On the Timing of Signals in Multisensory Integration and Crossmodal Interactions: a Scoping Review

Philip Sanders [1,2,3], **Benjamin Thompson** [2,4,5], **Paul Corballis** [2,6] **and Grant Searchfield** [1,2,3,*]

[1] Section of Audiology, University of Auckland, Auckland, New Zealand
[2] Centre for Brain Research, University of Auckland, New Zealand
[3] Brain Research New Zealand – Rangahau Roro Aotearoa, New Zealand
[4] School of Optometry and Vision Science, University of Auckland, Auckland, New Zealand
[5] School of Optometry and Vision Science, University of Waterloo, Waterloo, Canada
[6] Department of Psychology, University of Auckland, Auckland, New Zealand

Abstract

A scoping review was undertaken to explore research investigating early interactions and integration of auditory and visual stimuli in the human brain. The focus was on methods used to study low-level multisensory temporal processing using simple stimuli in humans, and how this research has informed our understanding of multisensory perception. The study of multisensory temporal processing probes how the relative timing between signals affects perception. Several tasks, illusions, computational models, and neuroimaging techniques were identified in the literature search. Research into early audiovisual temporal processing in special populations was also reviewed. Recent research has continued to provide support for early integration of crossmodal information. These early interactions can influence higher-level factors, and vice versa. Temporal relationships between auditory and visual stimuli influence multisensory perception, and likely play a substantial role in solving the 'correspondence problem' (how the brain determines which sensory signals belong together, and which should be segregated).

---

* To whom correspondence should be addressed. E-mail: g.searchfield@auckland.ac.nz

1. Background

Traditionally, sensory research was focused on single senses with little consideration for other modalities, and this led to interesting findings about the architecture of sensory processing for each sense in isolation (e.g., Evans and Whitfield, 1964; Hubel and Wiesel, 1962, 1968). Signals from different modalities were thought to be processed separately in the primary sensory cortices, and then integrated later on in processing in the association areas of the cortex, such as the superior temporal sulcus (Benevento *et al.*, 1977; Felleman and Van Essen, 1991; Jones and Powell, 1970; Treisman and Gelade, 1980). For a more detailed description of this history and the development of multisensory research see Alais *et al.* (2010). Shimojo and Shams (2001) reviewed converging psychophysical and neuroimaging evidence that challenged the view that perception was modular (i.e., that different modalities were processed separately at early stages of processing). They concluded that emerging evidence at the time suggested that crossmodal interactions could occur early in processing, and in regions previously purported to be unisensory.

Crossmodal interactions describe the influence of a signal in one modality over the processing and perception of a signal from another modality without necessarily integrating the signals (Spence *et al.*, 2009). Multisensory integration (MSI) refers to the processes by which the brain combines information from different sensory modalities. For the behaving organism, these processes help to determine whether sounds and visual events derive from the same source so that the organism can respond accordingly. One of the central questions in MSI research is how the brain determines which sensory signals belong together, and which should be segregated. This is known as the 'correspondence problem' (Ernst, 2006) and is sometimes referred to as 'causal inference' in the multisensory perception literature (Parise *et al.*, 2012), although the latter term can also refer to a more general perceptual and cognitive process.

Three principles of MSI emerged from animal research on multisensory cells in the superior colliculus that helped to explain how cellular mechanisms contribute to solving the correspondence problem. These were (1) the spatial principle: the closer in space that stimuli of different modalities occur, the more likely they are to be integrated, (2) the temporal principle: the closer in time that stimuli of different modalities occur, the more likely they are to be integrated, and (3) the principle of inverse effectiveness: relative gains from MSI are proportionally greater when the input signals are weak compared to when they are strong (Meredith and Stein, 1986a, b; Meredith *et al.*, 1987). Multisensory interactions can produce non-linear cellular responses that are superadditive (greater than the sum of the unisensory responses) or subadditive

depending on temporal and spatial relationships between component stimuli (Meredith and Stein, 1986a, b; Meredith *et al*., 1987).

The principles of MSI have been applied to the study of human sensory processing examining neuronal activity (Liu *et al*., 2011; Nozaradan *et al*., 2012; Senkowski *et al*., 2007) and neural networks (Calvert and Thesen, 2004; Dhamala *et al*., 2007; Marchant *et al*., 2012), through to perception and behaviour (Nidiffer *et al*., 2016; Zampini *et al*., 2005a). MSI and perception can be modulated at various levels of processing by the experimenter's choice of stimuli, stimulus manipulations, and task demands (Denison *et al*., 2013; Mégevand *et al*., 2013; Setti and Chan, 2011; Stevenson *et al*., 2012a). For example, low-level factors such as intensity and spatiotemporal relationships between component stimuli, and high-level factors such as stimulus complexity and meaning can all affect perception. High- and low-levels of processing can also interact and influence each other (Setti and Chan, 2011).

The spatial principle has received a lot of attention in multisensory research (e.g., Bolognini *et al*., 2010; Frassinetti *et al*., 2002; Holmes and Spence, 2005; Macaluso and Driver, 2005; Mahoney *et al*., 2015; Meredith and Stein, 1986a, 1996; Nidiffer *et al*., 2016). However, Spence (2013) questioned its generality because robust human behavioural evidence supporting the spatial principle comes only from tasks that require an orienting (or other spatial) response, or spatial attention. On the other hand, temporal judgement and target identification tasks appear to be less sensitive or not sensitive at all to the spatial principle. Interest in temporal aspects of MSI in humans has expanded and the temporal principle may be more robust across different task demands than the spatial principle (see Spence, 2013).

Temporal offsets in the arrival times of visual and auditory signals occur naturally due to the difference between the speeds of light and sound. The respective signals also have different transduction times once they have reached the body (King and Palmer, 1985; Spence and Squire, 2003). To maintain the perception of synchrony at various distances, the perceptual system has some tolerance for these offsets. An important concept in temporal MSI research is the temporal binding window (TBW). The TBW refers to the time range in which stimuli of different modalities must fall in order to be integrated, or the amount of tolerance that the perceptual system has for temporal discrepancies before segregation occurs. The concept of the TBW applies to various levels of MSI from cellular responses to human perception (e.g., Meredith *et al*., 1987; Spence and Squire, 2003; Wallace *et al*., 2004). For example, superadditive effects from bimodal stimulation on a multisensory cell may only occur if the second stimulus falls within the TBW (Meredith *et al*., 1987). On a perceptual level, lightning and thunder are perceived as a single event if the sound and light reach the observer close together in time, whereas distant storms produce segregated percepts. Perception of multimodal synchrony (and asynchrony)

can be measured in the laboratory with arbitrary stimuli using tasks such as
the simultaneity judgement (SJ) task, discussed below (e.g., Zampini *et al.*,
2005a).

Two important phenomena identified in MSI research are spatial and tem-
poral ventriloquism. In spatial ventriloquism a sound, such as a ventriloquist's
voice, is 'captured' by a visual stimulus, such as the dummy's correspond-
ing mouth movements, so that the sound appears to emanate from the same
location as the visual stimulus (Alais and Burr, 2004; Pick *et al.*, 1969). In
temporal ventriloquism, it is usually a visual stimulus that is captured by a
sound, and it is the perception of timing rather than the location in space that
is distorted (Fig. 1). The visual stimulus is perceived to occur earlier or later
than its physical presentation depending on whether the sound occurs before or
after it respectively (Fendrich and Corballis, 2001; Morein-Zamir *et al.*, 2003;
Scheier *et al.*, 1999). Both types of ventriloquism are examples of effects of
MSI on perception and lend support to the modality-appropriateness hypoth-
esis: the modality that dominates perception will be the one that provides the
most accurate information depending on the nature of the task at hand (Welch
and Warren, 1980; Welch *et al.*, 1986). In AV research, spatial tasks gener-
ally appear to be dominated by vision, whereas temporal tasks appear to be
dominated by audition. However, recent studies found that the dominance of
vision over audition in spatial ventriloquism (Alais and Burr, 2004), and audi-
tion over vision in temporal ventriloquism (Vidal, 2017), could be modulated
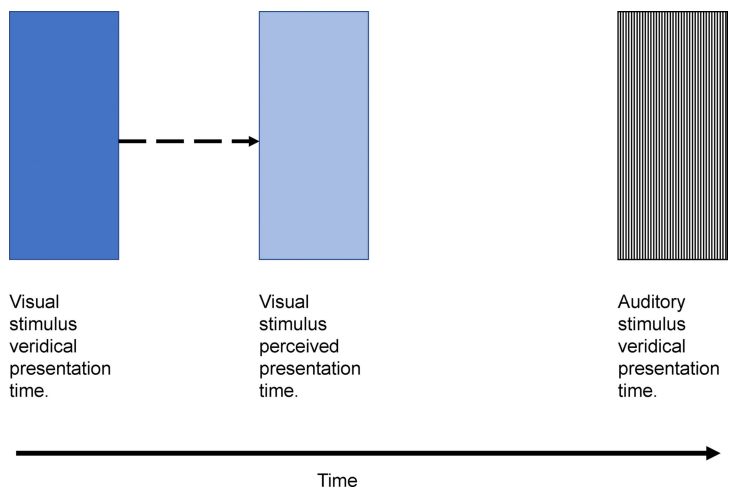by reducing the reliability of information afforded by the dominant stimulus.

Figure 1. Temporal ventriloquism. The perception of the visual stimulus is shifted in time
(dashed arrow) toward the auditory stimulus (black striped block). In this example the dark
blue block represents the actual time of presentation of the visual stimulus. The light blue block
represents the perceived time of the visual stimulus after capture by the auditory stimulus.

These results suggest that the perceptual system optimally integrates auditory and visual information (Alais and Burr, 2004).

The focus of this scoping review was on research from the last two decades into early crossmodal and multisensory temporal processing in humans, using simple auditory and visual stimuli. Our principal interest in this review was in the low-level (bottom up) stimulus driven factors, rather than higher-level (top down) cognitive factors, influencing multisensory processing. Although higher-level cognitive factors do play an important role in MSI, we opted to limit the scope of this review to early, low-level crossmodal interactions. This was because prior to the turn of the millennium, the prevailing view was that different modalities only interacted late in processing, yet the evidence since then (reviewed below) has largely refuted this view.

A scoping review can be used to quickly map key concepts, as well as sources and types of evidence in a research field (Mays *et al.*, 2001). Scoping reviews provide broad coverage of a field, but a limitation of the large breadth of scope is a sacrifice in the depth of discussion. This review was undertaken for two of the four potential reasons for carrying out a scoping review set out by Arksey and O'Malley (2005): "To examine the extent, range and nature of research activity" and "To summarize and disseminate research findings" (p. 21). The research question guiding this review was: What methods have been used to study low-level multisensory temporal processing in humans and how has this research informed our understanding of multisensory perception?

2. Methods

Our methods were based on the recommendations of Arksey and O'Malley (2005), with additional guidance from Pham *et al.* (2014). The five stages outlined by Arksey and O'Malley (2005) were followed: (1) identify the research question, (2) identify relevant studies, (3) select studies, (4) chart the data, and (5) collate, summarize, and report results. We did not include the optional consultation exercise.

2.1. Search Strategy, Study Identification/Selection, and Citation Management

The PubMed electronic database was used to identify relevant studies on 23/02/2016 using the search terms "(multisensory integration) AND (temporal OR timing)". This returned 426 articles. Initial inspection of the titles indicated a large body of research focused on speech and autism spectrum disorder (ASD). As we were concerned with early, low-level processing of simple stimuli we added "NOT speech" (in the title or abstract) to the search term, as speech stimuli are complex. This brought the number of returned articles to

256. After reviewing the titles and abstracts for relevance to the study, 97 full-text articles were selected and imported into the citation manager Mendeley (https://www.mendeley.com) for review. Additional relevant articles identified from reference lists in the selected articles were also stored for review. Mendeley periodically suggests articles based on content in the user's library and these were also considered for relevance. These steps led to 11 additional articles being included for review. On 02/05/2017, a subsequent search using the same search terms, but limited to articles published from the year of the initial search onwards, returned 55 articles. Twenty-six of these were selected for review, bringing the total number of articles to 134.

Some irrelevant studies were captured in the search, mostly due to the use of the term 'temporal'. Brief descriptions of the principles of MSI are common in the literature, so even when the research is not concerned with temporal factors the term is often mentioned. References to the term also occurred in irrelevant studies that referred to the temporal lobe of the brain. Due to time and personnel constraints, the scope of this review was limited to a subsection of the published (and e-published) temporal MSI research. We focused on human studies that utilized simple auditory and visual stimuli to study multisensory interactions at early stages of processing. It should be noted that timing plays an important role in the integration of other senses as well, but these were outside the scope of this review. Studies that were not concerned with low-level auditory-visual temporal processing were excluded. It should also be noted that our search terms may have been overly restrictive. Some pertinent studies that used terms such as 'crossmodal' or 'audiovisual' rather than 'multisensory integration' may not have been captured in the search.

2.2. Charting and Collating the Data

A spreadsheet was created to chart the studies (see Supplementary Material). Information was entered for authors, year of publication, study type, aims/concepts, methods/measures, area of interest, and a summary of results. Information on samples, tasks used, and brain regions identified was included where applicable.

3. Results

Of the 134 articles that met our initial search criteria, 89 were deemed relevant after examining the full text, and were selected for further review. Several key articles that had not been captured in our search were identified and added during the peer review process, bringing the total number included in this review to 107. The selection process is illustrated in Fig. 2. Of the articles selected, 12 were review articles and 95 were human and/or computational
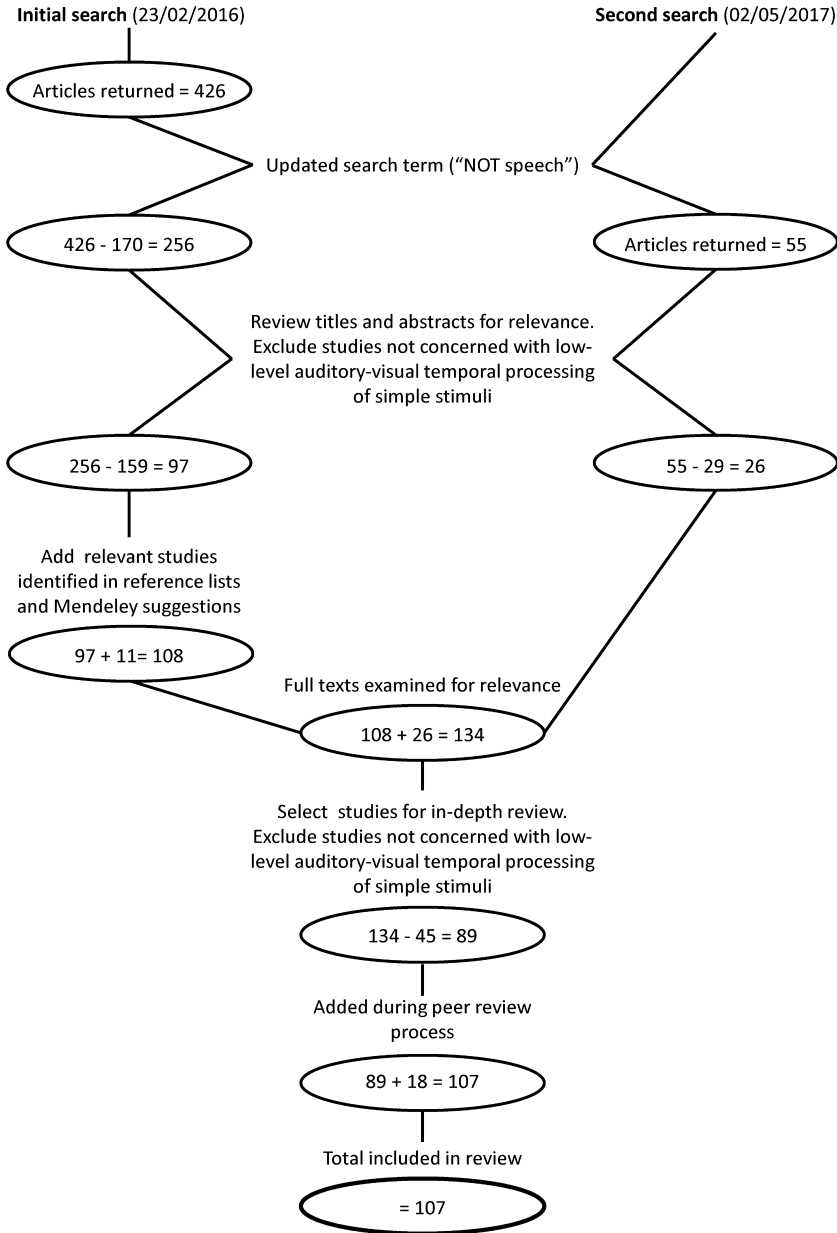
Initial search (23/02/2016)   **Second search** (02/05/2017)

Articles returned = 426

Updated search term ("NOT speech")

426 - 170 = 256

Articles returned = 55

Review titles and abstracts for relevance.
Exclude studies not concerned with low-
level auditory-visual temporal processing
of simple stimuli

256 - 159 = 97

55 - 29 = 26

Add relevant studies
identified in reference lists
and Mendeley suggestions

97 + 11= 108

Full texts examined for relevance

108 + 26 = 134

Select studies for in-depth review.
Exclude studies not concerned with low-
level auditory-visual temporal processing
of simple stimuli

134 - 45 = 89

Added during peer review
process

89 + 18 = 107

Total included in review

= 107

Figure 2. Illustration of the article selection process used in this review.

research studies (see Table 1 for a breakdown of article types and study con-
tent). Nineteen of the studies included a brain imaging or stimulation compo-
nent (electroencephalography [EEG], magnetoencephalography [MEG], func-
tional magnetic resonance imaging [fMRI], transcranial magnetic stimulation

Table 1.

Breakdown of article types and study content reviewed

	Number of articles
Article type	
Reviews	12
Research studies	95
Total	107
Study content	
Brain imaging/stimulation	19
Special population samples	18
Healthy adult samples only	73
Computational model studies	16

[TMS], and transcranial direct current stimulation [tDCS]). Eighteen of the studies included special population samples: musicians, the elderly, elderly fallers, synaesthetes, obese people, people with ASD, schizophrenia, unilateral spatial neglect (USN), and visual field defects (VFD). Seventy-three studies included healthy adult participants only (not specifically recruited for special ability or age). Of sixteen computational model studies, eleven included human participants. Models studied were the time-window-of-integration model (TWIN), maximum likelihood estimation (MLE), Bayesian causal inference, and the multisensory correlation detector (MCD). The most common behavioural tasks used were response time (RT), simultaneity judgment (SJ), and temporal order judgment (TOJ) tasks. The most commonly studied multisensory illusion was the sound-induced flash illusion (SIFI).

3.1. Review Articles

Review articles captured in the search covered various topics relating to MSI including the perception of synchrony of crossmodal signals (Keetels and Vroomen, 2012; King, 2005; Spence and Squire, 2003; Vroomen and Keetels, 2010), multisensory aspects in temporal perception (Fujisaki *et al.*, 2012), the occurrence of MSI in early cortical processing (Calvert and Thesen, 2004; Musacchia and Schroeder, 2009), neural mechanisms underlying MSI and cross modal interactions (Van Atteveldt *et al.*, 2014), the use of illusions to study multisensory processing (Recanzone, 2009), the spatial principle in human behaviour (Spence, 2013), temporal MSI in ASD (Chan *et al.*, 2016), and the unity assumption, which describes when a number of crossmodal signals are considered to emanate from the same event (Chen and Spence, 2017). Most of these topics are covered to some degree in this review. For more in-depth discussions on a topic, the reader is directed to the above reviews. The unity

assumption was beyond the scope of this review as it is a 'top down' influence on processing, and is not covered further.

3.2. Human Behavioural Studies

3.2.1. Multisensory Illusions

Illusions objectively demonstrate that sensory perception is not necessarily an accurate reflection of the physical world and they can be used to investigate the limits of the perceptual system. Illusions have been used to study both uni- and multisensory processing (Recanzone, 2009). Multisensory illusions can help elucidate the interplay between modalities, levels of processing, and how the brain integrates different senses to create a unified and adaptive perception of the physical world.

The flicker–flutter illusion (FFI) was an early example of auditory capture of a visual stimulus supporting the modality-appropriateness hypothesis (Gebhard and Mowbray, 1959; Welch and Warren, 1980; Welch *et al.*, 1986). In the FFI, streams of visual and auditory stimuli are presented concurrently but at different temporal rates. Observers' perceptions of the temporal rate of the visual stream tend to be biased towards the rate of the auditory stream (Gebhard and Mowbray, 1959). Recanzone (2003) found the expected auditory dominance in the FFI and also showed that audition dominated temporal ventriloquism aftereffects, analogous to the visual dominance of spatial ventriloquism aftereffects. No other illusion studies captured in our search utilised the FFI, with the majority opting instead for the related sound-induced flash illusion (SIFI) (Shams *et al.*, 2002).

In the SIFI, visual and auditory stimuli are presented concurrently, but the number of stimuli presented in each modality differs (Fig. 3). The perceived number of visual stimuli is altered towards the number of sounds presented (Shams *et al.*, 2000, 2002). For example, in the 'fission' version of the SIFI, a flash accompanied by two beeps will often be incorrectly perceived as two flashes. In the 'fusion' version, the visual percept is fused by the sound so that, for example, two flashes accompanied by one beep may be incorrectly perceived as a single flash (Andersen *et al.*, 2005). Increasing the complexity and familiarity of the stimuli used can reduce susceptibility to the illusion, indicating that high-level processing can modulate low-level MSI (Setti and Chan, 2011). Susceptibility can also be affected by low-level stimulus factors such as temporal relationships between stimuli, and can be further modulated by visual acuity. Kostaki and Vatakis (2016) found that the fusion illusion was stronger at shorter stimulus onset asynchronies (SOA), and when the beep preceded the first flash than when it followed. Those with lower visual acuity were more susceptible to the illusion. The SIFI has been utilised to study MSI in special populations (Bolognini *et al.*, 2016; McGovern *et al.*, 2014;
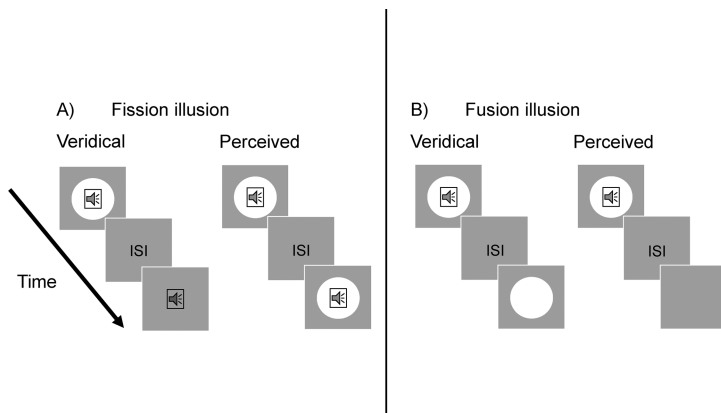
Figure 3. Sound-induced fission and fusion illusions. (A) In the fission illusion a single visual flash accompanied by two auditory tones is perceived as two flashes. (B) In the fusion illusion two visual flashes accompanied by a single auditory tone are perceived as one flash. The inter-stimulus interval (ISI) denotes the amount of time between the offset of the first stimuli and the onset of the second. Adapted from McGovern *et al.* (2014).

Neufeld *et al.*, 2012; Setti *et al.*, 2011, 2014), in computational model studies (Odegaard and Shams, 2016; Odegaard *et al.*, 2016), and in neuroimaging studies (Hamilton *et al.*, 2013; Keil *et al.*, 2014) and is discussed further in the appropriate sections.

Duration judgments can be influenced by MSI-driven illusory perceptions such as temporal ventriloquism. Romei *et al.* (2011) found that the perceived duration of visual stimuli was altered towards the duration of concurrently presented auditory stimuli (i.e., the visual stimulus appeared to be displayed for longer or shorter than its physical presentation time if paired with a longer or shorter auditory stimulus respectively). Bausenhart *et al.* (2014) found the same pattern for duration judgements of empty intervals marked by two visual stimuli presented concurrently with intervals marked by auditory stimuli. The effect was also apparent, although smaller, when visual stimuli acted as the accessory stimuli to auditory-marked intervals. The auditory dominance over visual perception often found in temporal experiments, such as those described above, supports the modality-appropriateness hypothesis. However, Vidal (2017) found that reducing the reliability of auditory stimuli, by introducing task-irrelevant pink noise, eliminated this dominance in interval judgement tasks. The effects of auditory capture were reduced to an extent that they did not differ from the effects of visual capture, suggesting that stimulus reliability dictated the dominant modality for the task, rather than 'appropriateness' (Alais and Burr, 2004; Vidal, 2017). Stimulus reliability may constrain the conditions under which the modality-appropriateness hypothesis holds.

Duration adaptation occurs as a result of repeated exposure to unvarying duration information, and produces repulsive aftereffects. If a subsequently presented test stimulus is shorter than a repeating adaptation stimulus, the perceived duration of the test stimulus is even shorter than its veridical duration. If it is longer than the adaptation stimulus, the perceived duration of the test stimulus is longer than its veridical duration. The perceived duration of a visual stimulus can also be lengthened or shortened through MSI, by concurrently presenting an auditory stimulus of longer or shorter length (Romei *et al.*, 2011). To determine whether duration adaptation occurred prior to MSI, Heron *et al.* (2013) distorted the perceived durations of visual adaptation stimuli with concurrently presented auditory stimuli. They found that the adaptation aftereffects on visual duration judgements were consistent with veridical rather than perceived durations of the adaptation stimuli. These results suggested that visual duration adaptation occurred earlier in sensory processing than MSI.

Temporal ventriloquism can enhance or diminish perception depending on the temporal relationships between auditory and visual stimuli. Vroomen and De Gelder (2004) demonstrated this with the flash-lag effect, where a flashed stimulus presented at the same location as a moving stimulus appears to lag behind it. The variability and magnitude of the effect was diminished if a sound was presented simultaneously with the flashed stimulus. A sound presented before the flash decreased the effect further, whereas a sound presented after the flash increased the effect relative to the synchronous condition. Variability remained similar in all conditions (all had sound). Simultaneous presentation sharpened temporal perception and temporal conflict appeared to draw the flashed visual stimulus toward the sound.

Studies that utilised illusions associated with higher-level processes such as feature binding and speech processing were outside of our focus on lower-level multisensory processes. However, some studies examined the interplay between low-level temporal processing and higher-level effects. In the stream/bounce illusion, two-dimensional images of a pair of identical balls are presented that travel along trajectories towards each other from one side of the display to the other, usually crossing in the centre of the display. The balls can be perceived as bouncing off one another, each reversing its original trajectory, or they can be perceived as moving past one another and continuing on their original trajectories (Fig. 4). Presenting a sound at the time the balls meet increases the likelihood of the 'bounce' perception (Sekuler *et al.*, 1997). Adaptation to AV lags (see temporal recalibration section) shifted the point of subjective simultaneity (PSS) in the direction of the lag (Fujisaki *et al.*, 2004). The PSS is an estimate of the temporal distance between stimuli at which they are perceptually bound together and experienced as synchronous. The same study found that AV lag adaptation also affected the proportion of bounce percepts at various SOA, suggesting that subjective simultaneity was
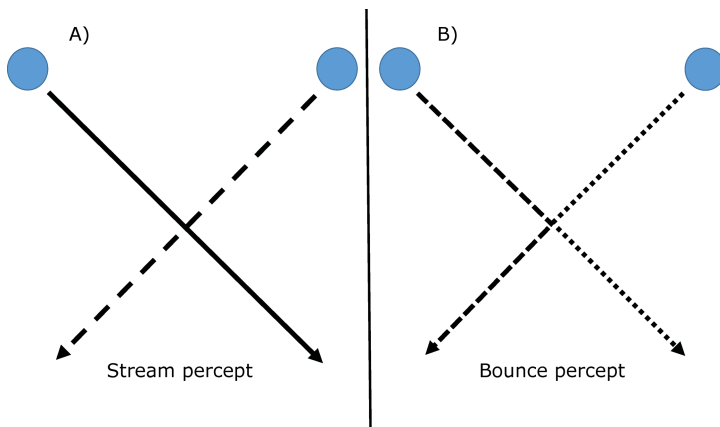
Figure 4. The stream/bounce illusion. Identical balls travel along trajectories toward each other. They can be perceived as (A) passing each other when they meet and continuing along the same trajectory (stream) or (B) bouncing of each other and changing trajectory. If a sound is introduced at the time the balls meet, the proportion of 'bounce' percepts is increased. Adapted from Shimojo and Shams (2001).

determined by early sensory processing rather than (or as well as) a higher-level cognitive shift. Bedard and Barnett-Cowan (2016) found evidence for a larger TBW in the elderly than the young in the stream/bounce illusion (i.e., they still perceived the collision at longer SOA than the young), although an earlier study found that the elderly were less susceptible to the bounce percept in the stream/bounce illusion overall (Roudaia *et al.*, 2013). Further research will be required to explain this discrepancy.

In the McGurk illusion, a video of a human face mouthing a syllable is presented with audio of a human voice saying a different syllable. The integration of the incongruent auditory and visual information can lead to the percept of an illusory third syllable (McGurk and MacDonald, 1976). Stevenson *et al.* (2012b) examined links between the TBW and MSI. They found that individuals with a wide TBW on the vision-leading side were more susceptible to the SIFI, whereas those with a narrow TBW on the vision-leading side were more susceptible to the McGurk illusion. These results suggested that those with a narrow TBW on the vision-leading side were better at multimodal temporal discrimination (indexed by the SIFI) and were also more likely to integrate synchronous stimuli (indexed by the McGurk illusion; Stevenson *et al.*, 2012b).

The multisensory illusion literature suggests that sound can alter the temporal perception of visual stimuli, and similarly visual stimulation can alter the spatial perception of sounds. Ostensibly, it appears that vision and sound dominate each other in the spatial and temporal realms, respectively. However, the

modality appropriateness-hypothesis is challenged when the reliability of the dominant stimulus signal is degraded, suggesting that the perceptual system optimally integrates signals across modalities (i.e., perception is influenced more by information from the more reliable signal) (Alais and Burr, 2004; Vidal, 2017). This idea of optimal integration has been successfully modelled in the maximum likelihood estimation (MLE) model (Ernst, 2006; Parise *et al.*, 2012) discussed below. Only one study captured in the search (Vidal, 2017) investigated the effects of degrading the auditory signal on visual temporal perception. This may be a fruitful area for future studies to explore, and they may consider using some of the multisensory illusions discussed above to do so.

3.2.2. Temporal Order Judgment and Simultaneity Judgment Tasks

In temporal order judgment (TOJ) tasks, two stimuli that differ along some categorical dimension are presented sequentially and the observer must indicate which was presented first (e.g., Love *et al.*, 2013; Morein-Zamir *et al.*, 2003; Stevenson and Wallace, 2013). Stimulus onset asynchrony (SOA) is varied to find the just noticeable difference (JND) — the smallest amount of time between the stimuli that an observer is sensitive to before their responding reaches chance levels. TOJ is a test of temporal discrimination, and can be used to assess both unisensory and multisensory temporal processing (Fig. 5).
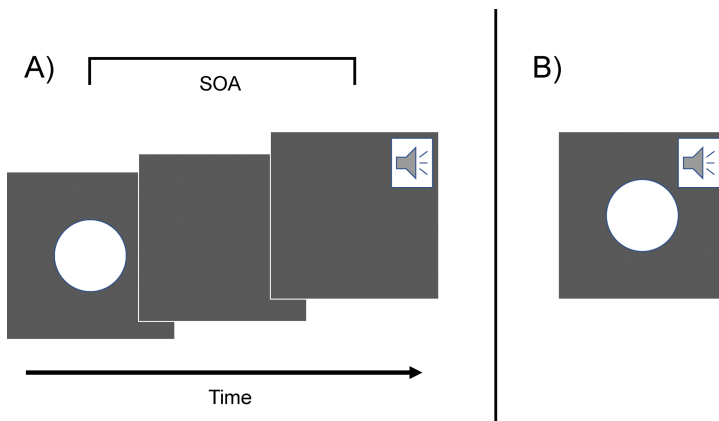

Figure 5. In audiovisual temporal order judgment (TOJ) tasks, observers indicate whether the auditory or visual stimulus came first. In simultaneity judgment (SJ) tasks, observers indicate whether the auditory and visual stimuli were presented synchronously or not. The white circle represents the visual stimulus and the speaker symbol represents the auditory stimulus in this example. TOJ and SJ tasks can use identical component stimuli. (A) In the SJ task the correct response to this trial would be 'asynchronous'. In the TOJ task, the correct response would be 'visual first'. A range of stimulus onset asynchronies (SOA) are used in both tasks. (B) Auditory and visual stimuli are presented simultaneously in the synchronous condition of the SJ task.

Simultaneity judgment (SJ) tasks are similar to TOJ tasks but must include a synchronous condition in the range of SOA used and observers report whether two stimuli were perceived as occurring simultaneously or not (Fig. 5) (e.g., Zampini *et al.*, 2005a). This is a test of sensitivity to asynchrony and the propensity to bind stimuli into a single percept. Both SJ and TOJ can be used to calculate estimates of an individual's PSS and TBW size.

There are at least two ways that multisensory processing can be assessed using TOJ. The first is using bimodal test stimuli to assess crossmodal temporal sensitivity (the observer judges which modality was presented first). The second way is by investigating the effect of accessory (task-unrelated) stimuli of a different modality on unimodal TOJ task performance (Barrett and Krumbholz, 2012; Morein-Zamir *et al.*, 2003). Including accessory auditory stimuli modulates performance on visual TOJ tasks depending on the temporal position of the accessory in relation to the test stimuli. Modulations have been attributed to temporal ventriloquism pulling the perception of the visual stimuli toward the sound (de Boer-Schellekens *et al.*, 2014; Morein-Zamir *et al.*, 2003). Temporal ventriloquism may be modulated by high-level factors; Parise and Spence (2008) found that synaesthetically congruent stimuli (low sounds presented with large visual stimuli or high sounds with small visual stimuli) produced larger effects of auditory capture than incongruent pairs in a TOJ task. However, Keetels and Vroomen (2011) reported that this effect disappeared when compared with a synchronous AV condition that controlled for response biases.

Greater tolerance for vision-leading AV asynchrony has often been reported in temporal multisensory literature: PSS is often vision-leading (see Vroomen and Keetels, 2010; Keetels and Vroomen, 2012 for reviews), although some studies have reported the opposite (Bald *et al.*, 1942; Rutschmann and Link, 1964). One group reported asymmetrical TBWs, where the vision-leading side of the window was wider than the auditory-leading side, meaning that visual-leading AV stimuli were bound at larger SOA than audio-leading stimuli (Stevenson *et al.*, 2012a, b). It makes intuitive sense for the perceptual system to be more tolerant of visual asynchronies, as visual information from an event will reach an observer before auditory information from the same event. If audition leads vision, the signals are unlikely to have emanated from the same event and the perceptual system should segregate them. However, other explanations have been proposed for asymmetries (see Keetels and Vroomen, 2012; Vroomen and Keetels, 2010) including (but not limited to) factors such as stimulus duration (Boenke *et al.*, 2009), attention (Zampini *et al.*, 2005b), and apparent causality (Kohlrausch *et al.*, 2013).

The extent to which underlying neural mechanisms are shared between SJ and TOJ tasks is not clear from current evidence. In our search, one study found that PSS were audio-leading for TOJ tasks, but video-leading for SJ

tasks (Love *et al.*, 2013), another found TOJ to have a video-leading PSS (Bedard and Barnett-Cowan, 2016), and a third found no significant differences in PSS between tasks (Binder, 2015). Furthermore, Van Eijk *et al.* (2008) found no correlation between the PSS of SJ and TOJ tasks. Similarly, Love *et al.* (2013) found no correlations between the tasks on either PSS or TBW performance measures, whereas Bedard and Barnett-Cowan (2016) found correlations in TBW size in a young group, and task-selective age-related impairments in an elderly group.

fMRI blood oxygen level-dependent (BOLD) measurements revealed that the SJ task did not show any regions where activity was greater than in the TOJ task (Binder, 2015). The TOJ task showed activity additional to the common areas between the tasks. Extra activity was confined to the left hemisphere in the frontal, parietal, and temporo-occipital areas. These results supported a two-stage model of TOJ where simultaneity is determined before temporal ordering of sensory events. This suggested that some neural mechanisms were shared between SJ and TOJ tasks but the latter required additional processing. More research will be required to confirm this and resolve discrepancies in behavioural results.

Stimulus characteristics affect performance on both SJ and TOJ tasks, and similarities in these effects further support the argument for shared mechanisms between them. Stevenson and Wallace (2013) found that TBW size differed between the two tasks but was larger for speech than non-speech stimuli in both. Correlations between each task-stimulus condition revealed a high degree of consistency within subjects. Leone and McCourt (2015) found that the PSS occurred at larger visual-leading asynchronies in both tasks for scotopic stimuli that activate only the rod photoreceptors in the retina, than phototopic stimuli that activate cones and saturate rods. Low-intensity AV stimuli have been reported to produce more perceptions of synchrony in the SJ task than high-intensity AV stimuli (Krueger Fister *et al.*, 2016). These studies suggested that the tasks were similarly affected by low- and high-level factors such as stimulus characteristics and stimulus complexity, respectively. Understanding the effects of stimulus choice has implications for study design as well as for comparisons between tasks and studies. Stimulus choice and presentation conditions may vary depending on the research question studied but attempts should be made to equate them with relevant studies (or account for differences) if valid comparisons are to be drawn.

The brain is highly adaptable to different environments and situations. Brain plasticity refers to changes induced by experience. TOJ and SJ tasks have proven useful in studying the adaptability and plasticity of the perceptual system, which may provide insight into these processes more generally throughout the brain. Mégevand *et al.* (2013) suggested that the TBW was adaptable and changed according to task demands. They showed that MSI in a

response time (RT) task and segregation in a TOJ task could occur at the same SOA. Other studies found that feedback training on TOJ tasks narrowed the TBW, and the improved temporal precision generalised to other tasks (Setti *et al.*, 2014; Stevenson *et al.*, 2013).

TOJ training has also been used to assess whether temporal perception is governed by a supramodal 'clock' mechanism or a set of modality-specific clocks (Alais and Cass, 2010). Participants were tested on auditory, visual and AV TOJ tasks before and after training on only one of the conditions. Each group improved in the condition that they were trained on. Only training on the AV task transferred; this group also showed post-training improvement on the visual task. Visual and AV training improvements also generalised to visual TOJs that utilised differently oriented grating stimuli to the trained stimuli, and to stimuli presented at a different location to training stimuli. Auditory training did not transfer to test tones that differed in pitch to the trained stimuli. The authors concluded that both supramodal and modality-specific clocks exist that are independent or quasi-independent of each other.

The SJ and TOJ studies described above indicate that the perception of cross modal simultaneity and temporal discrimination can differ between, and even within different individuals depending on task demands and stimulus characteristics. Temporal sensitivity appears to be not only subjective, but fluid — a point further evidenced in the temporal recalibration section below.

3.2.3. Temporal Recalibration

Repeated exposure to consistent temporal relationships between auditory and visual stimuli (AV lags) resulted in adaptation effects that caused the PSS for SJ and TOJ tasks to shift in the direction of the lag (Fujisaki *et al.*, 2004; Vroomen *et al.*, 2004). Navarra *et al.* (2013) reported that the shift occurred even when component auditory and visual stimuli were presented at different locations, as well as at SOA outside of the TBW. In their study the shifts were only observed after exposure to vision-leading AV stimuli. They suggested that the perceptual system may be more likely to recalibrate to vision-leading AV temporal asynchrony, as it is the order more frequently experienced in everyday life due to the faster transmission time of light compared to sound (Navarra *et al.*, 2013).

Yarrow *et al.* (2011) found that spatial cues could be used to group AV stimulus pairs when temporal cues were ambiguous and this influenced recalibration effects. Streams of auditory and visual stimuli were presented at a constant rate, and the side of presentation alternated after each modality had been presented. The mean PSS was larger (56 ms) when spatial grouping implied vision-leading pairs, than vision-lagging pairs (19.5 ms; PSS was vision-leading for both). The grouping of stimuli has implications for maintaining the perception of synchrony of multiple concurrent events that may

occur at different distances (and therefore will have different auditory and visual arrival time discrepancies).

Two studies found evidence to suggest that multiple recalibrations could be maintained concurrently if stimuli could be grouped by featural correspondence (e.g., male voice and male face speaking a syllable; Roseboom and Arnold, 2011), or spatial correspondence (Heron *et al.*, 2012). Participants adapted to AV stimuli with opposing temporal lags divided along featural (e.g., male speaker/vision-leading, female speaker/audio-leading) or spatial dimensions (e.g., left AV pair/audio-leading, right/vision-leading). Concurrent, opposing PSS shifts were observed in both studies; they differed for the male and female identities (Roseboom and Arnold, 2011), and for the left and right pairs (Heron *et al.*, 2012).

However, Heron *et al.* (2012) did not find evidence for featural constraints when opposing AV lags were presented from the same (central) location. They used visual sine gratings with high and low spatial frequencies paired with high and low pitched tones. They postulated that their relatively arbitrary contextual cues could not replicate the effects seen in Roseboom and Arnold (2011) because links between them were not as strong as higher-level links between faces and voices. However, they also pointed out that Roseboom and Arnold (2011) always presented the same pairing from the same side in the adaptation trials. This meant that both spatial and contextual cues were available and so the extent to which each played a role in the observed effects was unclear.

To reconcile the differences between the studies Roseboom *et al.* (2013) ran a similar experiment where they presented male and female speaker identities from the same (central) location. They ran a second experiment with arbitrary pairings similar to Heron *et al.* (2012), except the visual stimuli differed by orientation rather than spatial frequency, and there was a bigger difference in the pitch of the tones. Interestingly, in both experiments opposing recalibration effects were observed, suggesting that contextual cues alone were enough to constrain recalibration effects. They proposed three possible reasons why their arbitrary stimuli were able to produce these effects when those in Heron *et al.* (2012) could not: (1) the differences between unimodal stimuli were more salient, (2) a larger interval between the presentation of AV stimulus pairs in the adaptation phase, and (3) explicit information was given to participants on which stimuli formed a pair. All of these reasons could theoretically result in a stronger pairing of the auditory and visual stimuli.

Another recent advancement in the literature was the discovery that recalibration does not require prolonged periods of adaptation. Van der Burg *et al.* (2013) found that temporal perception in a given trial in an SJ task was influenced by exposure to the temporal relationship between stimuli in the preceding trial. The PSS on a given trial was shifted in the direction of the

preceding trial. The subsequent shift was larger and occurred over a greater range of SOA for vision-led than audio-led pairs. A later study found that both prolonged and rapid recalibration effects could occur concurrently (Van der Burg *et al*., 2015). Prolonged recalibration decayed over time but did not affect rapid recalibration, suggesting that the two effects were independent of each other. Rapid recalibration effects appear to be contingent upon the preceding trial's veridical rather than perceived temporal order, which suggests that it is an early, sensory effect not a higher-level effect (Van der Burg *et al*., 2018). Rapid recalibration effects have been shown to change throughout the lifespan, with children and the elderly showing larger effects than younger adults (Noel *et al*., 2016). Children and adolescents with ASD were found to have abnormal rapid recalibration (smaller changes in the PSS) for simple and complex non-speech stimuli compared to controls, but groups performed similarly for speech stimuli (Noel *et al*., 2017). Rapid recalibration highlights the dynamic and flexible nature of our sensory processing system that allows for coherent perception in a constantly changing physical world.

3.2.4. Response Time Tasks

RT tasks provide a simple method for assessing behavioural multisensory response enhancement (MRE). MRE describes better performance on tasks that utilise multimodal stimuli compared to unimodal versions of the same tasks. In multimodal RT tasks, MRE is said to occur if the race-model is violated. The race-model posits that different modalities are processed separately and multimodal processing times are determined by the fastest of the modalities (Calvert and Thesen, 2004; Miller, 1982; Raab, 1962). If the multisensory RT is faster than that predicted by the summed probabilities of the unisensory RT, then the race-model is violated. On the other hand, coactivation models allow input from all component signals to influence processing, and are generally favoured in recent literature (Calvert and Thesen, 2004).

The timing and location of the components of multimodal stimuli are important factors for MRE. RTs were faster, and sensitivity greater in a target detection task for synchronously presented AV stimulus trains than asynchronous trains (Marchant *et al*., 2012). Results from tasks requiring an orienting response suggested that component stimuli presented synchronously (Ho *et al*., 2013) and in the same location (Steenken *et al*., 2014) produced the greatest MRE. This has implications for driver assistance systems that may benefit from multimodal displays that adhere to the principles of MSI, over unimodal displays.

A recent study examined whether MRE of RTs occurred when asynchronous auditory and visual stimulus pairs were perceived as synchronous, or when they were physically synchronous (Harrar *et al*., 2017). Perception was manipulated through adaptation to AV lags. In unisensory trials, RTs sped

up or slowed down in the direction predicted by changes in perceived synchrony after AV adaptation. In multimodal trials, MRE only occurred when AV pairs were physically synchronous, and not when asynchronous pairs were perceived as synchronous. The authors suggested that perception of synchrony is independent of multisensory integration, since the former was affected by AV adaptation but MRE was not. They proposed that, for action, there may be early processes unaffected by adaptation, and for perception, later processes that are susceptible to adaptation (Harrar *et al.*, 2017).

There is debate as to whether MRE occurs due to MSI or crossmodal temporal preparation (Los and Van der Burg, 2013; Van der Stoep *et al.*, 2015). Temporal preparation refers to when one stimulus increases arousal or directs attentional resources to enhance processing of the second stimulus. Los and Van der Burg (2013) used a target detection RT task to assess this. A warning stimulus (S1) and target visual stimulus (S2) were presented at different SOA. At short SOA, S2 detection times were faster if S1 was auditory than if it was visual. However, they also found that auditory targets were detected faster than visual targets when only one stimulus was presented per trial. Therefore, an auditory S1 may result in a more effective preparation period in multimodal trials. When the auditory component of multimodal trials was delayed (to control for the discrepancy in S1 detection times) the intersensory facilitation effect vanished, suggesting that facilitation in the initial experiment was driven by temporal preparation rather than MSI (Los and Van der Burg, 2013).

However, the driver of multisensory facilitation may be task-dependent. Van der Stoep *et al.* (2015) assessed RT on unimodal auditory and visual target detection tasks (baseline), a bimodal redundant-target task, and a bimodal spatial cueing task. Redundant-target tasks are thought to measure MSI (Miller, 1986), whereas spatial cueing is thought to measure exogenous spatial attention (Ward, 1994). In the bimodal tasks, stimuli could be spatially aligned or misaligned, and asynchronous trials were always audio-leading. In the redundant-target task, participants responded to the first target stimulus detected in each trial, regardless of modality. RT increased as SOA increased and MRE occurred at 0 and 50 ms SOA but not at 100 or 200 ms. In the cueing task, participants responded to the visual stimulus only, and the auditory cue acted as a valid or invalid spatial cue on spatially aligned and misaligned trials respectively. RT decreased as SOA increased and MRE occurred at 50, 100 and 200 ms SOA. The effect of spatial alignment (faster RT for aligned than misaligned stimuli) increased as SOA increased for the cueing task, but the effect remained constant across SOA in the redundant-target task. The opposite pattern of results in the two tasks suggested that MRE was driven by MSI at 0 ms, crossmodal exogenous spatial attention at 100 and 200 ms, and a combination of the two at 50 ms SOA (Van der Stoep *et al.*, 2015). These results

suggested that MSI could produce MRE even when temporal preparation was unlikely, as participants could respond to whichever stimulus was first in the redundant-target task, whereas Los and Van der Burg (2013) measured facilitation of responses to the slower (visual) signal only. Furthermore, temporal preparation effects tend to increase with SOA, as was apparent in the cueing task, yet the redundant-target task did not follow this pattern.

The studies reviewed in this section suggest that crossmodal interactions and MSI can both facilitate RT. While spatial relationships between component multimodal stimuli can modulate MRE, temporal relationships may interact with task demands to dictate the underlying mechanisms that drive MRE.

3.2.5. Localization Tasks

In spatial localization tasks, observers indicate the perceived position of a target stimulus, or the perceived or expected termination point of a moving stimulus. In bimodal (AV) localization tasks, one modality serves as the target while the other serves as a distractor or accessory. Temporal relationships between auditory and visual stimuli affected the speed of target localization and perceived target location (Van Wanrooij *et al*., 2009). Spatially and temporally aligned AV stimuli produced more accurate localization performance than unimodal stimuli. RTs were improved in bimodal trials relative to visual trials for visual-leading stimuli, and slowed relative to auditory-only trials for audio-leading bimodal trials (Van Wanrooij *et al*., 2009). For spatially unaligned stimuli, biasing effects arising from crossmodal interactions and the spatial ventriloquist effect were modulated by temporal discrepancies between component stimuli (Schmiedchen *et al*., 2012; McGovern *et al*., 2016). Spatial ventriloquism effects tend to be smaller at larger SOA, and training-induced narrowing of the TBW (through an SJ task with feedback) improved spatial localization and reduced the ventriloquism effect (McGovern *et al*., 2016). Relative timing between crossmodal signals is a factor in solving the correspondence problem, so improving temporal discrimination can affect MSI of spatial cues (McGovern *et al*., 2016).

Correlations between the temporal structures of crossmodal signals may be more important than synchrony in solving the correspondence problem (Denison *et al*., 2013). Trains of auditory and visual stimuli with correlated complex temporal structures are likely to have emanated from the same event, even if they are asynchronous. Spatial ventriloquism effects were enhanced for spatially disparate and asynchronous AV trains if they had correlated temporal structures (Denison *et al*., 2013; Parise *et al*., 2012, 2013).

Van der Burg *et al*. (2008) found that a non-spatial auditory cue aided visual search in a cluttered display, reducing RTs. Participants searched for either a

horizontal or vertical target among sets of 24 to 48 distractors at other orientations. Targets and distractors changed colour continuously between red and green. Participants had to indicate the orientation of the target. RTs were facilitated when the auditory 'pip' and the colour change of the target were synchronized. The authors proposed that MSI of the pip and the colour change caused the target to pop out from the display. Follow-up experiments with the pip and target change presented at SOA between –150 and 100 ms showed that the closer to synchronous the AV pairs occurred, the greater the RT facilitation for both auditory- and visual-leading pairs. This suggested that the 'pip and pop' effect was not due to simple cueing, which would be expected to produce greater RT facilitation at larger auditory-leading SOA (due to more preparation time), and would not be possible when auditory stimuli occurred after the colour change. Rather, the enhanced localization was likely due to an increase in saliency due to MSI of the target and auditory cue.

While spatial congruency may not be vital in temporal AV tasks (Spence, 2013), the studies captured in our search suggest that temporal relationships between auditory and visual stimuli play an important role in AV spatial localization tasks.

3.3. Special Populations

A common finding in many of the studies captured in the search was a wider TBW for special populations than healthy controls on at least some of the measures and tasks used. This was true for elderly people (Bedard and Barnett-Cowan, 2016; de Boer-Schellekens and Vroomen, 2014; McGovern *et al.*, 2014), especially those with a history of falls (Setti *et al.*, 2011), as well as people with ASD (de Boer-Schellekens *et al.*, 2013; Kwakye *et al.*, 2011), and obese people (Scarpina *et al.*, 2016). As the wide TBW has been shown to be prevalent in various special populations, assessments of temporal processing of multisensory stimuli may have potential to be developed into tools for predicting or identifying neurological conditions and cognitive decline.

3.3.1. Elderly

Only one study captured in the search found no differences between young and elderly groups in crossmodal temporal sensitivity, where both groups performed similarly on a TOJ task (Fiacconi *et al.*, 2013). Another study that found that the elderly performed worse than young adults on a TOJ task, but also found that both groups performed similarly on an SJ task (Bedard and Barnett-Cowan, 2016). The difference in TOJ results between studies may have been due to the SOA range used in each experiment. Fiacconi *et al.* (2013) used only 0, 50, 100, and 250 ms SOA, whereas Bedard and Barnett-Cowan (2016) used a more sensitive measure with 0, 25, 50, 100, 150, 200, 300 ms SOA for both tasks. The lack of age-related differences in the SJ task

suggests that either the two tasks are governed by different processes or that TOJ requires another stage of processing that is prone to age-related decline, supporting the two-stage model described above.

Setti *et al*. (2014) found that the size of the TBW in the TOJ for an elderly group predicted susceptibility to the fission SIFI. The elderly were shown to be more susceptible to the fission illusion than the young at longer SOA, especially if they had a history of falls (McGovern *et al*., 2014; Setti *et al*., 2011). Setti *et al*. (2011) suggested that this was due to inefficient multisensory processing rather than visual acuity because fallers, non-fallers, and young groups performed similarly on unimodal and congruent multimodal trials. Increased fission illusion susceptibility appears to be related to the larger TBW in the elderly but can be changed through training on a TOJ task, indicating that multisensory temporal processing remains plastic into old age (Setti *et al*., 2014).

McGovern *et al*. (2014) found that the elderly were also more prone to the SIFI fusion illusion than the young at short SOA. However, the elderly group often reported only one flash on congruent two-flash–two-beep trials, indicating that slowed visual processing might account for fusion results rather than MSI. Fusion results did not differ between groups when differences in visual processing speeds were controlled for. These results indicated that susceptibility to the fission illusion increases with age but susceptibility to the fusion illusion remains stable. Bolognini *et al*. (2016) also found different results for the two illusions in the elderly but their results differed slightly from McGovern *et al*. (2014). The elderly were susceptible to the fission illusion but not the fusion illusion. The studies differed in their stimulus presentation. McGovern *et al*. (2014) always had one or two flashes that could be accompanied by zero to two beeps, whereas in Bolognini *et al*. (2016) the number of flashes and beeps could both vary between zero and four in each trial. SOA also differed between studies. Overall, both studies found a greater susceptibility to the fission illusion than the fusion illusion in the elderly. This suggests that the two illusions may be driven by different neural networks that are differently affected by aging (McGovern *et al*., 2014).

MSI can enhance sensory processing in the elderly and improve performance in certain tasks. In a visual TOJ task people over 50 years old performed worse than people under 50 on visual alone trials but showed greater improvements with the addition of task irrelevant sounds (de Boer-Schellekens and Vroomen, 2014). In a saccadic RT task, Diederich *et al*. (2008) found that MRE in bimodal conditions produced greater benefits for an elderly relative to a young group, when compared to unimodal conditions. However, even in bimodal conditions the elderly group were slower than the young in unimodal conditions. Weaker effects of MSI in the elderly than the young were shown

in a study that found RT MRE for the young but not the elderly group in a target detection task, and a smaller multisensory response in superior temporal gyrus for the elderly group as measured by magnetoencephalography (MEG) (Stephen *et al.*, 2010).

The studies described in this section suggest that the elderly can benefit from MSI but multimodal temporal sensitivity is impaired in normal aging. Some processes appear to be more resilient to the effects of aging than others. The plasticity of the TBW in the elderly demonstrated by Setti *et al.* (2014) suggests a potential for sensory training interventions to combat age-related sensory decline. Multisensory processing deficits may be indicative of other conditions that arise from age-related neural decline such as the propensity for falls (Setti *et al.*, 2011), and multisensory tasks may have the potential to be developed into simple and non-invasive tests for these.

3.3.2. Schizophrenia

De Boer-Schellekens *et al.* (2014) found that people with schizophrenia had impaired visual temporal processing compared to controls in a visual TOJ task. The addition of task-irrelevant sounds produced similar amounts of improvement in schizophrenia and control groups. These results indicated that early MSI processing was not impaired in the schizophrenia group (de Boer-Schellekens *et al.*, 2014).

3.3.3. Unilateral Spatial Neglect and Visual Field Disorders

In line with the aging research described above (McGovern *et al.*, 2014), Bolognini *et al.* (2016) found evidence supporting different underlying networks for the fusion and fission SIFI in patients with brain lesions. VFD result from primary visual cortex lesions, whereas USN results from lesions in fronto-temporo-parietal 'association' areas in the right hemisphere (Bolognini *et al.*, 2016). Regardless of the side of lesion in the occipital cortex, the fission illusion was reduced in the VFD group compared to controls and the USN group. The larger the lesion, the more the fission illusion was disrupted. Only the USN group reliably perceived the fusion illusion. A control experiment suggested that the fusion illusion is lost in healthy ageing, whereas the fission illusion endures into old age. The dissociable effects found between the groups suggested that the illusions were governed by different networks.

3.3.4. Synaesthesia

Neufeld *et al.* (2012) unexpectedly found evidence for a smaller TBW and less susceptibility to the SIFI in synaesthetes than in controls. In contrast, a previous study had reported greater susceptibility to the SIFI in synaesthetes, but did not include a temporal manipulation, so TBW size was not assessed (Brang *et al.*, 2012). These two studies provided conflicting evidence on whether synaesthesia results from a hypersensitive general binding

mechanism (Neufeld *et al.*, 2012). However, Brang *et al.* (2012) tested only grapheme–colour synaesthetes. On the other hand, Neufeld *et al.* (2012) tested grapheme–colour and auditory–visual synaesthetes, as well as people with both of these types of synaesthesia. The SIFI may not be a suitable tool for studying general underlying mechanisms of the heterogeneous condition, as the studies described in this section produced opposite results depending on whether the sample included only one type of synaesthesia or not. However, it may be useful in understanding underlying mechanisms within specific synaesthesia types.

3.3.5. Musical Training

Musicians were more sensitive to multisensory temporal asynchrony than non-musicians, and this was reflected in differential (MEG and fMRI) brain activity between the groups (Hodges *et al.*, 2005; Lu *et al.*, 2014). Musicians were more sensitive to auditory TOJ but performed similarly to non-musicians on visual TOJ tasks, even when sounds were included (Hodges *et al.*, 2005). On the other hand, musicians were more accurate than non-musicians at AV SJ (Lu *et al.*, 2014). MEG and behavioural results showed that musical training affects even early stages of AV temporal processing (Lu *et al.*, 2014). Paraskevopoulos *et al.* (2014) trained non-musicians to recognize deviations from a set of standard melodies with visual stimuli similar to musical notation and auditory five-note melodies, and compared post-training MEG activity with (long-term trained) musicians. Training occurred in either multisensory AV or separate auditory and visual conditions. After training, the AV-trained group showed differential activity to the separately-trained group on deviant trials, suggesting a crucial role for AV integration in plasticity. Musicians showed greater activity in the left superior temporal gyrus than both groups of non-musicians, and in anterior prefrontal regions compared to the AV-trained non-musician group.

Musical training can benefit auditory and crossmodal temporal precision. Information on the musical training history of participants may be useful during recruitment and when interpreting results in studies of temporal perception that are not specifically interested in musicians.

3.4. Human Neuroimaging and Brain Stimulation Studies

Traditionally, MSI was thought to occur only late in processing, in frontal and parietal association regions but this view has been largely refuted by recent evidence. Neuroimaging and brain stimulation research supports the case for crossmodal interactions and MSI at multiple stages of sensory processing, driven by neural networks that can differ according to the task at hand and include sensory cortices (Calvert and Thesen, 2004). Magnetic stimulation delivered to the visual cortex can elicit the perception of visual flashes

known as phosphenes (d'Arsonval, 1896; Thompson, 1910). An auditory stimulus presented 40 ms prior to the TMS pulse and spatially coincident with the expected phosphene location lowered the threshold of TMS intensity required for phosphene induction (Bolognini *et al*., 2010). These early interactions were greater in the periphery than the central visual field and the timing between the auditory stimulus and TMS pulse was critical. Phosphene perception did not significantly increase at any of the other interstimulus intervals (ISI) tested. Spatially inconsistent auditory stimuli produced no significant increases in phosphene perception. These results suggested that the auditory stimulus enhanced the excitability of the visual cortex in a spatially specific manner. The authors proposed that these effects may be due to anatomical connectivity between auditory and visual cortices (Bolognini *et al*., 2010) as these connections are stronger in peripheral than central visual field representations (Falchier *et al*., 2002).

The temporal precision of EEG and MEG makes them well suited to studying the time course of crossmodal interactions and MSI. Inferences can be made as to whether crossmodal effects are governed by sensory or cognitive processes depending on the latencies of components modulated by multimodal stimuli. The flash-lag effect describes a visual phenomenon where a stationary stimulus that is flashed onto a moving stimulus appears to lag behind it, even though both are presented at the same position at the same time (Nijhawan, 2002). When auditory stimuli preceded the flash, the flash-lag effect was reduced compared to visual-only trials (Stekelenburg and Vroomen, 2005). This temporal ventriloquism effect was reflected in modulations of the early visual N1 event-related potential (ERP) component, suggesting that it was governed by sensory processing rather than a response bias.

Evidence for MSI can be found in ERP at various SOA but the time courses and topographies of modulations may differ (Liu *et al*., 2011). Franciotti *et al*. (2011) found that event-related field (ERF) peak latencies of dipoles in both the primary auditory and primary visual cortex were longer for multimodal than unimodal stimuli. Peak latencies were further modulated by AV synchrony (Franciotti *et al*., 2011). Temporally congruent AV stimuli have been shown to enhance both auditory and visual steady-state evoked potential amplitudes relative to incongruent AV stimuli (Nozaradan *et al*., 2012). Temporally incongruent auditory stimuli can reduce visual stimulus locking of EEG (Schall *et al*., 2009). These results highlight the importance of temporal precision between the unimodal components of crossmodal stimuli on crossmodal interactions and MSI.

Oscillatory activity may be a mechanism by which the brain integrates information across uni- and multisensory regions (Senkowski *et al*., 2008). A MEG study revealed that perception of the fission SIFI was associated with increased ERF activity in the cingulate cortex but also greater beta-band power

before sound onset (Keil *et al.*, 2014). A complex pattern of alpha- and beta-band phase synchrony was observed when the illusion occurred. Another study found that early evoked gamma-band responses (GBR) were modulated by AV asynchrony regardless of the attended modality (Senkowski *et al.*, 2007). GBR may be related to temporal asynchrony perception processes that are independent of unimodal attentional processes (Senkowski *et al.*, 2007).

Neural oscillations also appear to play a role in temporal recalibration. Kösem *et al.* (2014) found that the phase of an entrained 1 Hz oscillatory MEG component shifted in different directions for auditory-led and visual-led lag adaptations but remained unchanged in a synchronous condition. Phase shifts in the auditory but not the visual cortex correlated with individuals' subjective perception (PSS) in an AV TOJ task. Perceived simultaneity linearly mapped onto the phase of auditory cortex oscillations. These results suggested that sensory cortices may use oscillatory mechanisms to encode and organize auditory and visual signals in time.

Many of the imaging and brain stimulation studies captured in the search supported evidence for the involvement of regions traditionally considered multisensory association areas (in the parietal and superior temporal cortices) in multisensory temporal processing. For example, Zmigrod and Zmigrod (2015) found that anodal transcranial direct current stimulation (tDCS) over the right posterior parietal cortex reduced the perception of synchrony in an SJ task. The angular gyrus (AG) was implicated in multisensory temporal processing as TMS over this region disrupted perception of the fission SIFI, whereas TMS to the supramarginal gyrus and vertex produced no reductions (Hamilton *et al.*, 2013). However, the association areas appear to work in concert with earlier processing areas. For example, SJ training resulted in a narrowing of the TBW that was reflected in decreased blood oxygen level-dependent (BOLD) responses in the posterior superior temporal sulcus, auditory cortex and visual cortex, as well as increases in coupling between these regions (Powers *et al.*, 2012). Another study presented bimodal streams of auditory and visual stimuli in a combination SJ/TOJ task (Dhamala *et al.*, 2007). Asynchronous perceptions were associated with a network of activation between the primary sensory, prefrontal, and inferior parietal cortices. For synchronous perceptions, the superior colliculus was engaged in the network rather than the inferior parietal cortex.

Modern neuroscience has moved away from a compartmental view of the brain towards the study of interregional neural networks (Medaglia *et al.*, 2015; Sporns, 2014). With the development of new techniques, evidence for different task-dependent networks supporting MSI will continue to increase our knowledge and understanding of the sensory processing system. The neuroimaging and brain stimulation studies captured in the search provide an example of how regions such as the auditory and visual cortices, once thought

to serve a limited purpose in sensory processing, are involved in complex inter-actions with each other and other regions from very early stages of processing. The study of perception is well established in unisensory research and is rel-atively easy to study objectively, compared to other cognitive systems. The study of multisensory networks may be a useful starting point towards a better understanding of cognitive neural networks more generally.

3.5. Models of MSI

Computational models have been developed that attempt to describe percep-tual processing on various multisensory tasks and illusions.

3.5.1. Time-Window-of-Integration

The TWIN model was developed to capture the effects of temporal rela-tionships between modalities in eye-orienting (saccadic) responses that pre-vious models of multisensory spatial integration were lacking (Colonius and Diederich, 2004). The model assumes two stages of processing. In the first stage, peripheral neural signals from different modalities 'race' each other. In the second stage, neural integration occurs and a response is prepared. Inte-gration only occurs if the peripheral processes (first stage) terminate within a certain time interval (Fig. 6). In this way, the TBW acts as a filter determining whether signals occurred close enough in time to be integrated (Colonius and Diederich, 2004). The size of the window is flexible and plastic and can be af-fected by prior knowledge and the probability of spatial alignment between component stimuli (Colonius and Diederich, 2011). Behavioural (saccadic RT) data from target detection tasks with visual and tactile stimuli supported predictions of the model (Colonius and Diederich, 2004). Unlike previous models of RT, the TWIN model included assumptions about target versus ac-cessory modalities. It made different predictions for different tasks, such as the focused attention versus the redundant target paradigm, due to its two-stage processing (Diederich and Colonius, 2009). The model has been applied to investigate a range of MSI research topics.
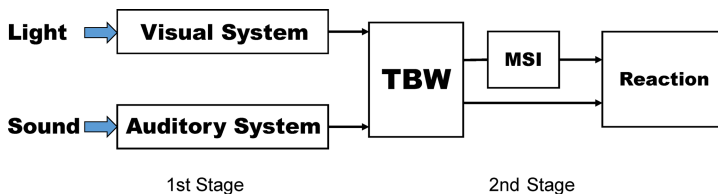

Figure 6. Two-stage schema for the TWIN model. In the first stage, multimodal signals race from the periphery. In the second stage, neural integration occurs and a response is prepared. In-tegration only occurs if the peripheral processes terminate within the temporal binding window. Adapted from Colonius and Diederich (2004).

Data from an AV-focused attention task fitted to the TWIN model implied that peripheral sensory processing took longer for the elderly than the young (Diederich *et al.*, 2008). The elderly had a wider TBW yet a lower probability of integration than the young. The model suggested that this was due to peripheral slowing in early processing rather than cognitive decline. Multisensory enhancements in the elderly may not be enough to fully compensate for peripheral slowing (Diederich *et al.*, 2008). The model was reasonably accurate at predicting behavioural data from a saccadic RT task in a simulated driving environment (Steenken *et al.*, 2014). Existing data (from Mégevand *et al.*, 2013) fitted to the TWIN model confirmed that the temporal window is generally wider in RT tasks than TOJ tasks (Diederich and Colonius, 2015).

3.5.2. *Maximum Likelihood Estimation Model, Bayesian Frameworks, and Causal Inference Models*

Noise in sensory signals can lead to variance in perceptual estimates of an environmental property, such as the number of stimuli presented or spatial location of a stimulus. So the same stimulus presented multiple times may produce different estimates each time. Multimodal signals can reduce variance and alter accuracy in perceptual tasks. Maximum likelihood estimation (MLE) models of MSI predict that optimal integration is achieved by taking into account the reliability (the inverse of variance) of unisensory signals and weighting them accordingly. An integrated estimate is then produced based on the weighted average of the unisensory signals, and the reliability of the integrated estimate is the sum of the unisensory reliabilities (Ernst, 2006). For example, in a temporal numerosity judgment study, participants tended to underestimate the number of stimuli presented in haptic, auditory, and visual conditions, whereas multisensory conditions resulted in reduced variance at long ISI and less underestimation at short ISI (Philippi *et al.*, 2008). While multisensory data from long ISI conditions matched MLE predictions, variance increased at short ISI, which was attributed to a cognitive range effect not accounted for in the model. In a target localization task, variance was greater for streams of auditory stimuli than in visual trials (Parise *et al.*, 2012). Temporally uncorrelated bimodal trials produced less variance than unimodal trials, but correlated trials produced the least variance and most accurate performances. Visual trials matched the results estimated by the MLE for the most reliable modality and correlated trials matched the MLE for optimal AV integration (Parise *et al.*, 2012).

While the MLE may be suited to modelling perceptual estimates, it does not take into account prior knowledge. Bayesian frameworks incorporate prior knowledge (based on the strength of coupling between signals) and sensory evidence to create the most probable perception of an environmental event

(Ernst, 2006). Temporal correlations between auditory and visual streams of stimuli can affect the strength of coupling and modulate spatial perception. In a target localization task, when streams were fully correlated, participants perceived stimuli to be spatially closer than when streams were less or not correlated (Parise *et al.*, 2013). Cuppini *et al.* (2014) designed a model that could account for both spatial and temporal illusions using a Bayesian framework with priors (expectations on the probability of stimulus characteristics). They proposed a network between auditory and visual unisensory areas that consisted of two layers (one for each modality) with bidirectional excitatory connections and lateral connections that could be excitatory or inhibitory. Receptive field, variance, and time constants differed between modalities. The model was tested against previous data and described them fairly well in most cases. The model did not account for higher-level factors but they suggested that it could be developed or incorporated into existing models to describe more complex networks (Cuppini *et al.*, 2014).

McGovern *et al.* (2016) found that SJ training improved AV temporal discrimination (narrowing the TBW) and reduced AV interactions (the ventriloquist effect). In synchronous conditions, an increase in temporal precision would be expected to narrow the width of the ventriloquist effect tuning function but increase its magnitude. However, the ventriloquist effect was reduced at all SOA. A Bayesian causal inference model with priors suggested that training improves AV temporal precision and decreases prior expectations of a common source for component AV stimuli, thereby reducing both the width and magnitude of the ventriloquist effect-tuning function (McGovern *et al.*, 2016).

Two studies examined within-subject variability in the strength of multisensory binding and the effects of selective attention on MSI (Odegaard and Shams, 2016; Odegaard *et al.*, 2016). They applied Bayesian causal inference models to data from AV target localization and temporal numerosity judgment tasks that included trials with incongruent AV streams, akin to the SIFI. Results indicated that participants had stable binding tendencies across time with similar results from sessions at least a week apart. However, binding tendencies did not generalize across tasks, suggesting that spatial and temporal MSI processes were governed by separate perceptual biases that were stable over time, rather than a global parameter in the brain (Odegaard and Shams, 2016). In the second study, the spatial task results indicated that selective attention improved visual but not auditory precision, and did not affect integration (Odegaard *et al.*, 2016). The temporal task results indicated that selective attention improved precision in both modalities but did not affect integration in either.

Daemi *et al.* (2016) proposed a computational model underlying causal inference in spatial localization that included working memory as controlled
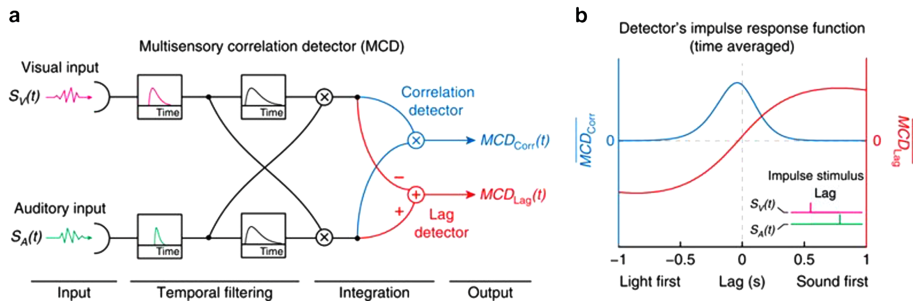
a

Multisensory correlation detector (MCD)

Visual input

$S_V(t)$

Time   Time

Correlation detector

$MCD_{Corr}(t)$

$MCD_{Lag}(t)$

Lag detector

Auditory input

$S_A(t)$

Time   Time

Input   Temporal filtering   Integration   Output

b

Detector's impulse response function (time averaged)

$\overline{MCD_{Corr}}$ $\overline{MCD_{Lag}}$

Impulse stimulus
Lag
$S_V(t)$
$S_A(t)$

−1   −0.5   0   0.5   1
Light first   Lag (s)   Sound first

Figure 7. Multisensory correlation detector (MCD) model. (a) Schematic representation of the model. The MCD integrates multisensory signals $[S_V(t), S_A(t)]$ through a set of low-pass temporal filters followed by linear operations. The MCD model yields two outputs, $MCD_{Corr}(t)$ and $MCD_{Lag}(t)$, representing, respectively, the temporal correlation and lag across the input signals. (b) Time-averaged impulse response function of the MCD. The *y*-axis represents the response of the model to visual and auditory impulses as a function of the lag across the senses (see inset). Blue line and axis represent the time-averaged response of the correlation detector (MCD_{Corr}), red line and axis represent the time-averaged response of the lag detector (MCD_{Lag}). Note how the correlation detector output (blue) peaks at low lags, whereas the output of the lag detector (red) changes sign depending on which modality comes first. Copyright Parise, C. V. and Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nat. Commun.*, **7**, 11543. https://doi.org/10.1038/ncomms11543. (Licensed under CC BY 4.0.)

leaky integrators. Spatial and temporal information from different sensory signals could be held and compared (spatiotemporal similarity measure), and the results fed into a decision-making framework to determine the extent of binding and perceived location (Daemi *et al*., 2016).

Parise and Ernst (2016) adapted the Hassenstein–Reichardt detector, originally a model of visual motion detection (Hassenstein and Reichardt, 1956), to develop the multisensory correlation detector (MCD) model of MSI. Signals from each modality go through a series of temporal filters and are then combined linearly to produce outputs for temporal correlation and lag between the inputs (Fig. 7). The MCD model was able to replicate causality judgment, TOJ, SJ, and AV correspondence detection task data. Results suggested that the model parsimoniously captured various aspects of MSI in line with other (unisensory) integration mechanisms (Parise and Ernst, 2016).

The computational model literature captured in the search illustrates how the field is constantly evolving to better explain multisensory perceptual phenomena and attempt to resolve issues such as the correspondence problem. Recent models suggest that low-level stimulus properties, including temporal relationships, help to determine signal reliability and interact with prior knowledge to influence perception. The integration of different MSI models, as well as the adaptation of models from other research fields is likely to continue to inform the study of multisensory processing, and vice versa. Computational

models may have practical uses in the development of artificial intelligence systems that require interpretation of sensory information, such as driverless vehicles.

4. Conclusions

In this scoping review, recent research into crossmodal interactions and MSI in early AV temporal processing was reviewed. Data collected from multisensory illusions and tasks, along with increasingly complex neuroimaging and brain stimulation techniques, as well as refined computational models have all contributed to our current understanding of multisensory processing.

The TBW and subjective simultaneity have been shown to change depending on both low-level and high-level factors, and have greater tolerance for asynchrony for complex stimuli such as speech (Krueger Fister *et al.*, 2016; Leone and McCourt, 2015; Stevenson and Wallace, 2013). In many of the studies captured in our search, TBW and subjective simultaneity measures indicated a higher propensity to bind vision-leading than auditory-leading AV stimuli (Stevenson *et al.*, 2012a, b; Yarrow *et al.*, 2011). Furthermore, AV training effects generalized to visual but not auditory TOJ (Alais and Cass, 2010), and adaptation effects on AV SJ and TOJ tasks were stronger when AV lags were vision leading (Navarra *et al.*, 2013; Van der Burg *et al.*, 2013). As noted above, this apparent asymmetry between processing of vision-led versus auditory-led stimuli may result from real-life experience, where light information from events consistently reaches the observer before sound waves. As vision generally provides less reliable temporal information than sound, it also stands to reason that it should be more flexible in temporal perception.

Understanding the factors underlying the larger TBW found in various clinical populations, and in healthy aging may provide insights into the workings of the brain in other processes. The TWIN model indicated that peripheral processing took longer for the elderly than the young (Diederich *et al.*, 2008). So the wider TBW common in elderly populations may be an adaptive response to compensate for slowed peripheral processing.

In support of the assertion made by Shimojo and Shams (2001), recent research captured in our search indicated that MSI can occur at early stages of processing, and multisensory perception is the result of networks spanning multiple regions in the brain, including the so-called unisensory areas. Studies of multisensory illusions that used arbitrary pairings of simple auditory and visual stimuli provided perceptual evidence suggesting early crossmodal interactions (e.g., Recanzone, 2003; Shams *et al.*, 2002). Both lag adaptation and rapid recalibration appear to be early sensory effects (Fujisaki *et al.*, 2004; Van der Burg *et al.*, 2013), and can affect AV SJ and TOJ performance (Fujisaki *et al.*, 2004; Navarra *et al.*, 2013; Vroomen *et al.*, 2004). Neuroimaging

studies also provided a range of evidence for early multisensory processing. Dipoles located in primary visual and auditory cortices showed modulations of ERF peak latencies with multimodal versus unimodal stimuli, and synchrony of component stimuli modulated this further (Franciotti *et al.*, 2011). Perceptions of synchrony and asynchrony involved networks that included sensory cortices (Dhamala *et al.*, 2007). Temporal ventriloquism decreased the flash-lag effect and the amplitude of the visual N1 (Stekelenburg and Vroomen, 2005), suggesting an early sensory process. Training-induced narrowing of the TBW affected activity and connections in sensory cortices and association areas (Powers *et al.*, 2012). In terms of the present state of multisensory research, the question is not *if* MSI involves early processing, but rather *how* different brain regions and neural processes interact, and synthesize information to solve the correspondence problem and produce coherent perception.

A promising line of research for how the brain may propagate and combine information is the study of oscillatory brain activity (Senkowski *et al.*, 2008). Perception of the SIFI (Keil *et al.*, 2014), AV asynchrony (Senkowski *et al.*, 2007), and recalibration (Kösem *et al.*, 2014) all affected oscillatory patterns. This is an area where multisensory research may provide insight into more general associative processes in the brain. Within the studies captured, different networks and/or stages of processing were suggested between the SJ and TOJ tasks (Binder, 2015), and between the SIFI fission and fusion illusions (Bolognini *et al.*, 2016; McGovern *et al.*, 2014). Interplay between these processes are likely to play a role in solving the correspondence problem, as temporal relationships appear to be an important factor, and emerging evidence suggests that it is not just synchrony, but also the correlation of temporal structure between signals that influences binding (Denison *et al.*, 2013; Parise *et al.*, 2012, 2013).

Some of the studies captured in the search supported the assertion put forth by Spence (2013) that the spatial principle of MSI does not always apply to non-spatial tasks (Fiacconi *et al.*, 2013; Navarra *et al.*, 2013). On the other hand, temporal factors modulated performance in a number of spatial localization tasks (Ho *et al.*, 2013; McGovern *et al.*, 2016; Parise *et al.*, 2012, 2013; Schmiedchen *et al.*, 2012; Van der Burg *et al.*, 2008; Van Wanrooij *et al.*, 2009). The apparent importance of temporal processing in MSI certainly warrants further research. While some models argue for separate mechanisms for temporal and spatial MSI (Odegaard and Shams, 2016), others argue for more general mechanisms (Parise and Ernst, 2016; Van Atteveldt *et al.*, 2014). Resolving this will be another interesting avenue for future research.

The human experience of the physical world is inherently multisensory. Multisensory research is uniquely poised as a relatively new but fast-growing field that has roots in the well-established unisensory research fields.

Understanding how sensory processing mechanisms, regions, and networks interact will not only be useful in developing interventions for sensory conditions but may also provide insight into more general brain function.

Supplementary Material

Supplementary material is available online at:
https://brill.figshare.com/s/5745e0b854650dd18ee0

References

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **14**, 257–262.

Alais, D. and Cass, J. (2010). Multisensory perceptual learning of temporal order: audiovisual learning transfers to vision but not audition, *PLoS One* **5**, e11283. DOI:10.1371/journal.pone.0011283.

Alais, D., Newell, F. and Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour, *Seeing Perceiving* **23**, 3–38.

Andersen, T. S., Tiippana, K. and Sams, M. (2005). Maximum likelihood integration of rapid flashes and beeps, *Neurosci. Lett.* **380**, 155–160.

Arksey, H. and O'Malley, L. (2005). Scoping studies: towards a methodological framework, *Int. J. Soc. Res. Methodol.* **8**, 19–32.

Bald, L., Berrien, F. K., Price, J. B. and Sprague, R. O. (1942). Errors in perceiving the temporal order of auditory and visual stimuli, *J. Appl. Psychol.* **26**, 382–388.

Barrett, D. J. K. and Krumbholz, K. (2012). Evidence for multisensory integration in the elicitation of prior entry by bimodal cues, *Exp. Brain Res.* **222**, 11–20.

Bausenhart, K. M., De La Rosa, M. D. and Ulrich, R. (2014). Multimodal integration of time: visual and auditory contributions to perceived duration and sensitivity, *Exp. Psychol.* **61**, 310–322.

Bedard, G. and Barnett-Cowan, M. (2016). Impaired timing of audiovisual events in the elderly, *Exp. Brain Res.* **234**, 331–340.

Benevento, L. A., Fallon, J., Davis, B. J. and Rezak, M. (1977). Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey, *Exp. Neurol.* **57**, 849–872.

Binder, M. (2015). Neural correlates of audiovisual temporal processing — comparison of temporal order and simultaneity judgments, *Neuroscience* **300**, 432–447.

Boenke, L. T., Deliano, M. and Ohl, F. W. (2009). Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment, *Exp. Brain Res.* **198**, 233–244.

Bolognini, N., Senna, I., Maravita, A., Pascual-Leone, A. and Merabet, L. B. (2010). Auditory enhancement of visual phosphene perception: the effect of temporal and spatial factors and of stimulus intensity, *Neurosci. Lett.* **477**, 109–114.

Bolognini, N., Convento, S., Casati, C., Mancini, F., Brighina, F. and Vallar, G. (2016). Multisensory integration in hemianopia and unilateral spatial neglect: evidence from the sound induced flash illusion, *Neuropsychologia* **87**, 134–143.

Brang, D., Williams, L. E. and Ramachandran, V. S. (2012). Grapheme-color synesthetes show enhanced crossmodal processing between auditory and visual modalities, *Cortex* **48**, 630–637.

Calvert, G. A. and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain, *J. Physiol. Paris* **98**, 191–205.

Chan, J. S., Langer, A. and Kaiser, J. (2016). Temporal integration of multisensory stimuli in autism spectrum disorder: a predictive coding perspective, *J. Neural Transm.* **123**, 917–923.

Chen, Y.-C. and Spence, C. (2017). Assessing the role of the 'unity assumption' on multisensory integration: a review, *Front. Psychol.* **8**, 445. DOI:10.3389/fpsyg.2017.00445.

Colonius, H. and Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model, *J. Cogn. Neurosci.* **16**, 1000–1009.

Colonius, H. and Diederich, A. (2011). Computing an optimal time window of audiovisual integration in focused attention tasks: illustrated by studies on effect of age and prior knowledge, *Exp. Brain Res.* **212**, 327–337.

Cuppini, C., Magosso, E., Bolognini, N., Vallar, G. and Ursino, M. (2014). A neurocomputational analysis of the sound-induced flash illusion, *Neuroimage* **92**, 248–266.

Daemi, M., Harris, L. R. and Crawford, J. D. (2016). Causal inference for cross-modal action selection: a computational study in a decision making framework, *Front. Comput. Neurosci.* **10**, 62. DOI:10.3389/fncom.2016.00062.

d'Arsonval, M. A. (1896). Dispositifs pour la mesure des courants alternatifs de toutes fréquences, *C. R. Soc. Biol.* **3**, 450–451.

de Boer-Schellekens, L. and Vroomen, J. (2014). Multisensory integration compensates loss of sensitivity of visual temporal order in the elderly, *Exp. Brain Res.* **232**, 253–262.

de Boer-Schellekens, L., Eussen, M. and Vroomen, J. (2013). Diminished sensitivity of audio-visual temporal order in autism spectrum disorder, *Front. Integr. Neurosci.* **7**, 8. DOI:10.3389/fnint.2013.00008.

de Boer-Schellekens, L., Stekelenburg, J. J., Maes, J. P., Van Gool, A. R. and Vroomen, J. (2014). Sound improves diminished visual temporal sensitivity in schizophrenia, *Acta Psychol. (Amst.)* **147**, 136–142.

Denison, R. N., Driver, J. and Ruff, C. C. (2013). Temporal structure and complexity affect audio-visual correspondence detection, *Front. Psychol.* **3**, 619. DOI:10.3389/fpsyg.2012.00619.

Dhamala, M., Assisi, C. G., Jirsa, V. K., Steinberg, F. L. and Scott Kelso, J. A. (2007). Multisensory integration for timing engages different brain networks, *Neuroimage* **34**, 764–773.

Diederich, A. and Colonius, H. (2009). Crossmodal interaction in speeded responses: time window of integration model, *Prog. Brain Res.* **174**, 119–135.

Diederich, A. and Colonius, H. (2015). The time window of multisensory integration: relating reaction times and judgments of temporal order, *Psychol. Rev.* **122**, 232–241.

Diederich, A., Colonius, H. and Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model, *Neuropsychologia* **46**, 2556–2562.

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration, in: *Human Body Perception From the Inside out*, G. Knoblich, I. M. Thornton, M. Grosjean and M. Shiffrar (Eds), pp. 105–131. Oxford University Press, New York, NY, USA.

Evans, E. F. and Whitfield, I. C. (1964). Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat, *J. Physiol.* **171**, 476–493.

Falchier, A., Clavagnier, S., Barone, P. and Kennedy, H. (2002). Anatomical evidence of multi-modal integration in primate striate cortex, *J. Neurosci.* **22**, 5749–5759.

Felleman, D. and Van Essen, D. (1991). Distributed hierarchical processing in primate cerebral cortex, *Cereb. Cortex* **1**, 1–47.

Fendrich, R. and Corballis, P. M. (2001). The temporal cross-capture of audition and vision, *Percept. Psychophys.* **63**, 719–725.

Fiacconi, C. M., Harvey, E. C., Sekuler, A. B. and Bennett, P. J. (2013). The influence of aging on audiovisual temporal order judgments, *Exp. Aging Res.* **39**, 179–193.

Franciotti, R., Brancucci, A., Della Penna, S., Onofrj, M. and Tommasi, L. (2011). Neuro-magnetic responses reveal the cortical timing of audiovisual synchrony, *Neuroscience* **193**, 182–192.

Frassinetti, F., Bolognini, N. and Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction, *Exp. Brain Res.* **147**, 332–343.

Fujisaki, W., Shimojo, S., Kashino, M. and Nishida, S. (2004). Recalibration of audiovisual simultaneity, *Nat. Neurosci.* **7**, 773–778.

Fujisaki, W., Kitazawa, S. and Nishida, S. (2012). Multisensory timing, in: *The New Handbook of Multisensory Processing*, B. E. Stein (Ed.), pp. 301–317. MIT Press, Cambridge, MA, USA.

Gebhard, J. W. and Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter, *Am. J. Psychol.* **72**, 521–529.

Hamilton, R. H., Wiener, M., Drebing, D. E. and Coslett, H. B. (2013). Gone in a flash: manipulation of audiovisual temporal integration using transcranial magnetic stimulation, *Front. Psychol.* **4**, 571. DOI:10.3389/fpsyg.2013.00571.

Harrar, V., Harris, L. R. and Spence, C. (2017). Multisensory integration is independent of perceived simultaneity, *Exp. Brain Res.* **235**, 763–775.

Hassenstein, V. and Reichardt, W. (1956). System theoretical analysis of time, sequence and sign analysis of the motion perception of the snout-beetle *Chlorophanus*, *Z Naturforsch B* **11**, 513–524.

Heron, J., Roach, N. W., Hanson, J. V. M., McGraw, P. V. and Whitaker, D. (2012). Audiovisual time perception is spatially specific, *Exp. Brain Res.* **218**, 477–485.

Heron, J., Hotchkiss, J., Aaen-Stockdale, C., Roach, N. W. and Whitaker, D. (2013). A neural hierarchy for illusions of time: duration adaptation precedes multisensory integration, *J. Vis.* **13**, 4. DOI:10.1167/13.14.4.

Ho, C., Gray, R. and Spence, C. (2013). Role of audiovisual synchrony in driving head orienting responses, *Exp. Brain Res.* **227**, 467–476.

Hodges, D. A., Hairston, W. D. and Burdette, J. H. (2005). Aspects of multisensory perception: the integration of visual and auditory information in musical experiences, *Ann. N. Y. Acad. Sci.* **1060**, 175–185.

Holmes, N. P. and Spence, C. (2005). Multisensory integration: space, time and superadditivity, *Curr. Biol.* **15**, R762–R764.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol.* **160**, 106–154.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex, *J. Physiol.* **195**, 215–243.

Jones, E. G. and Powell, T. P. S. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey, *Brain* **93**, 793–820.

Keetels, M. and Vroomen, J. (2011). No effect of synesthetic congruency on temporal ventriloquism, *Atten. Percept. Psychophys.* **73**, 209–218.

Keetels, M. and Vroomen, J. (2012). Perception of synchrony between the senses, in: *The Neural Bases of Multisensory Processes*, M. M. Murray and M. T. Wallace (Eds), pp. 147–177. CRC Press/Taylor and Francis, Boca Raton, FL, USA.

Keil, J., Müller, N., Hartmann, T. and Weisz, N. (2014). Prestimulus beta power and phase synchrony influence the sound-induced flash illusion, *Cereb. Cortex* **24**, 1278–1288.

King, A. J. (2005). Multisensory integration: strategies for synchronization, *Curr. Biol.* **15**, R339–R341.

King, A. J. and 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.

Kohlrausch, A., van Eijk, R., Juola, J. F., Brandt, I. and van de Par, S. (2013). Apparent causality affects perceived simultaneity, *Atten. Percept. Psychophys.* **75**, 1366–1373.

Kösem, A., Gramfort, A. and van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations, *Neuroimage* **92**, 274–284.

Kostaki, M. and Vatakis, A. (2016). Crossmodal binding rivalry: a "race" for integration between unequal sensory inputs, *Vis. Res.* **127**, 165–176.

Krueger Fister, J., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P. and Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing, *Neuropsychologia* **88**, 92–100.

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L. and Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders, *Front. Integr. Neurosci.* **4**, 129. DOI:10.3389/fnint.2010.00129.

Leone, L. M. and McCourt, M. E. (2015). Dissociation of perception and action in audiovisual multisensory integration, *Eur. J. Neurosci.* **42**, 2915–2922.

Liu, B., Jin, Z., Wang, Z. and Gong, C. (2011). The influence of temporal asynchrony on multisensory integration in the processing of asynchronous audio-visual stimuli of real-world events: an event-related potential study, *Neuroscience* **176**, 254–264.

Los, S. A. and Van der Burg, E. (2013). Sound speeds vision through preparation, not integration, *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1612–1624.

Love, S. A., Petrini, K., Cheng, A. and Pollick, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments, *PLoS One* **8**, e54798. DOI:10.1371/journal.pone.0054798.

Lu, Y., Paraskevopoulos, E., Herholz, S. C., Kuchenbuch, A. and Pantev, C. (2014). Temporal processing of audiovisual stimuli is enhanced in musicians: evidence from magnetoencephalography (MEG), *PLoS One* **9**, e90686. DOI:10.1371/journal.pone.0090686.

Macaluso, E. and Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain, *Trends Neurosci.* **28**, 264–271.

Mahoney, J. R., Molholm, S., Butler, J. S., Sehatpour, P., Gomez-Ramirez, M., Ritter, W. and Foxe, J. J. (2015). Keeping in touch with the visual system: spatial alignment and multisensory integration of visual-somatosensory inputs, *Front. Psychol.* **6**, 1068. DOI:10.3389/fpsyg.2015.01068.

Marchant, J. L., Ruff, C. C. and Driver, J. (2012). Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities, *Hum. Brain Mapp.* **33**, 1212–1224.

Mays, N., Roberts, E. and Popay, J. (2001). Synthesising research evidence, in: *Studying the Organisation and Delivery of Health Services: Research Methods*, N. Fulop, P. Allen, A. Clarke and N. Black (Eds), pp. 188–220. Routledge, London, UK.

McGovern, D. P., Roudaia, E., Stapleton, J., McGinnity, T. M. and Newell, F. N. (2014). The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration, *Front. Aging Neurosci.* **6**, 250. DOI:10.3389/fnagi.2014.00250.

McGovern, D. P., Roudaia, E., Newell, F. N. and Roach, N. W. (2016). Perceptual learning shapes multisensory causal inference via two distinct mechanisms, *Sci. Rep.* **6**, 24673. DOI:10.1038/srep24673.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748.

Medaglia, J. D., Lynall, M.-E. and Bassett, D. S. (2015). Cognitive network neuroscience, *J. Cogn. Neurosci.* **27**, 1471–1491.

Mégevand, P., Molholm, S., Nayak, A. and Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands, *PLoS One* **8**, e71608. DOI:10.1371/journal.pone.0071608.

Meredith, M. A. and Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus, *Brain Res.* **365**, 350–354.

Meredith, M. A. and Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration, *J. Neurophysiol.* **56**, 640–662.

Meredith, M. A. and Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons, *J. Neurophysiol.* **75**, 1843–1857.

Meredith, M. A., Nemitz, J. W. and Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors, *J. Neurosci.* **7**, 3215–3229.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals, *Cogn. Psychol.* **14**, 247–279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention, *Percept. Psychophys.* **40**, 331–343.

Morein-Zamir, S., Soto-Faraco, S. and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism, *Cogn. Brain Res.* **17**, 154–163.

Musacchia, G. and Schroeder, C. E. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex, *Hear. Res.* **258**, 72–79.

Navarra, J., Fernández-Prieto, I. and Garcia-Morera, J. (2013). Realigning thunder and lightning: temporal adaptation to spatiotemporally distant events, *PLoS One* **8**, e84278. DOI:10.1371/journal.pone.0084278.

Neufeld, J., Sinke, C., Zedler, M., Emrich, H. M. and Szycik, G. R. (2012). Reduced audio-visual integration in synaesthetes indicated by the double-flash illusion, *Brain Res.* **1473**, 78–86.

Nidiffer, A. R., Stevenson, R. A., Krueger Fister, J., Barnett, Z. P. and Wallace, M. T. (2016). Interactions between space and effectiveness in human multisensory performance, *Neuropsychologia* **88**, 83–91.

Nijhawan, R. (2002). Neural delays, visual motion and the flash-lag effect, *Trends Cogn. Sci.* **6**, 387–393.

Noel, J.-P., De Niear, M., Van Der Burg, E. and Wallace, M. T. (2016). Audiovisual simultaneity judgment and rapid recalibration throughout the lifespan, *PLoS One* **11**, e161698. DOI:10.1371/journal.pone.0161698.

Noel, J.-P., De Niear, M. A., Stevenson, R., Alais, D. and Wallace, M. T. (2017). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders, *Autism Res.* **10**, 121–129.

Nozaradan, S., Peretz, I. and Mouraux, A. (2012). Steady-state evoked potentials as an index of multisensory temporal binding, *Neuroimage* **60**, 21–28.

Odegaard, B. and Shams, L. (2016). The brain's tendency to bind audiovisual signals is stable but not general, *Psychol. Sci.* **27**, 583–591.

Odegaard, B., Wozny, D. R. and Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration, *Neurosci. Lett.* **614**, 24–28.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C. and Pantev, C. (2014). Multisensory integration during short-term music reading training enhances both uni- and multisensory cortical processing, *J. Cogn. Neurosci.* **26**, 2224–2238.

Parise, C. and Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect, *Neurosci. Lett.* **442**, 257–261.

Parise, C. V. and Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration, *Nat. Commun.* **7**, 11543. DOI:10.1038/ncomms11543.

Parise, C. V., Spence, C. and Ernst, M. O. (2012). When correlation implies causation in multisensory integration, *Curr. Biol.* **22**, 46–49.

Parise, C. V., Harrar, V., Ernst, M. O. and Spence, C. (2013). Cross-correlation between auditory and visual signals promotes multisensory integration, *Multisens. Res.* **26**, 307–316.

Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A. and McEwen, S. A. (2014). A scoping review of scoping reviews: advancing the approach and enhancing the consistency, *Res. Synth. Methods* **5**, 371–385.

Philippi, T. G., van Erp, J. B. F. and Werkhoven, P. J. (2008). Multisensory temporal numerosity judgment, *Brain Res.* **1242**, 116–125.

Pick, H. L., Warren, D. H. and Hay, J. C. (1969). Sensory conflict in judgments of spatial direction, *Percept. Psychophys.* **6**, 203–205.

Powers, A. R., Hevey, M. A. and Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning, *J. Neurosci.* **32**, 6263–6274.

Raab, D. H. (1962). Statistical facilitation of simple reaction times, *Trans. N. Y. Acad. Sci.* **24**, 574–590.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception, *J. Neurophysiol.* **89**, 1078–1093.

Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time, *Hear. Res.* **258**, 89–99.

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

Roseboom, W. and Arnold, D. H. (2011). Twice upon a time: multiple concurrent temporal recalibrations of audiovisual speech, *Psychol. Sci.* **22**, 872–877.

Roseboom, W., Kawabe, T. and Nishida, S. (2013). Audio-visual temporal recalibration can be constrained by content cues regardless of spatial overlap, *Front. Psychol.* **4**, 189. DOI:10.3389/fpsyg.2013.00189.

Roudaia, E., Sekuler, A. B., Bennett, P. J. and Sekuler, R. (2013). Aging and audio-visual and multi-cue integration in motion, *Front. Psychol.* **4**, 267. DOI:10.3389/fpsyg.2013.00267.

Rutschmann, J. and Link, R. (1964). Perception of temporal order of stimuli differing in sense mode and simple reaction time, *Percept. Mot. Skills* **18**, 345–352.

Scarpina, F., Migliorati, D., Marzullo, P., Mauro, A., Scacchi, M. and Costantini, M. (2016). Altered multisensory temporal integration in obesity, *Sci. Rep.* **6**, 28382. DOI:10.1038/srep28382.

Schall, S., Quigley, C., Onat, S. and König, P. (2009). Visual stimulus locking of EEG is modulated by temporal congruency of auditory stimuli, *Exp. Brain Res.* **198**, 137–151.

Scheier, C. R., Nijhawan, R. and Shimojo, S. (1999). Sound alters visual temporal resolution, *Invest. Ophthalmol. Vis. Sci.* **40**, 4169.

Schmiedchen, K., Freigang, C., Nitsche, I. and Rübsamen, R. (2012). Crossmodal interactions and multisensory integration in the perception of audio-visual motion — a free-field study, *Brain Res.* **1466**, 99–111.

Sekuler, R., Sekuler, A. B. and Lau, R. (1997). Sound alters visual motion perception, *Nature* **385**, 308.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S. and Woldorff, M. G. (2007). Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations, *Neuropsychologia* **45**, 561–571.

Senkowski, D., Schneider, T. R., Foxe, J. J. and Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing, *Trends Neurosci.* **31**, 401–409.

Setti, A. and Chan, J. S. (2011). Familiarity of objects affects susceptibility to the sound-induced flash illusion, *Neurosci. Lett.* **492**, 19–22.

Setti, A., Burke, K. E., Kenny, R. A. and Newell, F. N. (2011). Is inefficient multisensory processing associated with falls in older people?, *Exp. Brain Res.* **209**, 375–384.

Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A. and Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion, *Neuropsychologia* **61**, 259–268.

Shams, L., Kamitani, Y. and Shimojo, S. (2000). What you see is what you hear, *Nature* **408**, 788.

Shams, L., Kamitani, Y. and Shimojo, S. (2002). Visual illusion induced by sound, *Cogn. Brain Res.* **14**, 147–152.

Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions, *Curr. Opin. Neurobiol.* **11**, 505–509.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule, *Ann. N. Y. Acad. Sci.* **1296**, 31–49.

Spence, C. and Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony, *Curr. Biol.* **13**, R519–R521.

Spence, C., Senkowski, D. and Röder, B. (2009). Crossmodal processing, *Exp. Brain Res.* **198**, 107–111.

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience, *Nat. Neurosci.* **17**, 652–660.

Steenken, R., Weber, L., Colonius, H. and Diederich, A. (2014). Designing driver assistance systems with crossmodal signals: multisensory integration rules for saccadic reaction times apply, *PLoS One* **9**, e92666. DOI:10.1371/journal.pone.0092666.

Stekelenburg, J. J. and Vroomen, J. (2005). An event-related potential investigation of the time-course of temporal ventriloquism, *Neuroreport* **16**, 641–644.

Stephen, J. M., Knoefel, J. E., Adair, J., Hart, B. and Aine, C. J. (2010). Aging-related changes in auditory and visual integration measured with MEG, *Neurosci. Lett.* **484**, 76–80.

Stevenson, R. A. and Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies, *Exp. Brain Res.* **227**, 249–261.

Stevenson, R. A., Krueger Fister, J., Barnett, Z. P., Nidiffer, A. R. and Wallace, M. T. (2012a). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance, *Exp. Brain Res.* **219**, 121–137.

Stevenson, R. A., Zemtsov, R. K. and Wallace, M. T. (2012b). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions, *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 1517–1529.

Stevenson, R. A., Wilson, M. M., Powers, A. R. and Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing, *Exp. Brain Res.* **225**, 479–489.

Thompson, S. P. (1910). A physiological effect of an alternating magnetic field, *Proc. R. Soc. London B Biol. Sci.* **82**, 396–398.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention, *Cogn. Psychol.* **12**, 97–136.

Van Atteveldt, N., Murray, M. M., Thut, G. and Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations, *Neuron* **81**, 1240–1253.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W. and Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search, *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 1053–1065.

Van der Burg, E., Alais, D. and Cass, J. (2013). Rapid recalibration to audiovisual asynchrony, *J. Neurosci.* **33**, 14633–14637.

Van der Burg, E., Alais, D. and Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales, *Sci. Rep.* **5**, 14526.

Van der Burg, E., Alais, D. and Cass, J. (2018). Rapid recalibration to audiovisual asynchrony follows the physical — not the perceived — temporal order, *Atten. Percept. Psychophys.* **80**, 2060–2068.

Van der Stoep, N., Spence, C., Nijboer, T. C. W. and Van der Stigchel, S. (2015). On the relative contributions of multisensory integration and crossmodal exogenous spatial attention to multisensory response enhancement, *Acta Psychol. (Amst.)* **162**, 20–28.

van Eijk, R. L. J., Kohlrausch, A., Juola, J. F. and van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type, *Percept. Psychophys.* **70**, 955–968.

Van Wanrooij, M. M., Bell, A. H., Munoz, D. P. and Van Opstal, A. J. (2009). The effect of spatial–temporal audiovisual disparities on saccades in a complex scene, *Exp. Brain Res.* **198**, 425–437.

Vidal, M. (2017). Hearing flashes and seeing beeps: timing audiovisual events, *PLoS One* **12**, e172028. DOI:10.1371/journal.pone.0172028.

Vroomen, J. and De Gelder, B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect, *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 513–518.

Vroomen, J. and Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review, *Atten. Percept. Psychophys.* **72**, 871–884.

Vroomen, J., Keetels, M., de Gelder, B. and Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony, *Cogn. Brain Res.* **22**, 32–35.

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

Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention, *Can. J. Exp. Psychol.* **48**, 242–259.

Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy, *Psychol. Bull.* **88**, 638–667.

Welch, R. B., DuttonHurt, L. D. and Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception, *Percept. Psychophys.* **39**, 294–300.

Yarrow, K., Roseboom, W. and Arnold, D. H. (2011). Spatial grouping resolves ambiguity to drive temporal recalibration, *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 1657–1661.

Zampini, M., Guest, S., Shore, D. I. and Spence, C. (2005a). Audio-visual simultaneity judgments, *Percept. Psychophys.* **67**, 531–544.

Zampini, M., Shore, D. I. and Spence, C. (2005b). Audiovisual prior entry, *Neurosci. Lett.* **381**, 217–222.

Zmigrod, S. and Zmigrod, L. (2015). Zapping the gap: reducing the multisensory temporal binding window by means of transcranial direct current stimulation (tDCS), *Consc. Cogn.* **35**, 143–149.