shifts (relative to the “no adaptation” condition) were actually in the direction opposite to the actual delay. Similar results were found in Harrar and Harris (2005) (see Fig. 15.5). These results suggest that there may be a bias toward shifting the audio/visual PSS in the direction of sound first, but it is not yet known why. It could be because of attention shifts (Spence et al. 2001) that might only occur when both stimuli are present (Aschersleben 1999) or because of an expectancy (or “prior”) built into the system based on the usual relative arrival times of light and sound. Miyazaki et al. (2006) modeled the effect of prior expectancies on the adaptability of the simultaneity constancy mechanism and showed how PSS shifts for tactile stimuli can be predicted statistically either toward or away from an experienced time stagger. Despite these complications, the flexibility of the simultaneity system provides an experimental tool for testing numerous still outstanding questions about the simultaneity mechanism.

15.10 A single, global mechanism or multiple, specific mechanisms?

There may be a single, global mechanism responsible for simultaneity constancy both within and between modalities, or there may be multiple simultaneity mechanisms comprising separate systems for audio/visual, visual/tactile, and audio/tactile comparisons. These two alternatives can be distinguished by exploiting the flexibility demonstrated by Fujisaki et al. (2004), adapting the PSS of an audio/visual pair and looking for any effects in audio/tactile and visual/tactile pairs.

15.10.1 Implications from PSS shifts

Harrar and Harris (2008) tested TOJs before and after adaptation to each of three combinations of time-staggered stimuli (light/sound, light/touch, sound/touch), to distinguish a single global simultaneity constancy mechanism from multiple mechanisms for different

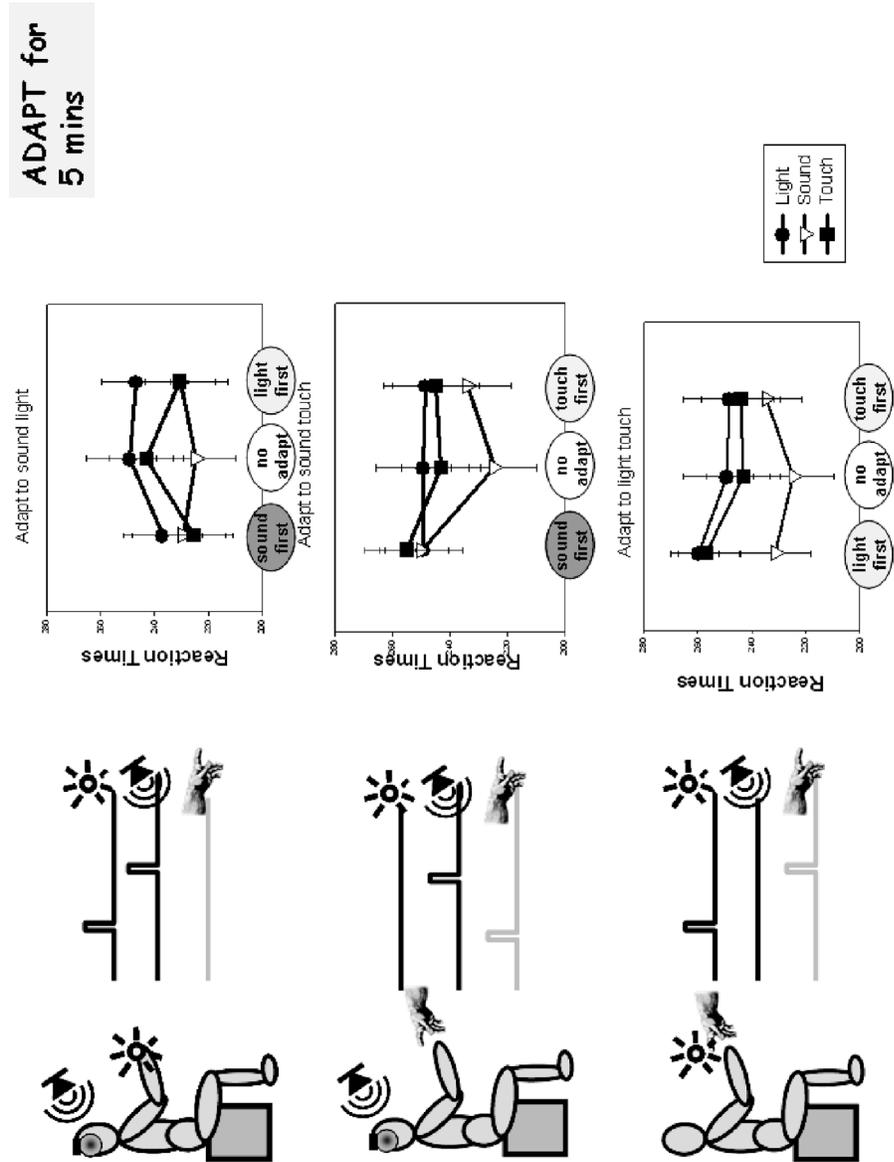


Fig. 15.7 Subjects were exposed to 5 min of time-staggered stimulus pairs either sound/light, sound/touch, or light/touch as shown in the cartoons on the left. Such exposure to sound/light pairs results in a shift of the PSS (see text). However, no shifts were found in the reaction times to sounds, lights, or touches after any pattern of exposure as shown by the graphs on the right.

stimulus combinations. A PSS shift in a pair that was not the one exposed would support a global mechanism, whereas a PSS shift of only the exposed pair would support multiple separate simultaneity constancy mechanisms. After adapting to a time-staggered sound/light combination, PSSs for the light/sound pair shifted but the PSS for the sound/touch and light/touch pairs did not change (Harrar & Harris 2008). This pattern of results renders a single, global simultaneity constancy mechanism quite unlikely. Therefore we suggest a multichannel system.

Although it appears that intermodal comparisons are accomplished through separate mechanisms, intramodal comparisons may each be executed by a single mechanism. There is some generalization across audiovisual stimuli, for example, from complex (watching a video of a person talking) to simple (beeps and flashes) examples of the same multimodal stimulus pair, suggesting that audiovisual comparisons share a simultaneity mechanism (Navarra et al. 2005; Vatakis & Spence 2006).

15.10.2 Implications from reaction times

Harrar and Harris (2008) also, for the first time, tested reaction times to individual, unimodal stimuli before and after adaptation to each of three combinations of time-staggered stimuli. Any reaction time change would indicate a global, low-level change in the functioning of the system that would blindly and without regard to context operate with a particular delay.

Even though the PSS of a light/sound pair shifted after repeated exposure, there were no systematic changes in reaction times to any of the individual stimuli (Fig. 15.7; Harrar & Harris 2008). This suggests that the timing corresponding to PSSs is constructed centrally and not tied to timing differences in the sensory signals. This is consistent with the simultaneity constancy mechanisms not working on the individual stimulus level.

15.11 Is touch special?

When Harrar and Harris (2008) tested TOJs for each of three combinations of time-staggered stimuli (light/sound, light/touch, sound/touch), the PSS of the light/touch pair or the sound/touch pair never changed, even after asynchronous exposure to light/touch or sound/touch pairs. In other words, the perception of the relative timing of stimuli where touch is involved seems more rigid. The touch system may be fundamentally different from the more passive auditory and visual systems because of its association with actions and knowledge of limb position, or it might be related to the fact that the temporal properties of a touch do not need to take into account outside factors (Miyazaki et al. 2006). The lack of plasticity in the tactile system may be evidence for a fixed temporal somatosensory homunculus (Bergenheim et al. 1996). Alternatively, the lack of immediate plasticity within the touch system may be related to the inhibition of one touch by another as described by von Békésy (1963).

15.12 Mechanisms of simultaneity constancy

Given that simultaneity constancy is engaged under many circumstances, it now behooves us to explain how variations in timing might be allowed for in a flexible simultaneity constancy mechanism. There are two broad classes of models that could achieve simultaneity constancy. Signals could be brought into line at some point in the brain using computations based on knowledge of the various factors that might desynchronize the signals. An alternative model is based on probability. Windows of acceptance for a range of time differences between stimulus pairs are generated based on previous experience. In the probabilistic model, if the time difference between signals falls in the acceptance window, the two signals are regarded as simultaneous. Both these classes of models can be implemented with single or multiple comparators.

15.12.1 Computational models

Computational models for resynchronizing individual stimuli have a long history that often smacks of the Cartesian theater in which an inner homunculus experiences what the senses present in some kind of “show in the brain” (see Dennett 1991 for a comprehensive debunking of this idea). Sternberg and Knoll (1973) described several variations. In these models, stimuli arrive at the brain at variable times. However, before they arrive at the decision center, the individual stimuli have their delays “corrected” (see Sternberg & Knoll’s Fig. 12A). Computational models require altering the timing of information as it passes through the nervous system such that by the time it arrives at the relevant decision-making site, any timing differences have already been removed.

If, for example, to achieve a central alignment in the representation of stimuli in the brain, the processing time of a particular modality needed to be made longer, this added delay in processing time would subsequently be applied in all circumstances in which that stimulation occurred regardless of which other stimuli were present. For example, if the processing of light were delayed to compensate for the time it takes for a sound coming from a distant event, the added delay would affect all subsequent perceptions involving visual stimuli. Delaying the processing of visual stimuli, while having the advantage of bringing it into synchrony with some particular sounds, would generally seem to be risky. Alternatively, computations could be simplified depending on which stimuli are being compared, adding different delays for touch if it is being compared with a light or a sound (see Sternberg & Knoll’s Fig. 12B,C) and limiting the operational range (Dennett & Kinsbourne 1992).

Reaction time and processing time evoke potential differences between various stimuli suggest empirically that no peripheral compensation for timing differences takes place at the individual stimulus location; however, this might be due to the dissociation between reaction times and perceived time (see above). Intuitively it seems unlikely that all the sources of variation in timing could be hardwired into the system or that it would be advantageous for it to be so. We therefore need to consider another class of model.

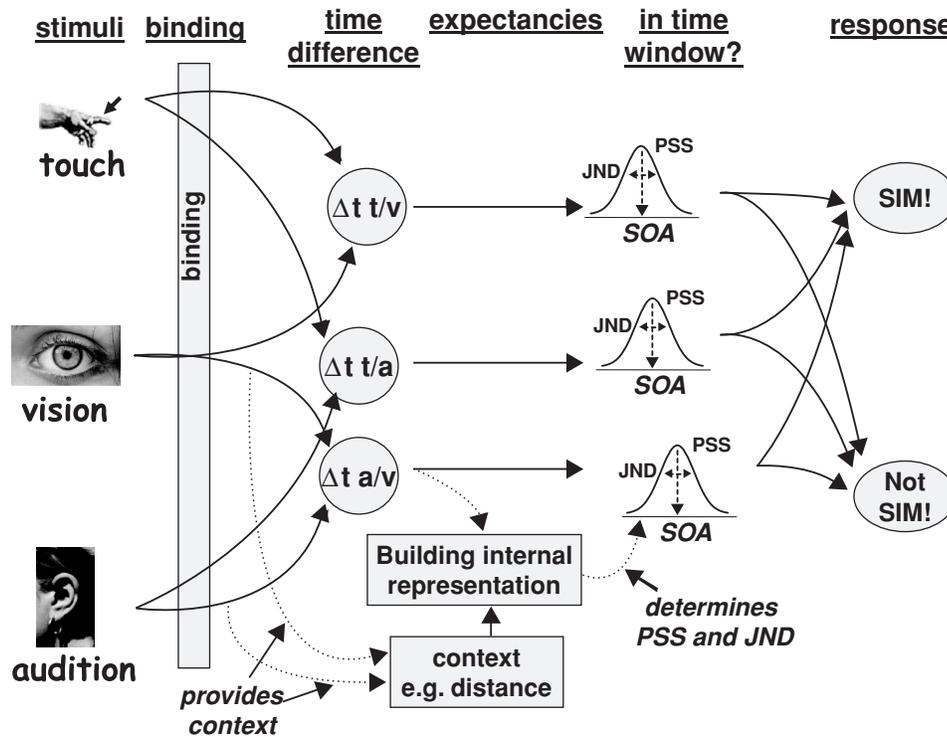


Fig. 15.8 Model for achieving simultaneity constancy. Information from different senses about which temporal judgments are to be made are first “bound,” meaning that they are regarded as coming from the same event. The time difference between them is noted and used for two purposes: (1) to decide if the difference falls within the range regarded as simultaneous (rightward pointing arrow), and (2) to contribute to the probabilistic assessment of what that range should be (down right pointing dotted arrow). For each stimulus combination an “internal representation” is built and available for comparison, dependent on experience and context. These processes are only drawn for audiovisual comparisons but are probably essentially the same for touch-vision and touch-audition, although the “contributing to the assessment” arrow from the tactile system may be weak or nonexistent. The input from the time difference assessment represents an instantaneous updating of this internal representation.

15.12.2 Probability models

Our suggested model for achieving simultaneity constancy is probabilistic rather than computational. Figure 15.8 summarizes the basic steps. The first part addresses the question of selecting appropriate stimuli for temporal compensation. The stimuli about which judgments are to be made are selected from the entire array of stimuli by a binding process. Although it is yet unclear how corresponding stimuli come to be bound together, there seems to be some spatial and temporal criteria (Spence et al. 2001; Bertelson & Aschersleben 2003).

The next stage in the model notes the timing difference between bound stimuli and uses this for two purposes. First, the time difference is used to generate an internal expectation of the time differences associated with those stimuli tagged for the particular context in which the stimuli occurred, for example, at a particular distance. It is the effect of context on this expectation that allows the “constancy” aspect of the model. Expectancies are continuously updated based on the context so that veridical and constant simultaneity can be perceived despite changes in the “context.” Note that the polarity of the time difference needs to be kept as part of this tagging procedure. Just as when describing experimental results, it is important to define that, for example, positive corresponds to “sound first.” Secondly, the time difference is also compared to this internal representation of the time differences associated with the stimuli in that particular context (shown in Fig. 15.8). If the time difference of the signals in question falls within the relevant time window, a decision of “simultaneous” is made. The time window determines the probability with which the stimuli are perceived as simultaneous. Because there is evidence for multiple simultaneity mechanisms (see *Implications from PSS shifts* previously), these separate mechanisms are distinguished within our model (Fig. 15.8).

To work through an example then, the processing time of touches on the body is generally faster than that of lights. So when a multimodal stimulus occurs that involves tactile and visual stimulation, such as watching something touch the skin or looking at an object being manipulated in the hand, the tactile input that leads the visual input, by roughly 40 ms, will be compared with the probability function (which has a peak at 40 ms based on previous experience of this combination) and will thus be within the simultaneity window. Lights and touches with a delay of 40 ms will thus be identified as simultaneous.

15.12.3 Recalibration achieved by JND increase followed by PSS shift

How does our model become “recalibrated” during exposure to a new time stagger introduced experimentally as demonstrated by Fujisaki et al. (2004)? Navarra et al. (2005, 2007) were able to find an adaptation effect for both audio/tactile and visuo/tactile pairs, even though their pairs included touch. The JNDs increased after adaptation thus tending to include a larger range of times (potentially including the experienced time stagger) and causing more SOAs to be perceived as “simultaneous.” Navarra et al. (2005) proposed that expanding of the JND is a precursor for all adaptive changes of the PSS. Figure 15.9 shows the internal expectation for time delays (“ ΔT old” in Fig. 15.9: The sum of previous experience – dotted curve) changes when repeatedly exposed to a new specific delay (labeled “ ΔT new” in Fig. 15.9). It assumes that the stored expectation declines at some rate and can be replaced as new data accumulate about a new probability of what time differences correspond to simultaneous. An interesting feature is that during the transition phase, the PSS shift is, initially, the result of an asymmetric increase in JND (simply caused by adding the distributions of the initial and the new representation). Such an asymmetric expansion is visible in the data of Fig. 15.2C (taken from Fujisaki et al. 2004 during an adaptation regime). Over time, and with further experience, the model predicts that the JNDs get

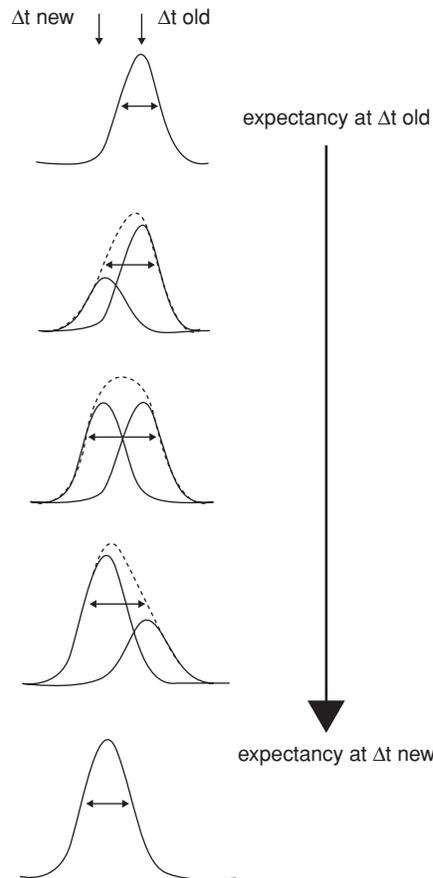


Fig. 15.9 Building an internal representation: changing one expectation to another. The curve at the top represents the expectation that Δt old corresponds to simultaneity. After Δt new is repeatedly experienced, the expectancy changes to this new value. In between the window becomes wider and asymmetric before settling down to the new values.

smaller again, but the largest PSS shift (as compared with the PSS shifts during the adaptation) still remains. This model therefore explains why similar experiments find different results with regard to JND changes and PSS shifts. PSS shifts found without JND changes (Fujisaki et al. 2004; Vroomen et al. 2004) signify the end of the recalibration process, whereas small PSS shifts accompanied by large JND increases (Navarra et al. 2005, 2007) suggest an early stage of the recalibration process.

15.13 Conclusions

This chapter has reviewed the occurrence of an important perceptual phenomenon: simultaneity constancy. There appear to be several parallel simultaneity constancy mechanisms,

each working on the timing of different combinations of stimuli. We have developed a model to explain how this may be done. The rules of when to activate these systems are not clear, but binding within a spatial window seems to be required. In this model, once a simultaneity mechanism is activated, the timing difference is compared with an internal representation of the expected time difference for those particular stimuli. SJ and TOJ estimates can then be made about the relative timing of stimuli. The model allows simple contextual factors, such as distance, to affect the internal expectancy and thus the perceived relative time of the stimuli. Highly cognitive tasks, such as watching a ball bounce, do not seem to engage simultaneity constancy, whereas simple beeps, flashes, and videos do. This may have to do with the level of complexity of the context within which the stimuli occur or might reflect the operation of other principles about which we are currently ignorant. Understanding the brain's abilities and limitations and the way in which some of the principles outlined here could be neurally implemented will advance our understanding of the perception of multimodal time.

15.14 Acknowledgments

This work was supported by a grant from the Natural Science and Engineering Research Council of Canada to Laurence Harris. Vanessa Harrar and Philip Jaekl hold NSERC graduate scholarships. We are grateful to Sugirhini and Shamini Selvanayagarajah for their help with the data collection.

References

- Alais, D., & Carlile, S. (2005). Synchronizing to real events: subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proc Natl Acad Sci U S A* **102**: 2244–2247.
- Allan, L. G. (1975). The relationship between judgments of successiveness and judgments of order. *Perception & Psychophysics* **18**: 29–36.
- Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Res* **45**: 1275–1284.
- Aschersleben, G. (1999). Task-dependent timing of perceptual events. In G. Aschersleben, T. Bachmann, & J. Müsseler (eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events* (293–318). North Holland: Elsevier.
- Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: cross-modal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. *Int J Psychophysiol* **50**: 157–163.
- Bergenheim, M., Johansson, H., Granlund, B., & Pedersen, J. (1996). Experimental evidence for a sensory synchronization of sensory information to conscious experience. In S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (eds.), *Towards a Science of Consciousness: The First Tucson Discussions and Debates* (301–310). Cambridge, MA: MIT Press.

- Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *Int J Psychophysiol* **50**: 147–155.
- Celesia, G. G. (1976). Organization of auditory cortical areas in man. *Brain* **99**: 403–414.
- Craig, J. C., & Baihua, X. (1990). Temporal order and tactile patterns. *Perception & Psychophysics* **47**: 22–34.
- Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown and Company.
- Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: the where and when of consciousness in the brain. *Behav Brain Sci* **15**: 183–201.
- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception* **9**: 719–721.
- Engel, G. R., & Dougherty, W. G. (1971). Visual-auditory distance constancy. *Nature* **234**: 308.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci* **7**: 773–778.
- Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature* **199**: 678–680.
- Harrar, V., & Harris, L. R. (2005). Simultaneity constancy: detecting events with touch and vision. *Exp Brain Res* **166**: 465–473.
- Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Exp Brain Res* **186**: 517–524.
- Jaekl, P. M., & Harris, L. R. (2007). Auditory-visual temporal integration measured by shifts in perceived temporal location. *Neurosci Lett* **417**: 219–224.
- Jaśkowski, P. (1999). Reaction time and temporal-order judgement as measures of perceptual latency: the problem of dissociations. In G. Aschersleben, T. Bachmann, & J. Müsseler (eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events* (265–282). North Holland: Elsevier.
- Jaśkowski, P., & Verleger, R. (2000). Attentional bias toward low-intensity stimuli: an explanation for the intensity dissociation between reaction time and temporal order judgment? *Conscious Cogn* **9**: 435–456.
- Jeffreys, D. A., & Axford, J. G. (1972). Source locations of pattern-specific components of human VEPs. *Exp Brain Res* **16**: 1–21.
- King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Exp Brain Res* **60**: 492–500.
- Kopinska, A., & Harris, L. R. (2004). Simultaneity constancy. *Perception* **33**: 1049–1060.
- Lesevre, N. (1982). Chronotopographical analysis of the human evoked potential in relation to the visual field. *Annals of the New York Academy of Science* **388**: 156–182.
- Lewald, J., & Guski, R. (2004). Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception. *Neurosci Lett* **357**: 119–122.
- Libet, B. (2004). *Mind Time: The Temporal Factor in Consciousness*. Cambridge, MA: Harvard University Press.
- Liegeois-Chauvel, C., Musolino, A., & Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain* **114**: 139–153.
- Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

- Macefield, G., Gandevia, S. C., & Burke, D. (1989). Conduction velocities of muscle and cutaneous afferents in the upper and lower limbs of human subjects. *Brain* **112**: 1519–1532.
- McKee, S. P., & Smallman, H. S. (1998). Size and speed constancy. In V. Walsh & J. J. Kulikowski (eds.), *Perceptual Constancy* (373–408). Cambridge: Cambridge University Press.
- Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nat Neurosci* **9**: 875–877.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cogn Brain Res* **17**: 154–163.
- Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neurosci Lett* **413**: 72–76.
- Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cogn Brain Res* **25**: 499–507.
- Nickalls, R. W. D. (1996). The influence of target angular velocity on visual latency difference determined using the rotating Pulfrich effect. *Vision Res* **36**: 2865–2872.
- Pöppel, E. (1988). *Mindworks*. Boston: Harcourt Brace Jovanovich.
- Pöppel, E., Schill, K., & von Steinbüchel, N. (1990). Sensory integration within temporally neutral systems states: a hypothesis. *Naturwissenschaften* **77**: 89–91.
- Schneider, K. A., & Bavelier, D. (2003). Components of visual prior entry. *Cogn. Psychol* **47**: 333–366.
- Shore, D. I., Spry, E., & Spence, C. (2002). Confusing the mind by crossing the hands. Brain research. *Cogn Brain Res* **14**: 153–163.
- Soto-Faraco, S., Ronald, A., & Spence, C. (2004). Tactile selective attention and body posture: assessing the multisensory contributions of vision and proprioception. *Percept Psychophys* **66**: 1077–1094.
- Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: when two locations are better than one. *Percept Psychophys* **65**: 318–328.
- Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *J Exp Psychol General* **130**: 799–832.
- Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Curr Biol* **13**: R519–521.
- Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: fundamental issues and a general model. In S. Kornblum (ed.), *Attention and Performance IV* (629–685). New York: Academic Press.
- Stone, R. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., & Beanland, M., et al. (2001). When is now? Perception and simultaneity. *Proc Roy Soc Lond B* **268**: 31–38.
- Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. *Nature* **421**: 911.
- Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neurosci Lett* **393**: 40–44.
- von Békésy, G. (1963). Interaction of paired sensory stimuli and conduction in peripheral nerves. *J Applied Physiol* **18**: 1276–1284.
- von Békésy, G. (1967). *Sensory Inhibition*. Princeton, NJ: Princeton University Press.
- Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cogn Brain Res* **22**: 32–35.

- Walsh, V., & Kulikowski, J. (1998). *Perceptual constancy: why things look as they do*. Cambridge: Cambridge University Press.
- Wilson, J. A., & Anstis, S. M. (1969). Visual delay as a function of luminance. *Am J Psychol* **82**: 350–358.
- Zampini, M., Brown, T., Shore, D. I., Maravita, A., Roder, B., & Spence, C. (2005). Audiotactile temporal order judgments. *Acta Psychol (Amst)* **118**: 277–291.
- Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Percept Psychophys* **67**: 531–544.
- Zampini, M., Shore, D. I., & Spence, C. (2005). Audiovisual prior entry. *Neurosci Lett* **381**: 217–222.