Intersensory binding across space and time: A tutorial review

Lihan Chen · Jean Vroomen

Published online: 25 May 2013
© Psychonomic Society, Inc. 2013

Abstract Spatial ventriloquism refers to the phenomenon that a visual stimulus such as a flash can attract the perceived location of a spatially discordant but temporally synchronous sound. An analogous example of mutual attraction between audition and vision has been found in the temporal domain, where temporal aspects of a visual event, such as its onset, frequency, or duration, can be biased by a slightly asynchronous sound. In this review, we examine various manifestations of spatial and temporal attraction between the senses (both direct effects and aftereffects), and we discuss important constraints on the occurrence of these effects. Factors that potentially modulate ventriloquism—such as attention, synesthetic correspondence, and other cognitive factors—are described. We trace theories and models of spatial and temporal ventriloquism, from the traditional *unity assumption* and *modality appropriateness hypothesis* to more recent Bayesian and neural network approaches. Finally, we summarize recent evidence probing the underlying neural mechanisms of spatial and temporal ventriloquism.

Keywords Multisensory Processing · Temporal processing · Spatial localization

Introduction

At the opening ceremony of the 2008 Beijing Summer Olympic Games, the successful debut performance of a lovely little girl had attracted widespread attention. Accompanying her on-stage performance, indeed, an off-stage voice got "attached" to the lip-syncing girl. Her vivid vocal performance had been finally revealed as a successful implementation of spatial ventriloquism (http://www.youtube.com/watch?v=8BlEiCqQRxQ&feature=related). Obviously, there are many more mundane examples of spatial ventriloquism, such as simply watching a TV. Here, the audio is perceived where the action is seen, rather than at the actual location of the sound (i.e., the position of the loudspeaker). A similar illusion occurs in the temporal domain (i.e., temporal ventriloquism), because the sound and the video of the TV appear to be synchronous despite delays between the two signals. The spatial and temporal ventriloquist effects have also received considerable attention in the scientific literature, because they demonstrate a more general phenomenon—namely, that sensory modalities such as vision, audition, and touch interact and sometimes change the percept of each other (Calvert, Spence & Stein, 2004; Stein, 2012; Stein & Meredith, 1993). The resulting cross-modal illusions have turned out to be extremely useful tools for probing how the brain combines information from different modalities. In essence, it appears to be the case that when information from two different modalities are in slight conflict with each other, cross-modal combinational fusions arise that produce multisensory illusions that can be every bit as compelling as those within a given sense (Stein, 2012).

Here, we review the literature on these intersensory illu-

expanding literature on these topics, we had to be very selective. Table 1 provides a selection of studies that we considered a representative example of the paradigms currently used, while Table 2 summarizes the basic findings of this literature. We apologize for all those that have not been mentioned.

Spatial ventriloquism: Immediate effect

Probably the best known example of intersensory binding is the visual bias of auditory location, here referred to as *spatial ventriloquism*. In a typical demonstration of spatial ventriloquism, the performing artist would synchronize the movements of a puppet's mouth with his own speech while avoiding movements of his/her own head or lips. The source of the sound is then mislocalized toward the position of the puppet's mouth. Fortunately, experimental psychologists do not need to be as artistic as that, because they can use a stripped-down version of this setup that quite often consists of a single beep from one location delivered with a synchronized flash from another location. The task of the observer might be to point or make a saccade toward the (apparent) location of the sound or to decide whether the sound came from the left or the right of a reference point, while at the same time trying to ignore the visual distractor (see Fig. 1). Alternatively, observers may also be asked to judge whether the flash and beep originated from a single location (whether they "fused") or not, in which case the visual stimulus cannot be ignored but is task relevant.

The spatial ventriloquist illusion manifests itself when the immediate pointing response toward the sound is shifted toward the visual stimulus despite instructions to ignore the latter (Alais & Burr, 2004b; Bertelson, 1999; Bertelson & Radeau, 1981; Brancazio & Miller, 2005; Howard & Templeton, 1966; Munhall, Gribble, Sacco & Ward, 1996; Radeau & Bertelson, 1987) or when, in the case of a fusion response, despite spatial separation, synchronized audiovisual stimuli fuse and are perceived as coming from a single location (Bertelson & Radeau, 1981; Godfroy, Roumes & Dauchy, 2003). This illusion has been demonstrated not only in human observers, but also in species such as cats, ferrets, and birds (Kacelnik, Walton, Parsons & King, 2002; King, Doubell, & Skaliora, 2004; Knudsen, Knudsen & Esterly, 1982; Knudsen & Knudsen, 1985, 1989; Meredith & Allman, 2009; Wallace & Stein, 2007).

Cross-modal mutual biases in localization responses have also been found in other modalities than the auditory and visual. In the visuomotor domain, there are famous prism adaptation studies that have been known since the late 19th century, when von Helmholtz published his seminal work in optics (von Helmholtz, 1962). During the mid-1960s, Held (1965) demonstrated that prism adaptation depends on the interaction between the motor and the visual systems and that such interaction normally induces a plastic change in the brain. There are also more recent studies reporting spatial attraction between the visual and somatosensory modalities (Blakemore, Bristow, Bird, Frith & Ward, 2005; Dionne, Meehan, Legon & Staines, 2010; Forster & Eimer, 2005; Rock & Victor, 1964; Serino, Farnè, Rinaldesi, Haggard & Làdavas, 2007 Taylor-Clarke, Kennett & Haggard, 2002) and between the auditory and tactile modalities (Bruns & Röder, 2010a, b; Caclin, Soto-Faraco, Kingstone & Spence, 2002; Occelli, Bruns, Zampini & Röder, 2012). Spatial ventriloquism can also be found with dynamic stimuli. In apparent motion, visual motion direction can attract the perceived direction of auditory motion (Kitajima & Yamashita, 1999; Mateeff, Hohnsbein & Noack, 1985; Soto-Faraco, Lyons, Gazzaniga, Spence & Kingstone, 2002; Soto-Faraco, Spence & Kingstone, 2004a, b, 2005; Stekelenburg & Vroomen, 2009), and auditory motion can attract visual motion (Alais & Burr, 2004a; Chen & Zhou, 2011; Meyer & Wuerger, 2001; Wuerger, Hofbauer & Meyer, 2003).

From a theoretical point of view, it is important to realize that in spatial ventriloquism, there is not a complete *capture* of sound by vision but, rather, a *mutual* attraction in space. The effect of vision on sound location is usually robust, whereas the reverse effect—sound attracting visual location—is usually quite subtle and has mostly been observed with visual displays that are difficult to localize (Alais & Burr, 2004b; Bertelson & Radeau, 1981, 1987). More recently, though, a particularly clear effect of sound on visual localization has been reported by Hidaka et al. (2009; see also Teramoto, Hidaka, Sugita, Sakamoto, Gyoba, Iwaya & Suzuki, 2012). These authors presented a blinking visual stimulus at a fixed location against a nontextured dark background. This static blinking stimulus is perceived to be moving laterally when the flash onsets are synchronized with an alternating left–right sound source. This illusory visual motion is particularly powerful when retinal eccentricity is increased, and it also works in the vertical dimension when sounds alternate in upper and lower space (for a demo, see www.journal.pone.0008188.s003.mov).

Temporal ventriloquism: Immediate effect

Less well-known is that an analogous phenomenon of intersensory binding occurs in the time dimension, referred to as *temporal ventriloquism*. Here, temporal aspects of a visual stimulus, such as its onset, interval, or duration, can be shifted by slightly asynchronous auditory stimuli (Alais & Burr, 2004a; Bertelson, 1999; Burr, Banks & Morrone, 2009; Chen, Shi & Müller, 2010; Fendrich & Corballis, 2001; Freeman & Driver, 2008; Getzmann, 2007; Morein-Zamir, Soto-Faraco & Kingstone, 2003; Recanzone, 2009; Scheier, Nijhawan & Shimojo, 1999; Sekuler, Sekuler & Lau, 1997; Vroomen & de Gelder, 2004b; Watanabe &

Table 1 Selection of representative studies on spatial and temporal ventriloquism

Effect	Task/Paradigm	Main Question	Finding	Sample Study
Spatial ventriloquist effect	Sound localization/ fusion	Whether bias occurred only with fusion	Vision attracts auditory location without fusion	Bertelson and Radeau (1981)
	Speech identification / sound localization	Role of "unity assumption"	Ventriloquism unaffected by face orientation	Bertelson et al. (1994)
	Sound localization	Role of endogenous/ exogenous attention	Visual attention no effect on ventriloquism	Bertelson et al. (2000b); Vroomen et al. (2001a, b)
	Sound localization	Spatial criteria	<15° separation necessary	Slutsky and Recanzone (2001)
	Sound localization	Temporal criteria	From −100 to +300 ms	Slutsky and Recanzone (2001)
	Sound localization	Cue combination of audiovisual signals	Near optimal integration	Alais and Burr (2004b)
	Sound localization	Common cause of sensory cues	Ideal observer model	Körding et al. (2007)
	Sound localization	Contribution of prior and likelihood	Prior modulates audiovisual integration	Van Wanrooij et al. (2010)
	Sound localization	Role of visuospatial attention	Arrows and gaze shift sound location	Borjon et al. (2011)
	Fusion response	Properties of audiovisual fusion	Audiovisual fusion areas larger in periphery	Godfroy et al. (2003)
	Direction of auditory motion	Capture with dynamic stimuli	Visual and tactile stimuli capture auditory motion	Soto-Faraco et al. (2002, 2004)
	Sound localization	Neural mechanism of ventriloquism	Affects auditory cortex	Bonath et al. (2007); Stekelenburg and Vroomen (2009)
	Sound localization	Neural mechanism of pitch–size synesthesia	Right parietal area involved	Bien et al. (2012)
	Direction of auditory motion	Neural mechanism of syneasthetic congruency	Low- and high-level mechanisms	Sadaghiani et al. (2009)
Spatial ventriloquist aftereffect	Sound localization	Generalization across frequency	No transfer between 750 and 3 kHz	Recanzone (1998); Lewald (2002)
	Sound localization	Generalization across frequency	Partial or complete transfer between 400 and 6.4 kHz	Frissen et al. (2003, 2005)
	Sound localization	Generalization across space	Greater at adapted location	Bertelson et al., (2006); Kopco et al. (2009)
	Sound localization	Effect of perceptual load	No influence or bigger effect with central load	Eramudugolla et al. (2011)
	Sound localization	Time course/dissipation	Fast/single exposure	Wozny and Shams (2011b); Frissen et al. (2012)
Temporal ventriloquist effect	Visual TOJ	Auditory capture of vision	Clicks improve visual TOJ	Scheier et al. (1999); Morein-Zamir et al. (2003)
	Visual TOJ	Role of spatial discordance	No effect of audiovisual spatial discordance	Vroomen and Keetels (2006)
	Visual TOJ	Role of auditory grouping	Auditory grouping precedes intersensory binding	Keetels et al. (2007)
	AV-TOJ	Role of cross-modal correspondence	Poor JNDs for audiovisual congruent speech	Vatakis et al. (2008)
	AV-TOJ	Role of cross-modal correspondence	Poor JNDs for congruent pitch–size	Parise and Spence (2009)
	AV-TOJ	Role of cross-modal correspondence	No effect of binding with audiovisual sine wave speech	Vroomen and Stekelenburg (2011)
	Visual motion judgment	Role of temporal grouping	Bias in apparent motion	Freeman and Driver (2008); Chen et al. (2011)
	Finger tapping	Auditory/visual dominance	Clicks influence synchronization with flashes	Aschersleben and Bertelson (2003); Repp (2005)
	TOJ/interval estimate	Bayesian approach	Less perfect quantitative fit	Burr et al. (2009); Ley et al. (2009);
Temporal ventriloquist aftereffect	AV TOJ/synchrony judgments	Temporal recalibration	PSS shifts in the direction of exposure lag	Fujisaki et al. (2004); Vroomen et al. (2004)
	Motor–sensory TOJ		Reversal of motor–sensory order	Stetson et al. (2006)

Table 1 (continued)

Effect	Task/Paradigm	Main Question	Finding	Sample Study
		Change in perception of causality		
	AV-TOJ	Storage/dissipation of aftereffect	Counterevidence (not delay) affects dissipation	Machulla, Di Luca, Froehlich and Ernst (2012)
	AV-TOJ	Role of spatial and contextual factors	Concurrent estimations and no location constraints	Roseboom and Arnold (2011)
	AV-TOJ	Attention modulation of temporal pattern	Spatially specific	Heron et al. (2012)
	Unimodal stimulus detection	Recalibration via change in processing speed	Auditory processing faster after sound-lag adaptation	Navarra et al. (2009)
	Finger tapping	Generalization of aftereffect	Synchronized finger tapping to flashes/clicks changed	Sugano et al. (2012)
	Magnitude estimation	Population coding in timing	Adaptation not uniform for each SOA	Roach et al. (2011)

Shimojo, 2001). One particularly clear manifestation of temporal ventriloquism is that an abrupt sound attracts the apparent onset of a slightly asynchronous flash (see Fig. 2). As in the spatial case, temporal ventriloquism can also be evoked by touch, and it can also become manifest in motor–sensory illusions (Bresciani & Ernst, 2007; Keetels & Vroomen, 2008a).

In general, researchers have interpreted temporal ventriloquism in terms of "*capture*" of auditory time onsets (or time intervals) over corresponding visual time onsets (or time intervals), rather than as a mutual bias between vision and audition, as in the case of spatial ventriloquism (e.g., Recanzone, 2009). An early demonstration of what one might, arguably, refer to as an example of temporal ventriloquism was reported by Gebhard and Mowbray (1959) in a phenomenon called *auditory driving*. They presented observers with a flickering light (5–40 Hz) and a fluttering sound (varying between ~5 and ~40 Hz). Observers reported that a constant flicker rate altered when the flutter changed, whereas the reverse effect (visual flicker altering auditory

Table 2 Summary of basic characteristics of audiovisual binding in space and time

	Space	Time
Relative strength	–Vision usually dominates audition, but mutual attraction can be demonstrated	–Audition captures vision
Temporal window	–Audiovisual stimuli need to be presented within ~−100 ms (sound-first) to ~+300 ms (sound-late)	–Somewhat narrower than for space
Spatial window	~±15° of horizontal separation, but with large variation	–Unconstraint by spatial disparity
Stimulus features	–Greater effect when sounds are difficult to localize	–Sounds with sharp transition
	–Visual stimuli can be presented in focus or periphery	–Visual stimuli preferably in periphery
		–Audiovisual rate <6 Hz
Aftereffect	–Space- and eye-specific (greater at adapted position)	–Modality-specific change in processing speed
	–Greater at adapted frequency, but with mixed evidence about transfer to other frequencies	–Smaller at adapted delay
	–Fast (after single exposure)	–Frequency specific
		–Space specific (simultaneous adaptation to sound-lead and sound-lag possible)
		–Probably fast (possibly after a few exposures)
Role of attention	–Direction of endogenous/exogenous shift of attention and shift in sound location can be dissociated	–Sounds preferably segregated with sharp onsets
	–But arrows and gaze can induce shift sound location as well	–Attention to the audiovisual timing relation increases aftereffect
	–Dual task with focused attention does not decrease the aftereffect	
Audiovisual congruence	–Phonetic congruency in speech: no effect	–Gender-matched speech: more fusion
	–Face orientation: no effect–	–Pitch/size congruence: more fusion for congruent pairs
	–Speech/nonspeech mode with sine wave speech: no effect	–Nonspeech like musical instruments: no effect of audiovisual congruency
	–Pitch/size congruence: greater effect for congruent pairs	

flutter) could not be observed. In more recent years, temporal ventriloquism has been demonstrated in a number of other paradigms: Besides auditory driving (Bresciani & Ernst, 2007; Gebhard & Mowbray, 1959; Recanzone, 2003; Shipley, 1964; Welch, DuttonHurt & Warren, 1986), or a variant of this called the *double-flash illusion* (Shams, Kamitani & Shimojo, 2000), researchers have used the flash-lag effect with accompanying sounds (Vroomen & de Gelder, 2004b), visual temporal order judgment (TOJ) tasks with accompanying sounds (Bertelson & Aschersleben, 2003; Morein-Zamir et al., 2003; Vroomen & Keetels, 2006), sensorimotor synchronization (Aschersleben & Bertelson, 2003; Repp, 2005; Repp & Penel, 2002; Stekelenburg, Sugano & Vroomen, 2011; Sugano, Keetels & Vroomen, 2010, 2012), and other variants of cross-modal temporal capture (Alais & Burr, 2004b; Bruns & Getzmann, 2008; Chen & Zhou, 2011; Freeman & Driver, 2008; Getzmann, 2007; Kafaligonul & Stoner, 2010; Shi, Chen & Müller, 2010; Vroomen & de Gelder, 2000, 2003).

A particularly useful setup that has provided a relative bias-free measure of temporal capture was first described by

sounuda

modality (visual or tactile) can be modulated by spatially uninformative but temporally irrelevant grouping stimuli in the distractor (auditory) modality (Chen, Shi & Müller, 2011; Freeman & Driver, 2008; Kafaligonul & Stoner, 2010). Cross-modal temporal capture in motion perception has also been demonstrated in a task of categorizations of visual motion percepts (Getzmann, 2007; Shi et al., 2010). Temporal capture has also been demonstrated in synchronization tasks in which observers are quite capable of tapping a finger in synchrony with a click while ignoring a temporally misaligned flash, but when trying to tap in synchrony with a flash, participants have great difficulty ignoring a temporally misaligned click (Aschersleben & Bertelson, 2003; Repp, 2005).

Spatial ventriloquist aftereffects

Another signature of a "true" merging of the senses is that prolonged exposure to an intersensory conflict leads to compensatory aftereffects. For spatial ventriloquism, it consists of postexposure shifts in auditory localization toward the visual distractor (Bertelson, Frissen, Vroomen, & de Gelder, 2006; Canon, 1970; Frissen, Vroomen, de Gelder & Bertelson, 2003, 2005; Lewald, 2002; Radeau, 1973, 1992; Radeau & Bertelson, 1969, 1974, 1976, 1977, 1978; Recanzone, 1998; Zwiers, Van Opstal & Paige, 2003) and sometimes also in visual localization (e.g., Radeau, 1973; Radeau & Bertelson, 1974, 1976, Experiment 1). A simple procedure for measuring aftereffects is depicted in Fig. 4, where, after exposure to an audiovisual spatial conflict, *unimodally* presented test sounds are displaced in the direction of the conflicting visual stimulus seen during the exposure phase.

It is generally agreed that these aftereffects reflect a recalibration process that is evoked to reduce the discrepancy between the senses. Most likely, this kind of recalibration is essential in achieving and maintaining a coherent intersensory representation of space, as in the case of prism adaptation (Held, 1965; Redding & Wallace, 1997; Welch, 1978). On a long-term scale, recalibration may compensate for growth of the body, head, and limbs, while on a short-term scale, it likely accommodates all kinds of changes in the acoustic environment that occur when, for example, one enters a new room.

Examining aftereffects has several interesting properties that are not available when testing immediate effects: One is that, during the posttest, observers do not need to ignore a visual distractor, because the test stimuli are presented unimodally. The advantage of this is that Stroop-like response conflicts between modalities, like an observer who points by mistake to a flash rather than a target sound, do not contaminate the picture. Another advantage is that one can

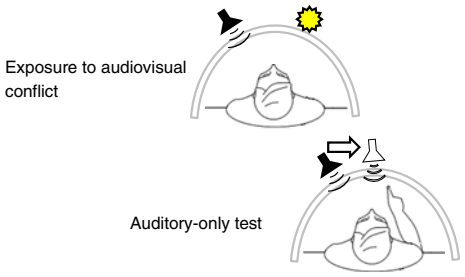

Fig. 4 Setup for measuring a spatial ventriloquist aftereffect: Observers are first exposed for a prolonged time to an audiovisual spatial conflict (here, a train of flashes to the right of sounds). In an auditory posttest, the apparent location of the sound is shifted in the direction of the previously experienced conflict

probe for the occurrence of aftereffects at different stimulus values than the one used during exposure. This can tell one whether the changes were specific to the values used in the exposure situation or, instead, generalize to a range of neighboring values.

The magnitude of the aftereffect typically depends on the number of exposure trials and the spatial discrepancy experienced during exposure (usually between 5° and 15°). Usually, it is a fraction of that discordance, although it can vary considerably by about 10 %–50 % in humans (Bertelson et al., 2006; Frissen, Vroomen & de Gelder, 2012; Kopco, Lin, Shinn-Cunningham & Groh, 2009) and 25 % in monkeys (but see Recanzone, 1998, who obtained the same amount of aftereffect as the adapting displacement). When observers are adapted in a single location in space, visual recalibration of apparent sound location does not shift uniformly to the left or right, but the effect is bigger at the trained than at the untrained location (Bertelson et al., 2006). This location-specific aftereffect partly shifts with eye gaze (Kopco et al., 2009).

The transfer of the aftereffect has also been examined in the auditory frequency domain to investigate whether adaptation is specific for sound localization cues on the basis of interaural time differences (mainly used for low-frequency tones) and interaural level differences (mainly used for high-frequency tones). The critical examination is to use the same or different auditory frequencies in the exposure and test phases. The picture here is not entirely clear: Recanzone (1998) and Lewald (2002) reported that aftereffects did not transfer across frequencies of 750 and 3000 Hz (Bruns & Röder, 2012; Lewald, 2002; Recanzone, 1998), while Frissen and collaborators obtained transfer across even wider frequency differences of 400 and 6400 Hz (Frissen et al., 2003, 2005)

Spatial ventriloquism and its aftereffect are also effective in improving spatial hearing in monaural conditions when interaural difference cues are not available. For example, Strelnikov, Rosito and Barone (2011) had observers wear an ear plug for 5 days, during which time they were trained in five 1-h sessions to localize monaural sounds. Sound

Another prediction of this sensory-specific criterion-shift account is that the adjusted modality causes a shift of equal magnitude in other cross-modal combinations. For example, Sugano et al. (2010) compared lag adaption in motor–visual and motor–auditory pairings (i.e., a finger tap followed by a delayed click or a delayed flash) and reported that the PSS was uniformly shifted within and across modalities. Adaptation to a delayed tap–click thus shifted not only the perceived timing of a tap–click test stimulus, but also the perceived timing of a tap–flash test stimulus (and vice versa). They argued that this pattern was most easily accounted for by assuming that the timing of the motor component (when was the tap?) shifted.

Yet another prediction of the sensory-specific criterion-shift account is that adaptation to asynchrony produces uniform recalibration across a whole range of SOAs. Observers who shift their PSS by 30 ms to sound-leading thus should do this across a whole range of SOAs. Interestingly, contrary to this prediction, it has been reported that the magnitude of the induced shift is not equal for each SOA but actually *increases* as the SOA of the test stimulus moves away from the adapted delay[1] (Roach, Heron, Whitaker & McGraw, 2011). Roach et al. examined this by adapting observers to an audiovisual delay of either 100-ms sound-first or 100-ms light-first trials and then measured the perceived magnitude of temporal separation at a wider range of SOAs. The authors found that the magnitude of the induced bias was practically zero at the adapted delays themselves (if compared with no delay—i.e., 0-ms adaptation baseline) but increased as the SOA of the test stimulus moved away from that of the adaptor. To explain these findings, Roach et al. proposed that multisensory timing is represented by a dedicated population code of neurons that are each specifically tuned to different asynchronies. Intersensory timing is represented by the distributed activity across these neurons. When observers adapt to a specific delay—say, audiovisual pairs of 100 ms sound-first—it results in a reduction of the response gain of the neurons around the adapted delay of 100-ms sound-first. A simultaneous sound–light pair (at 0-ms lag) then causes a repulsive shift of the population response profile away from the adapted SOA, and a simultaneous pair is then perceived as 'light-first' (see also Cai, Stetson & Eagleman, 2012, for a similar model).

Another line of research has examined whether temporal recalibration is stimulus specific or, rather, generalizes across different stimulus values. Navarra, García-Morera and Spence (2012) reported that audiovisual temporal adaptation only partly generalizes across auditory frequencies, since exposure to lagging sounds of 250 Hz caused shifts of the PSS in an SJ task if the test sounds were of the same frequency (250 Hz) or slightly different (300 Hz) but the effect was smaller (although still significant) if the test sound was 2500 Hz. In a further quest for stimulus specificity, Roseboom and Arnold (2011) adapted observers to a male actor on the left of the center of a screen whose lip movements *lagged* the sound track, whereas a female actor was shown on the right of the screen whose lip movements *preceded* the soundtrack. Results showed that audiovisual synchrony estimates for each actor were shifted toward the preceding audiovisual timing relationship for that actor. Temporal recalibrations thus occurred in positive and negative directions concurrently. This refutes the idea of a generalized timing mechanism but, rather, supports the idea that perceivers can form multiple concurrent estimates of appropriate timing for audiovisual synchrony. In a similar vein, Heron, Roach, Hanson, McGraw and Whitaker (2012) showed that observers were able to simultaneously adapt to two opposing temporal relationships, provided stimuli were segregated in space. Perceivers thus could concurrently be adapted to "sound-first on the left" and "flash-first on the right." Interestingly, no stimulus-specific recalibration was found when the spatial segregation was replaced by contextual stimulus features like "high-pitched sound-first" and "low-pitched sound late." This may suggest that audiovisual timing is spatially selective or, alternatively, that adapters need to be sufficiently different from each other so that separate timing relations can be maintained.

Spatial and temporal criteria for intersensory pairing

The underlying notion for both spatial and temporal ventriloquism is that the brain integrates, despite small deviances in space and time, signals from different modalities into a single multisensory event. Critically, when the deviance between the signals in space or time is too large, the signals likely originate from different events, in which case there is no reason to "bind" the information streams. Consequently, there is then also no reason to fuse, integrate, or recalibrate, because two separate events are perceived. This notion raises the question of whether spatial and temporal ventriloquism actually depends on the same a priori criteria for intersensory binding. Surprisingly, from the literature, it appears that this is most likely *not* the case. There is a well-established finding that a variety of multisensory illusions are preserved over a time window of several hundred milliseconds surrounding simultaneity, giving rise to the notion of a "temporal window of integration" (Colonius & Diederich, 2004; Dixon & Spitz, 1980; van Wassenhove, Grant & Poeppel, 2007). In the same vein, one can adopt a "spatial window of integration" for when multisensory

[1] Note that this is different from the spatial ventriloquist aftereffect, where shifts in localization peak at the adapted position (Frissen et al., 2012; Frissen et al., 2003, 2005; see also Bedford, 1989).

illusions are likely to occur. The question is whether these putative spatial and temporal windows of integration are the same for spatial and temporal ventriloquism. From the literature, it appears that they are quite different. For spatial ventriloquism, several behavioral and physiological studies have shown that the spatial ventriloquist effect disappears when the audiovisual temporal alignment is outside a −100 –+300 ms window (−100 ms=sound before vision; +300 ms=sound after vision), while the horizontal spatial alignment should not exceed ~15° (Godfroy et al., 2003; Hairston, Wallace, Vaughan, Stein, Norris & Schirillo, 2003; Lewald & Guski, 2003; Slutsky & Recanzone, 2001; Radeau & Bertelson, 1977), although the specific degree of tolerated disparities could take a wide range (Wallace, Roberson, Hairston, Stein, Vaughan & Schirillo, 2004).

For temporal ventriloquism, though, these windows are quite different, since the temporal audiovisual asynchrony should not exceed ~200 ms, whereas spatial disparity plays almost no role. Concerning the temporal window, Morein-Zamir et al. (2003) reported that accessory auditory stimuli could shift the perceived time of occurrence of visual stimuli if presented within ~200 ms (Morein-Zamir et al., 2003). The *double-flash* illusion (the illusory flashing in the presence of multiple beeps) also declines when audiovisual asynchrony exceeds ~70 ms (Shams et al., 2000). Jaekl and Harris (2007) reported that temporal cross-capture of audiovisual stimuli takes place within a temporal disparity of ~125 ms (Jaekl & Harris, 2007). These relatively narrow temporal criteria for temporal ventriloquism to take effect may reflect the narrow integration time of polysensory neurons in the brain (Meredith, Nemitz & Stein, 1987; Recanzone, 2003).

The most striking difference with spatial ventriloquism, though, is that temporal ventriloquism is hardly affected by spatial discordance (Bruns & Getzmann, 2008; Keetels & Vroomen, 2008a; Recanzone, 2003; Vroomen & Keetels, 2006). Vroomen and Keetels (2006) examined this in a setup shown in Fig. 6.

Observers were asked to judge whether a lower or upper LED was presented first (a visual TOJ task) while two accessory sounds were sandwiched in an AVVA style at ±100-ms SOAs such that they improved the visual JND (= temporal ventriloquism). Crucially, the improvement by the sounds was equal when sounds came from the same location as or a different location than the lights (Fig. 6a), for static sound sounds from the same location or for dynamic sounds with apparent motion from left to right or right to left (Fig. 6b), and for sounds and lights coming from the same side or the opposite sides of central fixation (Fig. 6c). In the setup of Fig. 6c, it could also be demonstrated in a visual detection task that the lateral sounds actually captured visuospatial attention, because observers were faster to detect a flash when the sounds came from the same, rather than the opposite, side of fixation. The sounds thus captured visuospatial attention (see, e.g., Driver & Spence, 1998). Temporal ventriloquism thus appears to be independent of the spatial separation between sound and flash, despite the fact that the location of the sounds was potent enough to capture visuospatial attention (see also Keetels & Vroomen, 2008a, for similar effects with tactile–visual stimuli). For audiovisual temporal recalibration, it also appears that spatial misalignment between sound and flashes does not decrease temporal recalibration (Keetels & Vroomen, 2007), although audiovisual spatial alignment can be of importance when sounds and flashes are presented in continuous streams with an ambiguous temporal ordering (Yarrow, Roseboom & Arnold, 2011). To summarize, it appears that temporal ventriloquism has, in comparison with spatial ventriloquism, a somewhat smaller window of temporal integration but a much wider, if not a nonexisting, window of spatial integration.

The role of attention for the spatial ventriloquist effect

An important controversy regarding the mechanism of multisensory binding is the degree to which it operates automatically, without the need for attention (e.g., Talsma, Senkowski, Soto-Faraco & Woldorff, 2010). In the visual domain, adaptation effects that were once thought to be entirely stimulus-driven have since proved to be remarkably susceptible to the attentional state of the observer (e.g., Verstraten & Ashida, 2005). Does the same apply to spatial and temporal ventriloquism?

The initial evidence about the role of attention suggested that spatial ventriloquism is largely an automatic phenomenon (e.g., Alais & Burr, 2004a; Bertelson & Aschersleben, 1998; Bertelson, Pavani, Ladavas, Vroomen & de Gelder, 2000a; Bertelson, Vroomen, de Gelder & Driver, 2000b; Bonath, Noesselt, Martinez, Mishra, Schwiecker, Heinze & Hillyard, 2007; Driver, 1996; Vroomen, Bertelson & de Gelder, 2001; Arnold, 2007, s0TDa14cno1.59216scn8.2570TD(6)

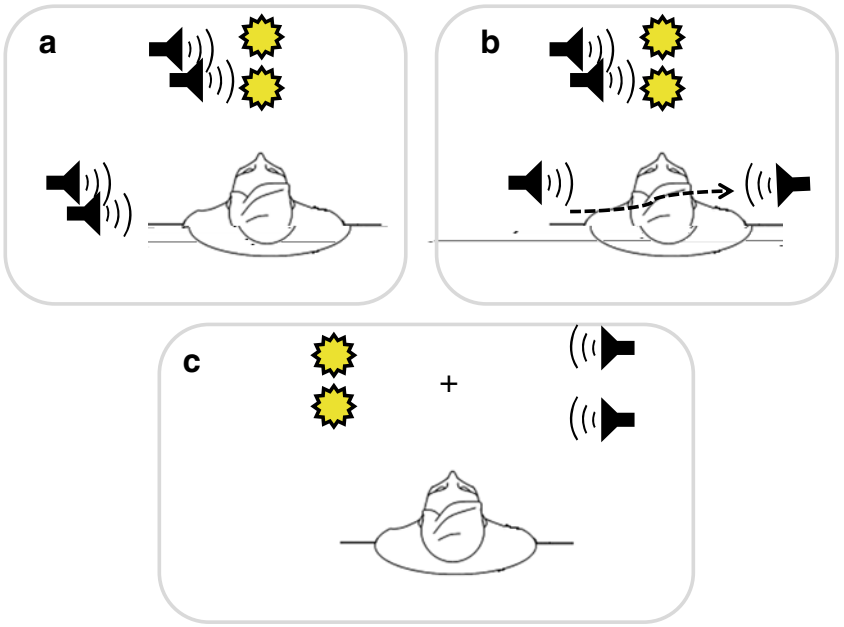

Fig. 6 Observers judged which of two flashes (*upper* or *lower*) appeared first. Sensitivity for visual temporal order improved, relative to a silent control condition, when clicks were presented in an AVVA style (temporal ventriloquism [TV]). This TV-effect was *not* affected by whether the sounds came from the same location as rather than a different location than the lights (**a**), were static rather than moving (**b**), and came from the same rather than the opposite side of fixation (**c**). The laterally presented sounds in panel c were potent because they did capture visuospatial attention, but this did not affect TV (Vroomen & Keetels, 2006)

simultaneously with the unseen flash was shifted in the direction of the visual stimulus (Bertelson et al., 2000a).

More recently, though, some authors have questioned the full automaticity of the ventriloquist effect and suggested that attention might at least have a modulating influence (Fairhall & Macaluso, 2009; Röder & Büchel, 2009; Sanabria, Soto-Faraco & Spence, 2007b). Maiworm, Bellantoni, Spence and Röder (2012) examined whether audiovisual binding, as indicated by the magnitude of the ventriloquist effect, is influenced by threatening auditory stimuli presented prior to the ventriloquist experiment. This emotional stimulus manipulation resulted in a reduction of the magnitude of the subsequently measured ventriloquist effect in both hemifields, as compared with a control group exposed to a similar attention-capturing but nonemotional manipulation. This piece of evidence was taken to show that the ventriloquist illusion is not fully automatic, although there is no straightforward explanation why it is reduced. Borjon, Shepherd, Todorov and Ghazanfar (2011) also reported a novel finding of visual gaze steering auditory spatial attention. Specifically, visual perception of eye gaze and arrow cues presented slightly before sounds shifted the apparent origin of these sounds (delivered through headphones) in the direction by the arrows or eye gaze. In both conditions, the shifts were equivalent, suggesting a generic, supramodal attentional influence by visual cues on immediate sound localization. The authors claimed that they could distinguish (in their mathematical model) a simple response bias from a genuine perceptual shift, although this seems questionable to us. A simple response bias might be that whenever an observer is unsure about sound location, he/she responds in the direction of the eyes or arrow. The sounds whose location is ambiguous should then show a shift by gaze or arrows, but not the nonambiguous clearly localizable sounds. This was the actual pattern in the data, and it seems, therefore, that future studies need to examine the exact nature of this gaze/arrow effect on sound location in more detail.

The putative role of attention for the spatial ventriloquist aftereffect

The extent to which the ventriloquist aftereffect depends on attention has also been examined. A study by Eramudugolla, Kamke, Soto-Faraco and Mattingley (2011) reported that a robust auditory spatial ventriloquist aftereffect could be induced when participants fixated on a central visual stimulus while the audiovisual adaptors (a click sound with a spatially discordant flash) were presented in the periphery. A ventriloquist aftereffect was thus obtained despite the fact that the visual inducer had not been in the focus of attention. Possibly, though, despite central fixation, attention could have slipped through to the visual inducer. In a further attempt to examine the role of focal attention, the authors used a dual-task paradigm in which participants had to detect, during the exposure phase, a visual target at central fixation that was either easy or difficult to detect. In contrast with the notion that the difficult task would deplete attentional resources necessary for adaptation, these authors found that increasing load from low to high levels did not abolish the aftereffect but, in some conditions, actually *enhanced* it. The underlying basis of this load effect on the ventriloquist aftereffect remains unknown, but the results are in clear contradiction with the notion that attention increases (or is necessary for) intersensory binding. Clearly,

though, further testing is necessary to resolve this controversy regarding the mechanisms of multisensory integration and the degree to which spatial ventriloquism might operate automatically.

The role of attention for the temporal ventriloquist effect

There is substantial evidence that the brain has "hard-wired" mechanisms for extracting spatial information from sound and vision, but the neural evidence for "hard-wired" mechanism to extract the relative timing from different senses is much weaker. In fact, explicit judgments about cross-modal temporal order can be very difficult and require high-level processing, which is quite unlike spatial judgments about sound and flashes, like pointing or saccades that are mainly driven in an automatic fashion. One simple but important observation is that detection of cross-modal synchrony can become almost impossible if the presentation rates are above 5 Hz, which is far below the temporal limits of the individual visual or auditory systems (Benjamins, van der Smagt & Verstraten, 2008; Fujisaki & Nishida, 2005, 2007). Fujisaki et al. proposed that audiovisual asynchrony perception is mediated by a 'mid-level" mechanism that first needs to extract salient auditory and visual features before making temporal cross-correlations across sensory channels (Fujisaki & Nishida, 2005, 2007, 2008, 2010), as opposed to earlier detection by specialized low-level sensors. Further evidence in support of the idea that intersensory binding needs temporally salient events comes from the "pip-and-pop" paradigm (Van der Burg, Olivers, Bronkhorst & Theeuwes, 2008). These authors reported that only transient pips (as opposed to slowly changing sounds) can make a synchronized color/luminance change in the visual periphery more salient (Van der Burg, Cass, Olivers, Theeuwes & Alais, 2010). In addition, the capturing effect of transient sounds requires that visual attention be spread over the visual field rather than focused on fixation. The visual stimulus thus needs some attentional resources before a sound can capture its onset (Van der Burg, Olivers & Theeuwes, 2012).

It also appears that the auditory capturing stimulus needs to be segregated from the background. Keetels, Stekelenburg and Vroomen (2007) examined this in a visual TOJ task in which paired lights were embedded in a train of auditory beeps (see Fig. 7).

The capturing sounds (those temporally closest to the visual flashes) had either the same features as or different features (like pitch, rhythm, or location) from the flanker sounds. The authors found that temporal ventriloquism occurred only when the two capturing sounds differed from the flankers, which made them stand out, thus demonstrating that (intramodal) grouping of the sounds in the auditory stream took priority over intersensory pairing. Audiovisual temporal ventriloquism thus requires salient auditory and visual stimuli with a sharp rise time of energy. The extent to which spatial ventriloquism requires similarly "sharp" auditory and visual transient stimuli to bias sounds in the spatial domain has, to the best of our knowledge, not been examined in a systematic way.

The role of attention for the temporal ventriloquist aftereffect

When studying audiovisual temporal recalibration, the majority of studies tried to ensure that observers actually saw the visual part of the audiovisual adapter by engaging observers in either a visual fixation task or a visual detection task of oddball stimuli that changed in luminance or size (Fujisaki et al., 2004; Keetels & Vroomen, 2007, 2008b; Navarra et al., 2009; Takahashi, Saiki & Watanabe, 2008; Vroomen et al., 2004). Heron et al. (2012; Heron, Roach, Whitaker & Hanson, 2010), though, had observers focus during adaptation on the temporal relationship of the adapting stimulus itself by using audiovisual oddball stimuli that differed in their temporal relation from the adapters. When observers selectively attended to the temporal relationship, the aftereffect was almost tripled relative to situations in which selective attention was focused on visual features of the same stimuli. Attending to the temporal order of the adapting stimuli itself thus appeared to be an effective booster of temporal recalibration, although it is important to note that diverting attention away from temporal order did not abolish the basic effect. This thus suggests that aftereffects of temporal recalibration may depend more on high-level processing than do aftereffects in spatial ventriloquism, but this idea needs further testing.

The role of cognitive factors for intersensory binding

When information is presented in two modalities, a decision has to be made (usually unconscious) about whether these two information sources represent a single object/event or multiple objects/events. This putative intersensory binding process likely involves an assessment of the degree of concordance of the total sensory inputs with a unitary source (Bertelson, 1998; Radeau, 1994a, b; Radeau & Bertelson, 1977, 1987). Only if the evidence points to a unitary source is information in the involved modalities integrated. A vexing question about this binding processing is whether higher-order cognitive factors play a role. Early studies examined factors that could bias intersensory binding or pairing toward "single" or "multiple" events. One of the factors considered was the compellingness of the situation. These early studies used fairly realistic situations,

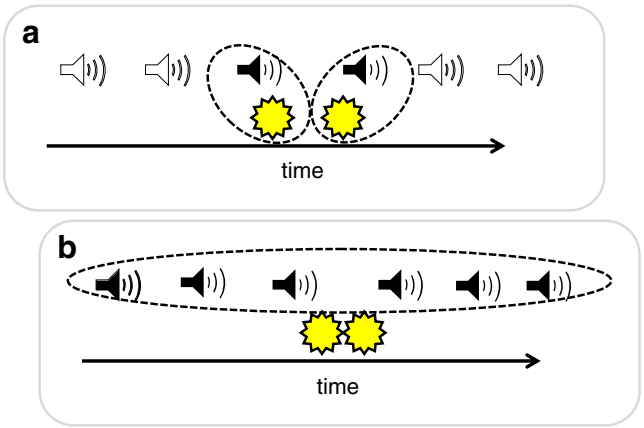

Fig. 7 **a** Sounds presented in a train of other sounds need to differ in pitch, rhythm, or location to capture flashes. **b** When the sounds are identical to the other sounds, there is no intersensory binding with the flashes, because sounds are now grouped as one stream. Within-modality auditory grouping then takes priority over cross-modal audiovisual binding (Keetels et al., 2007)

simulating real-life events such as a voice speaking and the concurrent sight of a face (e.g., Bertelson et al., 1994; Warren, Welch & McCarthy, 1981), or the sight of whistling kettles (Jackson, 1953) or of beating drums (Radeau & Bertelson, 1977, Experiment 1). Radeau and Bertelson (1977), for example, combined a voice with a realistic face (the sight of the speaker) and a simplified visual input (light flashes in synchrony with the amplitude peaks of the sound) and found that exposure to these two situations produced comparable spatial ventriloquist effects, suggesting that realism plays little if any role in spatial ventriloquism.

A critique on these older studies, though, is that comparisons were made across arbitrary stimulus classes (e.g., flashes vs. faces). Bertelson, Vroomen, Wiegeraad and de Gelder (1994) avoided this stimulus confound and obtained more direct evidence in an experiment in which observers heard, on each trial, a speech sound, either /ama/ or /ana/, from an array of seven hidden loudspeakers. At the same time, observers saw on a centrally located screen a face, either upright or inverted, that articulated /ama/ or /ana/ or remained still (the baseline; see Fig. 8).

Participants had two tasks: They pointed toward the apparent origin of the sound, and they reported what had been said. The orientation of the face had a large effect on "what" was perceived (i.e., the McGurk effect; McGurk & MacDonald, 1976), but not on "where" the sound came from, because the upright and inverted faces were equally effective in attracting the apparent location of the speech sound. The spatial ventriloquist effect was also equally big for when the sound and the face were congruent (e.g., hearing /ama/ and seeing /ama/) or incongruent (hearing /ana/ and seeing /ama/). For speech sounds, it thus appeared that the orientation of the face and statistical co-occurrence between sound and lip movements did not affect spatial ventriloquism.

Similar questions have been asked more recently by a number of investigators, but with a slightly different approach. One prediction is that for strongly paired stimuli, it is difficult to judge the relative temporal order or the relative spatial location of the individual components, because they are fused (e.g., Arrighi, Alais & Burr, 2006; Kohlrausch & van de Par, 2005; Levitin, MacLean, Matthews, Chu & Jensen, 2000; Vatakis, Ghazanfar & Spence, 2008; Vatakis & Spence, 2007, 2008). Vatakis was the first to report such a "unity" effect in the temporal domain with audiovisual speech stimuli (human voices and moving lips) that were either gender matched (i.e., voices and moving lips belonging to the same person) or mismatched (i.e., voices and moving lips belonging to a different person). When the voice and the face were gender congruent, more multisensory binding took place, leading to a "unity effect," which was evidenced by poor discrimination thresholds for audiovisual temporal asynchronies. Subsequent studies, though, showed that this phenomenon could not be observed when participants were presented with realistic nonspeech stimuli (Vatakis et al., 2008). Vroomen and Stekelenburg (2011) argued that these comparisons might suffer from the fact that stimuli were compared that differed on a number of low-level acoustic and visual dimensions. To address this concern, they used sine-wave speech (SWS) replicas of pseudowords and the corresponding video of a face that articulated these words. SWS is artificially degraded speech that, depending on instruction, is perceived as either speech or nonspeech whistles. Using these identical SWS stimuli, the authors found that listeners in speech and nonspeech

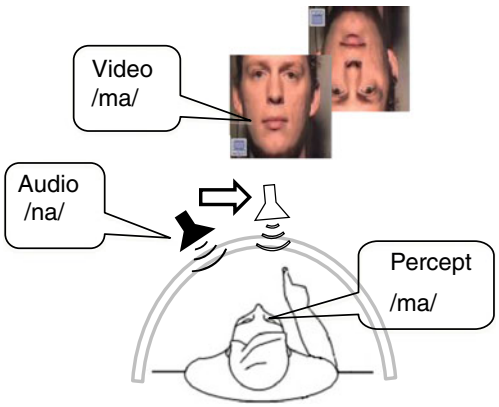

Fig. 8 Observers report what they hear (/ma/ or /na/) and point to where the sound comes from. The video of the speaker attracts, as compared with a static face, the apparent location of the sound (spatial ventriloquism), and it biases the identity of the perceived sound when sound and face are incongruent (McGurk effect). Inverting the face reduces the McGurk effect, but not spatial ventriloquism (Bertelson et al., 1994)

modes were *equally* sensitive at judging audiovisual temporal order. In contrast, when the same stimuli were used to measure the McGurk effect, they found that the phonetic content of the SWS was integrated with lipread speech only if the SWS stimuli were perceived as speech, but not if perceived as nonspeech. Listeners in speech mode, but not those in nonspeech mode, thus bound sound and vision, but judging audiovisual temporal order was unaffected by this. The evidence for or against a role of cognitive factors in intersensory binding is, at present, thus rather mixed and waits further testing.

The role of synesthetic congruency for intersensory binding

Another factor that may contribute to the "compellingness" of the situation and, thus, may affect intersensory binding is *synesthetic congruency*. Synesthetic congruency, also referred to as *crossmodal correspondences* (Spence, 2011), refers to natural or semantic correspondences between stimulus features such as pitch/loudness in the auditory dimension with size/brightness in the visual dimension (Evans & Treisman, 2010; Gallace & Spence, 2006; Guzman-Martinez, Ortega, Grabowecky, Mossbridge & Suzuki, 2012; Makovac & Gerbino, 2010; Parise & Spence, 2008, 2009, 2012; Spence, 2011; Sweeny, Guzman-Martinez, Ortega, Grabowecky & Suzuki, 2012). As an example, people usually associate higher-pitched sounds with smaller/higher/brighter/sharper objects and lower-pitched sounds with larger/lower/dimmer/rounder objects (Hubbard, 1996), although this is not always consistent across cultures (Dolscheid, Shayan, Majid & Casasanto, 2011). In accord with the unity assumption, Parise and Spence (2009) reported that the relative temporal order and relative spatial position of synesthetically congruent pitch/size pairs (high–small or low–large) were more difficult to judge than in incongruent pairs (high–large or low–small; see Fig. 9). These findings cannot be explained in terms of simple response biases or strategies, and the reduced sensitivity for congruent pairings in space and time likely may reflect a stronger binding between the unisensory signals (see also Bien, ten Oever, Goebel & Sack, 2012; Parise & Spence, 2008).

It might also be worth looking at what is known about the neural mechanisms of synesthetic congruency effects. The study by Bien et al. (2012) investigated pitch–size associations in spatial ventriloquism with transcranial magnetic stimulation (TMS) and event-related potentials (ERPs). As in Parise and Spence (2009), observers had more difficulty judging the location of low- or high-pitched tones when these tones were combined with small or large visual circles in a congruent fashion (small–high or large–low) rather than an incongruent fashion (small–low or large–high). ERP recordings showed that the P2 component at right parietal recording sites was, around 250 ms after stimulus onset, also more positive for congruent than for incongruent pairings. In addition, continuous theta-burst TMS applied over the right intraparietal sulcus diminished this congruency effect, thus indicating that the right intraparietal sulcus was likely involved in synesthetic congruency.

Audiovisual synesthetic congruency has also been examined in an fMRI study by Sadaghiani, Maier and Noppeney (2009), who examined whether sounds bias visual motion perception. They used a visual-selective attention paradigm in which observers discriminated the direction of visual motion at several levels of reliability while an irrelevant auditory stimulus was presented that was congruent, absent,

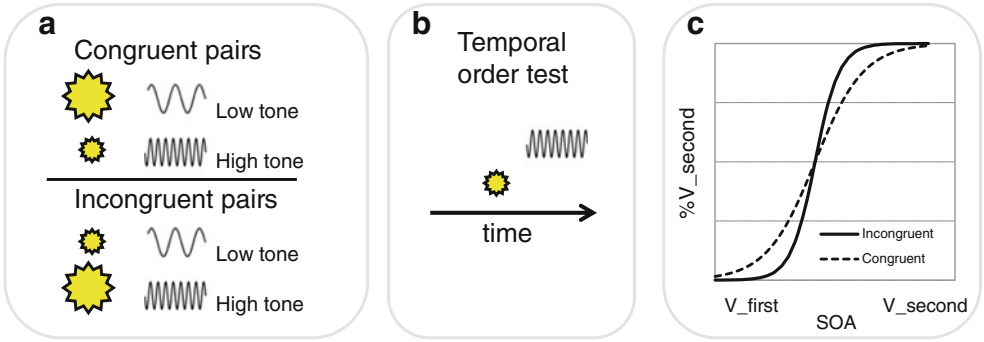

Fig. 9 **a** Large or small flashes were combined with high- or low-pitched tones in a synesthetically congruent (low–large, high–small) or incongruent (low–small, high–large) fashion. **b** Participants judged whether the sound or flalsh came second. **c** Judgments of temporal order were more sensitive (steeper curve=smaller just noticeable difference) with incongruent than with congruent pairs, presumably because congruent pairs fuse (Parise & Spence, 2009)

relative to an estimate of its noisiness, rather than one modality capturing the other (Alais & Burr, 2004b; Burr & Alais, 2006; Sato, Toyoizumi & Aihara, 2007).

Several studies have modeled audiovisual interactions with Bayesian inference. The general understanding of the Bayesian approach is that an inference is based on two factors, the likelihood and the prior. The likelihood represents the sensory noise in the environment or in the brain, whereas the prior captures the statistics of the events in the environment (Alais & Burr, 2004b; Battaglia, Jacobs & Aslin, 2003; Burr & Alais, 2006; Ernst & Banks, 2002; Sato et al., 2007; Shams & Beierholm, 2010; Shi et al., 2010; Witten & Knudsen, 2005). With a localization task, Alais and Burr (2004b) demonstrated that the ventriloquist effect results from near-optimal integration. When visual localization is good, vision dominates and captures sound, but for severely blurred visual stimuli, sound captures vision. Körding et al. (2007) extended this idea by formulating an ideal-observer model that first infers whether two sensory cues likely originate from the same event (the intersensory binding) and only then estimates the location from the two sources. They also argued that the capacity to infer causal structure is not limited to conscious, high-level cognition but is performed continually and effortless in perception (Körding et al., 2007).

Bayesian inference in spatial ventriloquism has also been examined with saccadic eye or head movements (Bell, Meredith, Van Opstal & Munoz, 2005; Van Wanrooij, Bell, Munoz & Van Opstal, 2009). When audiovisual stimuli are spatially aligned within ~20° of vertical separation and observers have to orient to an auditory or a visual target, both latency and accuracy improve, relative to the unisensory and misaligned conditions. This thus indicates a rule of "best-of-both-worlds" (Corneil, Van Wanrooij, Munoz & Van Opstal, 2002), in which observers benefit from the spatial accuracy provided by the visual component, and a shorter latency onset that is triggered by the auditory component. Prior expectations about the audiovisual spatial alignment also matter: That is, participants are faster on aligned trials when 100 % of the trials are aligned, rather than only 10 % of the trials (Van Wanrooij, Bremen & Van Opstal, 2010), thus suggesting that audiovisual binding may have a dynamic component that depends on the evidence for stimulus congruency as acquired from prior experience.

In the spatial ventriloquist effect, locations computed by vision, sound, and touch need to be coordinated. Each sensory modality, though, encodes the position of objects in different frames of reference. Visual stimuli are represented by neurons with receptive fields on the retina, auditory stimuli by neurons with receptive fields around the head, and tactile stimuli by neurons with receptive fields on the skin. A change in eye position, head position, or body posture will result in a change in the correspondence between the visual, auditory, and tactile neural responses that encode the same object. To combine these different information streams, the brain must therefore take into account the positions of the receptors in space. Pouget and colleagues have developed a framework to examine how these computations may be performed (Deneve, Latham & Pouget, 2001; Pouget, Deneve & Duhamel, 2002). This theory, targeting multisensory spatial integration and sensorimotor transformations, is based on a neural architecture that combines basis functions and attractor dynamics. Basis function units are used to solve the recoding problem and provide a biologically plausible solution for performing spatial transformations (Poggio, 1990; Pouget & Sejnowski, 1994, 1997), whereas attractor dynamics are used for optimal statistical inference. Most recently, Magosso, Cuppini and Ursino (2012) proposed a neural network model to account for spatial ventriloquism and the ventriloquist aftereffects, using two reciprocally interconnected unimodal layers (unimodal visual and auditory neurons). This model fits nicely with related biological mechanisms (Ben-Yishai, Bar-Or & Sompolinsky, 1995; Ghazanfar & Schroeder, 2006; Georgopoulos, Taira &

Lukashin, 1993; Martuzzi, Murray, Michel, Thiran, Maeder, Clarke & Meuli, 2007; Rolls & Deco, 2002; Schroeder & Foxe, 2005). Both neural network models thus suggest that neural processing may be based on probabilistic population coding. The models dovetail well with in vivo observations and nicely explain the spatial ventriloquist effect and its aftereffect

Computational approaches on temporal ventriloquism: Bayes and low-level neural models

The Bayesian approach has also been adopted to understand temporal ventriloquism (Burr et al., 2009; Hartcher-O'Brien & Alais, 2011; Ley et al., 2009; Shi et al., 2010). Interestingly, different from the findings in spatial ventriloquism, all these studies report that the quantitative fit (dominance of auditory over visual in temporal perception) was less perfect than predicted by maximum likelihood estimation, although temporal localization of audiovisual stimuli was better than for the visual sense alone.

Miyazaki, Yamamoto, Uchida and Kitazawa (2006) hypothesized that Bayesian calibration is at work during judgments regarding audiovisual temporal order but that the effect is concealed behind lag adaptation mechanism (Miyazaki et al., 2006). By canceling lag adaptation by using two pitches of sounds, they successfully uncovered "Bayesian calibration" that was working behind lag adaptation (Yamamoto, Miyazaki, Iwano & Kitazawa, 2012). In a recent study, Sato and Aihara (2009, 2011) proposed a unifying Bayesian model to account for temporal ventriloquism aftereffects. According to the model, both lag adaptation and "Bayesian calibration" can be regarded as Bayesian adaptation, the former as adaptation of the likelihood function and the latter as adaptation of the prior probability. The model incorporates trial-by-trial update rules for the size of lag adaptation and the estimate of the peak of prior distribution.

Another interesting approach to understanding the neural and computational mechanism underlying temporal recalibration has been proposed by Roach et al. (2011) for audiovisual timing and by Cai et al. (2012) for motor–visual timing. In both accounts, the relative timing of two modalities is represented by the distributed activity across a relatively small number of neurons tuned to different delays. There is an algorithm that reads out this population code. In Roach et al., exposure to a specific delay selectively reduces the gain of neurons tuned to this delay, resulting in a repulsive shift of the population response to subsequent stimuli away from the adapted value. Cai et al. adopted a somewhat different approach to simulate adaptation to delayed feedback by changing the input weights of the delay-sensitive neurons. It seems that both models are well able to explain temporal recalibration, but at this stage, there is not sufficient evidence to reject either of the models. An appeal of this approach is that it is computationally specific and that shifts in perceived simultaneity following asynchrony adaptation could actually arise from computationally similar processes known to underlie classic sensory adaptation phenomena, such as the tilt aftereffect.

Neural mechanisms in spatial ventriloquism

To what extent is a sound that originates from, say, 5° on the left neurally *identical* to a sound that actually originates from the center but is ventriloquized by 5° to the left? This question relates to an ongoing debate about whether spatial ventriloquism reveals a genuine sensory process or is just the by-product of response biases. The *sensory* account emphasizes that inputs from the two modalities are combined at early stages of processing and that only the product of the integration process is available to conscious awareness (Bertelson, 1999; Colin, Radeau, Soquet, Dachy & Deltenre, 2002; Kitagawa & Ichihara, 2002; Stekelenburg, Vroomen & de Gelder, 2004; Vroomen et al., 2001a, b; Vroomen & de Gelder, 2003). The *decisional account* suggests that inputs from the two modalities are available independently and that the behavioral results mainly originate from response interference, decisional biases, or cognitive biases (Alais & Burr, 2004a; Meyer & Wuerger, 2001; Sanabria, Spence & Soto-Faraco, 2007; Wuerger et al., 2003).

The ERP technique, owing to its excellent temporal resolution, may provide a tool to further assess the level at which the ventriloquism effect occurs. The mismatch negativity (MMN), which indexes the automatic detection of a deviant sound rarely occurring in a sequence of standard background sounds, is assumed to be elicited at a preattentive level (Carles, 2007; Näätänen, Paavilainen, Rinne & Alho, 2007). It is known that a shift in the location of a sound can evoke an MMN. From the literature, it appears that a similar MMN can be evoked by an *illusory* shift when the location of a sound is ventriloquized by a flash, thus emphasizing the sensory nature of the phenomenon. Several studies have indeed reported that the MMN is reduced when the illusory perception of auditory spatial separation is diminished by the ventriloquist effect (Colin et al., 2002), while others have reported that an MMN can be evoked when an illusory shift (but physically stationary) sound is induced by a concurrent flash (Stekelenburg & Vroomen, 2009; Stekelenburg et al., 2004).

Bonath et al. (2007) combined ERPs with event-related functional magnetic resonance imaging (fMRI) to demonstrate that a precisely timed biasing of the left–right balance of auditory cortex activity by the discrepant visual inputs underlies the ventriloquist illusion. ERP recordings showed

that the presence of the ventriloquist illusion was associated with a laterally biased cortical activity between 230 and 270 ms (i.e., the N260 component) as revealed in auditory–visual interaction waveforms. This N260 component was colocalized with a lateralized blood-oxygen-level-dependent (BOLD) response located in the posterior/medial region of the auditory cortex in the planum temporale (Bonath et al., 2007). Bruns and Röder (2010a) also substantiated the N260 as a signature for audiotactile spatial discrepancies. Furthermore, Bertini, Leo, Avenanti and Ladavas (2010) used repetitive transcranial magnetic stimulation (rTMS) to investigate an auditory localization task and replicated the finding that cortical visual processing in the occipital cortex modulates the ventriloquism effect (Bonath et al., 2007). From these results, it thus appears that sound localization is biased by vision and touch around 260 ms, with neural consequences in the auditory cortex (see also Bien et al., 2012). It should be noted, though, that this time course is relatively late if compared with the initial processing of sound location that already takes place at the level of the brainstem. A sound from central location that is ventriloquized by 5° to the left is, at least in the initial processing stages, thus not identical to a sound that originates from 5° on the left. Neural identity—or better, neural similarity—for these two sounds likely occurs later, around 260 ms, in the auditory cortex. It remains for future studies, though, to examine the exact nature of this.

Single neuron approach

Groundbreaking neurophysiological studies of the superior colliculus (SC) have established many principles of our understanding of multisensory processing at the level of single neurons (Meredith & Stein, 1983), and they continue to improve our understanding of multisensory integration in general (Stein & Stanford, 2008). This endeavor has made it possible to examine the mechanism of spatial ventriloquism on the single neuron basis. However, in order to probe the underlying neural mechanisms of the ventriloquist effect at the single neuron level, a suitable animal model where invasive studies can be conducted needs to be identified. Studies have documented several critical regions of the nonhuman primate brain that have multisensory responses, including the superior temporal sulcus (e.g., Benevento, Fallon, Davis & Rezak, 1977; Bruce, Desimone & Gross, 1981, Cusick, 1997), the parietal lobe (e.g., Cohen, Russ & Gifford, 2005, Mazzoni, Bracewell, Barash & Andersen, 1996, Stricanne, Andersen & Mazzoni, 1996), and the frontal lobe (Benevento et al., 1977; Russo & Bruce, 1994). Each of these areas could potentially be involved in multisensory processing that leads to the spatial ventriloquism effect. Direct evidence is lacking, though, due to lacking techniques and experimental methodology, given that it is difficult to have an animal perform a task in which an illusion may (or may not) occur, especially for a one-shot discrimination task of stimulus localization (Recanzone & Sutter, 2008).

Concluding remarks

The spatial and temporal ventriloquist illusions continue to serve as extremely valuable tools for studying multimodal integration. However, they also call for further endeavors of research. One challenge for the near future will be to design experiments that assess spatial and temporal ventriloquism in naturalistic environments so as to examine the role of cognitive factors for intersensory pairing. Another challenge is to specify how the brain represents stimulus reliability of different modalities and how it dynamically weighs stimulus timing and location. Bayesian observers and neural network models assume a correspondence between the architecture of the models and its underlying biological substrates (Ma & Pouget, 2008; Magosso et al., 2012; Stein & Stanford, 2008), but this awaits further proof. This "linking" holds the promise that a potential unifying theory based on an integrative "behavior–neurophysiology–computation" model is awaiting that will give a comprehensive depiction of spatial and temporal ventriloquism. Such a model would clearly facilitate our understanding of some of the basic principles of intersensory interactions in general, even if such understanding needs continuous refining.

Acknowledgements We would like to thank Chris Olivers and two anonymous reviewers for their comments that helped us to improve the article.

References

Alais, D., & Burr, D. (2004a). No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Research, 19,* 185–194.

Alais, D., & Burr, D. (2004b). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262.

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision, 6,* 260–268.

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology, 50,* 157–163.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 20,* 1391–1397.

Bedford, F. L. (1989). Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology. Human Perception and Performance, 15,* 232–248.

Benjamins, J. S., van der Smagt, M. J., & Verstraten, F. A. (2008). Matching auditory and visual signals: Is sensory modality just another feature? *Perception, 37,* 848–858.

Bell, A. H., Meredith, M. A., Van Opstal, A. J., & Munoz, D. P. (2005). Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *Journal of Neurophysiology, 93,* 3659–3673.

Benevento, L. A., Fallon, J., Davis, B. J., & 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. *Experimental Neurology, 57,* 849–872.

Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences, 92,* 3844–3848.

Bertelson, P. (1998). Starting from the ventriloquist: The perception of multimodal events. In M. Sabourin, F. I. M. Craik, & M. Robert (Eds.), *Advances in psychological science. Vol.2: Biological and cognitive aspects* (pp. 419–439). Sussex: Psychology Press.

Bertelson, P. (1999). Ventriloquism: A case of cross-modal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347–362). Amsterdam: Elsevier.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review, 5,* 482–489.

Corneil, B. D., Van Wanrooij, M., Munoz, D. P., & Van Opstal, A. J. (2002). Auditory-visual interactions subserving goal-directed saccades in a complex scene. *Journal of Neurophysiology, 88,* 438–454.

Cusick, C. G. (1997). The superior temporal polysensory region in monkeys. *Cerebral Cortex, 12,* 435–468.

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience, 4,* 826–831.

Di Luca, M., Machulla, T., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision, 9,* 1–16.

Dionne, J. K., Meehan, S. K., Legon, W., & Staines, W. R. (2010). Crossmodal influences in somatosensory cortex: Interaction of vision and touch. *Human Brain Mapping, 31,* 14–25.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9,* 719–721.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2011). The thickness of musical pitch: Psychophysical evidence for the Whorfian hypothesis. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 537–542.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech due to lip-reading. *Nature, 381,* 66–68.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences, 2,* 254–262.

Eramudugolla, R., Kamke, M. R., Soto-Faraco, S., & Mattingley, J. B. (2011). Perception load influences auditory space perception in the ventriloquist aftereffect. *Cognition, 118,* 62–74.

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

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10,* 1–12.

Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience, 29,* 1247–1257.

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

Forster, B., & Eimer, M. (2005). Vision and gaze direction modulate tactile processing in somatosensory cortex: Evidence from event-related brain potentials. *Experimental Brain Research, 165,* 8–18.

Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology, 18,* 1262–1266.

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2003). The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychologica, 113,* 315–327.

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2005). The aftereffects of ventriloquism: Generalization across sound-frequencies. *Acta Psychologica, 118,* 93–100.

Frissen, I., Vroomen, J., & de Gelder, B. (2012). The aftereffects of ventriloquism: The time course of the visual recalibration of auditory localization. *Seeing and Perceiving, 25,* 1–14.

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

Fujisaki, W., & Nishida, S. (2005). Temporal frequency characteristic of synchrony-asynchrony discrimination of audiovisual signals. *Experimental Brain Research, 166,* 455–464.

Fujisaki, W., & Nishida, S. (2007). Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision Research, 47,* 1075–1093.

Fujisaki, W., & Nishida, S. (2008). Top-down feature-based selection of matching features for audio-visual synchrony discrimination. *Neuroscience Letters, 433,* 255–230.

Fujisaki, W., & Nishida, S. (2010). A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proceedings of the Royal Society B: Biological Sciences, 277,* 2281–2290.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68,* 1191–1203.

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *The American Journal of Psychology, 72,* 521–529.

Georgopoulos, A. P., Taira, M., & Lukashin, A. (1993). Cognitive neurophysiology of the motor cortex. *Science, 260,* 47–52.

Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception, 36,* 1089–1103.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences, 10,* 278–285.

Godfroy, M., Roumes, C., & Dauchy, P. (2003). Spatial variations of visual-auditory fusion areas. *Perception, 32,* 1233–1245.

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual shape frequency and auditory amplitude-modulation rate. *Current Biology, 22,* 383–388.

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience, 15,* 20–29.

Hartcher-O'Brien, J., & Alais, D. (2011). Temporal ventriloquism in a purely temporal context. *Journal of Experimental Psychology. Human Perception and Performance, 37,* 1383–1395.

Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American, 213,* 84–94.

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

Heron, J., Roach, N. W., Whitaker, D., & Hanson, J. V. M. (2010). Attention regulates the plasticity of multisensory timing. *European Journal of Neuroscience, 31,* 1755–1762.

Hidaka, S., Manaka, Y., Teramoto, W., Sugita, Y., Miyauchi, R., Gyoba, J., ... Iwaya, Y. (2009). Alternation of Sound Location Induces Visual Motion Perception of a Static Object. *PLoS One, 4*(12), e8188.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. New York.: Wiley.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology, 109,* 219–238.

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology, 5,* 52–65.

Jaekl, P. M., & Harris, L. R. (2007). Auditory-visual temporal integration measured by shifts in perceived temporal location. *Neuroscience Letters, 417,* 219–224.

Kacelnik, O., Walton, M. E., Parsons, C. H., & King, A. J. (2002). Visual-auditory interactions in sound localization: From behavior to neural substrate. *Proceedings of the Neural Control of Movement Satellite Meeting*, 21

Kafaligonul, H., & Stoner, G. R. (2010). Auditory modulation of visual apparent motion with short spatial and temporal intervals. *Journal of Vision, 10,* 1–13.

Kitajima, N., & Yamashita, Y. (1999). Dynamic capture of sound motion by light stimuli moving in three-dimensional space. *Perceptual and Motor Skills, 89,* 1139–1158.

Keetels, M., Stekelenburg, J., & Vroomen, J. (2007). Auditory grouping occurs prior to intersensory pairing: Evidence from temporal ventriloquism. *Experimental Brain Research, 180,* 449–456.

Keetels, M., & Vroomen, J. (2008a). Tactile-visual temporal ventriloquism: No effect of spatial disparity. *Perception & Psychophysics, 70,* 765–771.

Keetels, M., & Vroomen, J. (2008b). Temporal recalibration to tactile-visual asynchronous stimuli. *Neuroscience Letters, 430,* 130–134.
Keetels, M., & Vroomen, J. (2012). Exposure to delayed visual feedback of the hand changes motor-sensory synchrony perception. *Experimental Brain Research, 219,* 431–440.
King, A. J., Doubell, T. P., & Skaliora, I. (2004). Epigenetic factors that align visual and auditory maps in the ferret midbrain. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes*, (pp. 599–612). MIT Press: Cambridge.
Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416,* 172–174.
Knudsen, E. I., Knudsen, P. F., & Esterly, S. D. (1982). Early auditory experience modifies sound localization in barn owls. *Nature, 295,* 238–240.
Knudsen, E. I., & Knudsen, P. F. (1985). Vision Guides the adjustment of auditory localization in young barn owls. *Science, 230,* 545–548.
Knudsen, E. I., & Knudsen, P. F. (1989). Vision calibrates sound localization in developing barn owls. *Journal of Neuroscience, 9,* 3306–3313.
Kohlrausch, A., & van de Par, S. (2005). Audio–visual interaction in

Radeau, M. (1994a). Auditory-visual interaction and modularity. *Current Psychology of Cognition., 13,* 3–51.

Radeau, M. (1994b). Ventriloquism against audio-visual speech: Or, where Japanese-speaking barn owls might help. *Current Psychology of Cognition, 13,* 124–140.

Radeau, M., & Bertelson, P. (1969). Adaptation à un déplacement prismatique sur la base de stimulations exafférentes en conflit. *Psychologica Belgica, 9,* 133–140.

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *Quarterly Journal of Experimental Psychology, 26,* 63–71.

Radeau, M., & Bertelson, P. (1976). The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception & Psychophysics, 20,* 227–235.

Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics, 22,* 137–146.

Radeau, M., & Bertelson, P. (1978). Cognitive factors and adaptation to auditory-visual discordance. *Perception & Psychophysics, 23,* 341–343.

Radeau, M., & Bertelson, P. (1987). Auditory–visual interaction and the timing of inputs: Thomas (1941) revisited. *Psychological Research, 49,* 17–22.

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences, 95,* 869–875.

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

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

Recanzone, G. H., & Sutter, M. L. (2008). The biological basis of audition. *Annual Review of Psychology, 59,* 119–142.

Redding, G. M., & Wallace, B. (1997). *Adaptive spatial alignment*. Hillsdale: Lawrence Erlbaum.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12,* 969–992.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology. Human Perception and Performance, 28,* 1085–1099.

Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2011). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society B: Biological Sciences, 278,* 1314–1322.

Rock, I., & Victor, J. (1964). Vision and Touch: An experimentally created conflict between the two senses. *Science, 143,* 594–596.

Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision*. Oxford: Oxford University Press.

Röder, B., & Büchel, C. (2009). Multisensory interactions within and outside the focus of visual spatial attention (commentary on Fairhall & Macaluso). *European Journal of Neuroscience, 29,* 1245–1246.

Roseboom, W., & Arnold, D. H. (2011). Twice Upon a Time: Multiple concurrent temporal recalibration of audiovisual speech. *Psychological Science, 22,* 872–877.

Russo, G. S., & Bruce, C. J. (1994). Frontal eye field activity preceding aurally guided saccades. *Journal of Neurophysiology, 71,* 1250–1253.

Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *Journal of Neuroscience, 29,* 6490–6499.

Sanabria, D., Spence, C., & Soto-Faraco, S. (2007a). Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: A signal detection study. *Cognition, 102,* 299–310.

Sanabria, D., Soto-Faraco, S., & Spence, C. (2007b). Spatial attention and audiovisual interactions in apparent motion. *Journal of Experimental Psychology. Human Perception and Performance, 33,* 927–937.

Sato, Y., & Aihara, K. (2009). Integrative Bayesian model on two opposite types of sensory adaptation. *Artificial life and Robotics, 14,* 289–292.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Computation, 19,* 3335–3355.

Sato, Y., & Aihara, K. (2011). A Bayesian Model of Sensory Adaptation. *PLoS One, 6*(4), e19377.

Scheier, C. R., Nijhawan, R., & Shimojo, S. (1999). Sound alters visual temporal resolution. *Investigative Ophthalmology & Visual Science, 40,* 4169.

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology, 15,* 454–458.

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

Serino, A., Farnè, A., Rinaldesi, M. L., Haggard, P., & Làdavas, E. (2007). Can vision of the body ameliorate impaired somatosensory function? *Neuropsychologia, 45,* 1101–1107.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cogntive Sciences, 14,* 425–432.

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

Shi, Z., Chen, L., & Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: The influence of time interval. *Experimental Brain Research, 203,* 723–735.

Shipley, T. (1964). Auditory Flutter-Driving of Visual Flicker. *Science, 145,* 1328–1330.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12,* 7–10.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research, 4,* 139–146.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004a). Congruency effects between auditory and tactile motion: Extending the phenomenon of cross-modal dynamic capture. *Cognitive, Affective, & Behavioral Neuroscience, 4,* 208–217.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004b). Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology. Human Perception and Performance, 30,* 330–345.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2005). Assessing automaticity in the audiovisual integration of motion. *Acta Psychologica, 118,* 71–92.

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Percepiton & Psychophysics, 73,* 971–995.

Stein, B. E. (2012). *The new handbook of multisensory Processes*. Cambridge: MIT Press.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge: MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9,* 255–266.

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters, 357,* 163–166.

Stekelenburg, J. J., Sugano, Y., & Vroomen, J. (2011). Neural correlates of motor-sensory temporal recalibration. *Brain Research, 1397,* 46–54.

Stekelenburg, J. J., & Vroomen, J. (2009). Neural correlates of audiovisual motion capture. *Experimental Brain Research, 198,* 383–390.

Stetson, C., Cui, X., Montague, R. R., & Eagleman, D. M. (2006). Motor-Sensory Recalibration leads to an illusory reversal of action and sensation. *Neuron, 51,* 651–659.

Strelnikov, K., Rosito, M., & Barone, P. (2011). Effect of audiovisual training on monaural spatial hearing in horizontal plane. *PLoS One, 6*(3), e18344.

Stricanne, B., Andersen, R. A., & Mazzoni, P. (1996). Eye-centered, head-centered and intermediate coding of remembered sound locations in area LIP. *Journal of Neurophysiology, 76,* 2071–2076.

Sugano, Y., Keetels, M., & Vroomen, J. (2010). Adaptation to motor-visual and motor-auditory temporal lags transfer across modalities. *Experimental Brain Research, 201,* 393–399.

Sugano, Y., Keetels, M., & Vroomen, J. (2012). The build-up and transfer of sensorimotor temporal recalibration measured via a synchronization task. *Frontiers in Psychology, 3,* 246.

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2012). Sounds exaggerate visual shape. *Cognition, 124,* 194–200.

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

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14,* 400–410.

Taylor-Clarke, M., Kennett, S., & Haggard, P. (2002). Vision modulates somatosensory cortical processing. *Current Biology, 12,* 233–236.

Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y., & Suzuki, Y. (2012). Sounds can alter the perceived direction of a moving visual object. *Journal of Vision, 12,* 1–12.

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology. Human Perception and Performance, 34,* 1053–1065.

Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS One, 5*(5), e10664.

Van der Burg, E., Olivers, C. N. L., & Theeuwes, J. (2012). The size of the attentional window modulates capture by audiovisual events. *PLoS One, 7*(7), e39137.

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

Van Wanrooij, M. M., Bremen, P., & Van Opstal, A. J. (2010). Acquired prior knowledge modulates audiovisual integration. *European Journal of Neuroscience, 31,* 1763–1771.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45,* 598–607.

Vatakis, A., Ghazanfar, A., & Spence, C. (2008). Facilitation of multisensory integration by the 'unity assumption': Is speech special? *Journal of Vision, 8,* 1–11.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics, 69,* 744–756.

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica, 127,* 12–23.

Verstraten, F. A. J., & Ashida, H. (2005). Attention-based motion perception and motion adaptation: What does attention contribute? *Vision Research, 45,* 1313–1319.

von Helmholtz, H. (1962). Treatise on physiological optics. Dover Publications

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica, 108,* 21–33.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001b). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics, 63,* 651–659.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology. Human Perception and Performance, 26,* 1583–1590.

Vroomen, J., & de Gelder, B. (2003). Visual motion influences the contingent auditory motion aftereffect. *Psychological Science, 14,* 357–361.

Vroomen, J., & de Gelder, B. (2004a). Perceptual Effects of Cross-modal Stimulation: Ventriloquism and the Freezing Phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The Handbook of multisensory processes* (pp. 141–150). Cambridge: MIT Press.

Vroomen, J., & de Gelder, B. (2004b). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology. Human Perception and Performance, 30,* 513–518.

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

Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology. Human Perception and Performance, 32,* 1063–1071.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72,* 871–884.

Vroomen, J., & Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: Not that special. *Cognition, 118,* 78–86.

Wallace, M. T., & Stein, B. E. (2007). Early experience determines how the senses will interact. *Journal of Neurophysiology, 97,* 921–926.

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

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics, 30,* 557–564.

Watanabe, K., & Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science, 12,* 109–116.

Welch, R. B. (1978). *Perceptual modification: Adapting to altered sensory environments*. New York.: Academic Press.

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

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

Wilson, E. C., Reed, C. M., & Braida, L. D. (2009). Integration of auditory and vibrotactile stimuli: Effects of phase and stimulus-onset asynchrony. *Journal of .Acoustic Society of America, 126,* 1960–1974.

Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron, 48,* 489–496.

Wozny, D. R., & Shams, L. (2011a). Computational characterization of visually induced auditory spatial adaptation. *Frontiers in Integrative Neuroscience, 5,* 75.

Wozny, D. R., & Shams, L. (2011b). Recalibration of auditory space following milliseconds of cross-modal discrepancy. *The Journal of Neuroscience, 31,* 4607–4612.

Wuerger, S. M., Hofbauer, M., & Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics, 65,* 1188–1196.

Yamamoto, S., Miyazaki, M., Iwano, T., & Kitazawa, S. (2012). Bayesian calibration of simultaneity in audiovisual temporal order judgment. *PLoS One, 7*(7), e40379.

Yarrow, K., Roseboom, W., & Arnold, D. W. (2011). Spatial grouping resolves ambiguity to drive temporal recalibration. *Journal of Experimental Psychology. Human Perception and Performance, 37,* 1657–1661.

Zwiers, M. P., Van Opstal, A. J., & Paige, G. D. (2003). Plasticity in human sound localization induced by compressed spatial vision. *Nature Neuroscience, 6,* 175–181.