



## The supramodal brain: implications for auditory perception

Lawrence D. Rosenblum, James W. Dias & Josh Dorsi

**To cite this article:** Lawrence D. Rosenblum, James W. Dias & Josh Dorsi (2016): The supramodal brain: implications for auditory perception, Journal of Cognitive Psychology, DOI: [10.1080/20445911.2016.1181691](https://doi.org/10.1080/20445911.2016.1181691)

**To link to this article:** <http://dx.doi.org/10.1080/20445911.2016.1181691>



Published online: 13 May 2016.



Submit your article to this journal [↗](#)



Article views: 3



View related articles [↗](#)



View Crossmark data [↗](#)

## The supramodal brain: implications for auditory perception

Lawrence D. Rosenblum, James W. Dias and Josh Dorsi

Department of Psychology, University of California, Riverside, CA, USA

### ABSTRACT

The perceptual brain is designed around multisensory input. Areas once thought dedicated to a single sense are now known to work with multiple senses. It has been argued that the multisensory nature of the brain reflects a cortical architecture for which task, rather than sensory system, is the primary design principle. This supramodal thesis is supported by recent research on human echolocation and multisensory speech perception. In this review, we discuss the behavioural implications of a supramodal architecture, especially as they pertain to auditory perception. We suggest that the architecture implies a degree of perceptual parity between the senses and that cross-sensory integration occurs early and completely. We also argue that a supramodal architecture implies that perceptual experience can be shared across modalities and that this sharing should occur even without bimodal experience. We finish by briefly suggesting areas of future research.

### ARTICLE HISTORY

Received 15 October 2015  
Accepted 17 April 2016

### KEYWORDS

Multisensory; speech;  
auditory

It may be time for us to broaden how we study audition. As evidence mounts for a perceptual brain designed around multisensory functioning, more research should examine audition in its multisensory context. Of course, audition's function as support for vision is a long-standing research area. Much research has examined how audition can drive visual attention and enhance degraded visual input. But new multisensory research suggests that not only do the senses support each other, in an important way, they can *become* each other.

Research is showing a profound neurophysiological and behavioural flexibility across the senses, including audition. The perceptual brain is often agnostic when it comes to the individual senses and can change its cross-sensory functioning with even minor sensory decrements. This fact could have critical implications for how we understand auditory function. Audition may be performing a much wider range of crossmodal functions than we typically consider. There is evidence that the brain can use audition for functions typically thought to be vision's domain. Evidence also shows that audition can effortlessly integrate with the other senses in often surprising ways. This fact may indicate a commonality in information accessed

across modalities. Finally, there is evidence that audition can share its perceptual experience with the other senses and vice versa. These facts could be a result of a perceptual brain whose organisation is based more on type of task than on type of sensory input.

In this review, we discuss the ramifications of crossmodal flexibility for our understanding of audition. We will review the recent research demonstrating this flexibility as manifest in both neurophysiological and behavioural contexts. In the first section, we present research showing neurophysiological crossmodal plasticity and activation in the brain. This research involves instances of audition activating visual brain areas, and vision and touch activating auditory brain areas. This research also shows that the visual cortex can use auditory input to recover many of the same dimensions it would normally glean from vision. These characteristics call for a consideration of a *supramodal* conception of the perceptual brain, for which task, rather than sensory modality, is the primary organising principle.

We follow this section by discussing the behavioural ramifications of the supramodal approach, with an emphasis on audition. We will present

behavioural research showing that audition can sometimes take on the role of vision, and vice versa. We then discuss the ease with which audition can integrate with the other senses. We will argue that the ubiquity, automaticity, and primacy with which audition integrates suggest a brain that treats information similarly across the senses. Finally, we discuss evidence showing that audition can share its perceptual experience with the other senses and vice versa. We argue that some degree of perceptual learning involves extracting amodal perceptual primitives that are not tied to a given sense modality. In each section, we argue that these findings call for a shift in emphasis in auditory perception research to examine more of its multisensory nature.

Throughout, we will be drawing heavily on speech perception research. The reason for this choice is that speech is one of the most developed topics of multisensory research both methodologically and conceptually. We will draw connections to demonstrate how the theoretical implications of a supramodal speech mechanism are relevant to other areas of perception.

## **Crossmodal compensatory plasticity and activation**

### ***Compensatory plasticity in blind individuals***

It has long been known that sensory loss can change the brain in compensatory ways (for a review, see Collignon, Voss, Lassonde, & Lepore, 2009). However, recent evidence demonstrates the surprising ubiquity and sophistication of this compensation. For audition, some of the most striking recent examples come from echolocation in blind individuals. Thaler, Arnott, and Goodale (2011) found that both early and late-blind perceivers recruit visual cortex when identifying objects from reflected sound. These researchers have also found that the area of visual cortex recruited depends on whether the echolocation task involves identifying an object's shape or surface material (Arnott, Thaler, Milne, Kish, & Goodale, 2013). Importantly, these areas were the same as those recruited during *visual* shape and surface material recognition (accordingly) by sighted perceivers. A more recent study by these researchers corroborated these findings using a more stringent test of surface material detection (Milne, Arnott, Kish, Goodale, & Thaler, 2015). Again, scans revealed that at least for expert

echolocators, the same areas typically involved in visual texture perception are recruited when judging surface material from echolocation.

Research on the use of auditory prostheses by blind individuals is consistent with these findings. When blind individuals use sound-based prostheses to make spatial judgments, areas of visual cortex typically involved in spatial judgments are active (e.g. Collignon, Lassonde, Lepore, Bastien, & Veraart, 2007). Other auditory prosthesis research has found evidence for the classic ventral "what" and dorsal "where" division in visual processing based on sound input (e.g. Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009; Striem-Amit et al., 2011). Next, upon learning to use auditory prostheses to recognise other people's body parts, blind individuals show activation in a visual area associated with human body recognition (Striem-Amit & Amedi, 2014). Finally, auditory prostheses allow blind individuals to recognise number forms, and this recognition activates visual areas known to be involved in visual recognition of the same (Abboud, Maidenbaum, Dehaene, & Amedi, 2015). Interestingly, the recruitment of visual cortex does not seem involved in just any auditory task for which blind individuals have shown enhancement. Research using repetitive transcranial magnetic stimulation to disrupt visual cortex has found that while auditory spatial tasks (sound localisation; spatial recognition using an auditory prosthesis) are impeded, simple intensity and frequency discrimination are spared (Collignon et al., 2007). These findings suggest that the compensatory function of visual cortex for blind individuals is specialised to functional and often spatial tasks.

The implications of these findings for audition are profound. Not only can an auditory skill (e.g. echolocation) make use of visual cortex, it can do so in a way that takes advantage of the *task-specificity* of that cortex (e.g. surface material vs. shape detection). In an important way, auditory input is treated as visual input. Certainly, the obvious differences in energy media, peripheral anatomy, perceptual resolution, and phenomenology cannot be ignored. But there is a very real sense in which the perceptual brain allows audition to act as vision.

There are other cross-sensory plasticity examples involving tactile stimulation. Braille reading by blind individuals recruits visual cortex areas known to be involved in normal text reading by the sighted (Reich, Maidenbaum, & Amedi, 2012). Braille

reading also activates the early visual areas typically associated with foveal vision (Cheung, Fang, & He, 2009). Other research shows that a tactile prosthesis can activate visual areas responsive to map learning and navigation as blind individuals use the prosthesis to navigate (Kupers, Chebat, Madsen, Paulson, & Ptito, 2010).

### ***Compensatory plasticity in deaf individuals***

Some analogous findings have been found in deaf individuals. It has been known for some time that visual stimulation (e.g. sign language) can induce activity in the secondary and tertiary auditory cortices of deaf individuals (see Bavelier, Dye, & Hauser, 2006 for a review). However, more recent research shows that *primary* auditory cortex is also activated by visual and somatosensory stimulation, as well as its combination (Allman, Keniston, & Meredith, 2009; Auer, Bernstein, Sungkarat, & Singh, 2007; Karns, Dow, & Neville, 2012; Meredith & Allman, 2012; Rebillard, Carlier, Rebillard, & Pujol, 1977; but see Kral, Schröder, Klinke, & Engel, 2003; Stewart & Starr, 1970; and see Butler & Lomber, 2013 for a review). Other research shows that the deaf individuals' well known enhanced sensitivity to the visual periphery is supported by activation in primary auditory cortex (Scott, Karns, Dow, Stevens, & Neville, 2014). Deaf individuals also invoke primary auditory cortex to better detect visual changes in motion direction (Bottari et al., 2014). Finally, auditory cortex is activated by face stimuli in deaf individuals, with the degree of activation correlated with face recognition skills (Stropahl et al., 2015).

These findings are consistent with animal research. Lomber, Meredith, and Kral (2010) report that congenitally deaf cats are more accurate at performing visual localisation tasks and have a lower visual movement detection threshold, compared to hearing cats. Importantly, these authors found that deactivating the auditory cortex with cooling reduced these advantages. There is also evidence that in deaf ferrets, the auditory cortex becomes active in response to tactile information, such as being touched by cotton probes (Allman et al., 2009; Meredith & Allman, 2012). These latter findings are consistent with imaging data collected from deaf humans that show activity in the auditory cortex in response to vibrotactile (Auer et al., 2007) as well as visual-tactile stimuli (e.g. Karns et al., 2012).

### ***Crossmodal activation in sensory-intact individuals***

While some of these crossmodal activation effects depend on the age of sensory loss, some effects can occur with very brief sensory deprivation. For example, blindfolded sighted individuals can be trained to use the tongue display unit—a tactile-based sensory substitution device—over the course of a few hours (e.g. Reich et al., 2012). Research shows that after this training, sighted perceivers will recruit areas sensitive to visual motion when using the device for motion detection (Matteau, Kupers, Ricciardi, Pietrini, & Ptito, 2010). Other research shows that as little as two hours of visual deprivation in sighted perceivers allows tactile stimulation to induce activity in visual cortex (Weisser, Stilla, Peltier, Hu, & Sathian, 2005). Thus, crossmodal plasticity is not limited to instances of long-term deprivation and can occur in normal observers.

Returning to auditory-related examples, the speech perception literature has provided some of the earliest examples of crossmodal activation in normal perceivers (for a review, see van Wassenhove, 2013). Early on, Calvert and her colleagues showed that lipreading from silent faces activates auditory cortex (e.g. Calvert et al., 1997; and see also Campbell, 2008; Pekkola et al., 2005). Later research showed that visual speech information modulates activity as early as auditory brainstem (Musacchia, Sams, Nicol, & Kraus, 2006). Other evidence shows that changes in visual speech information (e.g. from “pa” to “ka”) when paired with constant auditory segments, can modulate activity in auditory cortex in a way similar to that induced by changes in the auditory segments alone (e.g. Callan, Callan, Kroos, & Vatikiotis-Bateson, 2001; Colin et al., 2002; Sams et al., 1991). Finally, speeding of auditory cortex responses can occur when perceivers also see the face speaking, and, surprisingly, if they instead *feel* the face speaking (e.g. Treille, Vilain, & Sato, 2014). Regarding this latter finding, it is known that normal perceivers can touch an articulating face to perceive speech, and that this tactile speech can integrate with heard speech (Fowler & Dekle, 1991). Recent research shows that touching a face during articulation can speed auditory cortex reactions to auditory speech in the same way as can visual speech (Treille, Cordeboeuf, Vilain, & Sato, 2014; Treille, Vilain, et al., 2014). Regardless, these speech findings suggest that

crossmodal activation of primary sensory cortices occurs in normal perceivers.

These findings in normal human listeners are consistent with those in sensory-intact animals. For example, in rhesus and macaque monkeys, the caudomedial and posterior auditory association cortex respectively respond to somatic and visual stimuli (Schroeder et al., 2001; Schroeder & Foxe, 2002). Similarly, the auditory cortex of ferrets is responsive to visual stimuli (Bizley, Nodal, Bajo, Nelken, & King, 2007). Interestingly, these visual sensitive areas within the auditory cortex were just as responsive to auditory stimuli as were areas that exclusively responded to auditory stimuli. Finally, work with primates indicates that many of these crossmodal responses in the auditory cortex are supported by connections with multisensory processing areas (Hackett et al., 2007; Smiley et al., 2007; see also Smiley & Falchier, 2009 for a review).

### Supramodal brain architecture

Based on these striking compensatory plasticity and crossmodal activation findings, some researchers have offered a new interpretation of the perceptual brain (e.g. Pascual-Leone & Hamilton, 2001; Reich et al., 2012; Ricciardi, Bonino, Pellegrini, & Pietrini, 2014; Striem-Amit et al., 2011; see also Fowler, 2004; Rosenblum, 2013). These researchers propose that the perceptual brain is organised according to *task-specific*,<sup>1</sup> rather than *sensory-specific*, areas. This *supramodal* (also known as *meta-modal* and *task-machine*) principle has been used to explain many of the phenomena described above. Thus, the supramodal nature of the brain allows a specific area of visual cortex to detect surface material (vs. shape), regardless of whether the input comes from vision or audition (echolocation). Similarly, areas of auditory cortex are involved in speech recognition tasks regardless of whether the information comes from sound, sight, or even touch.

Of course, the flexibility of the task-based architecture should not be overstated. Much of the literature shows that degree of crossmodal plasticity is dependent on when sensory loss occurs (e.g. Merabet & Pascual-Leone, 2010). It is now well known that for congenitally deaf children, cochlear implantation is more successful if it occurs before the other senses have a chance to take over the functioning of auditory cortex (e.g. Sandmann et al., 2012). Also, the traditional literature supporting a sensory-system based cortical architecture

cannot be ignored. Visual cortex is most readily modulated by visual input, for example. However, it could be that these traditional associations are related to a type of input providing the most robust information for a task (e.g. visual cortex for distance perception; auditory cortex for speech). If so, then task, rather than form of input, is considered the first function of a perceptual brain region. In fact, there is recent evidence that in some cases, the perceptual brain's functional organisation can develop independent of sensory input. Visual cortices of congenitally blind adults develop retinotopic organisations similar to those of sighted adults (Striem-Amit et al., 2015). This suggests a surprising degree of independence between the perceptual cortices and their presumed sensory organ input.

The supramodal perspective has some profound implications for our study and understanding of audition. In the remaining sections, we will address some of these implications. Throughout, the emphasis will be on behavioural research, although associated neurophysiological findings will be discussed. In the next section, we argue that a supramodal account implies an auditory function that has a capacity to provide perceptual support typically provided by vision. Conversely, the account also predicts that vision, and possibly touch, should be able to provide perceptual support typically provided by hearing. Next, we will argue that the supramodal account predicts that audition would integrate easily with the other senses, and that this integration would occur early in the perceptual stream. Finally, the account would predict that audition should be able to share its experience with the other senses, and vice versa. Thus, perceptual learning through audition should be transferable to visual tasks, and visual learning should be transferable to auditory task.

We will argue that two additional implications to the supramodal account make behavioural predictions. First, a supramodal mechanism could extract a form of information that is, itself, supramodal. The concept of modality-neutral information suggests that the relevant information in each modality takes a similar, higher-order form. The concept has been used to explain multisensory speech perception, and we will argue, is consistent with a supramodal architecture. Second, if, as argued, a supramodal mechanism extracts common perceptual primitives available in multiple modalities, then experience plays a different role in crossmodal perception. While experience can



certainly fine-tune both unimodal and crossmodal perception, associative experience between the senses is *not necessary* for either integration or crossmodal learning. As will be argued, this claim should have some implications for Bayesian approaches to cue integration.

## Crossmodal parity

### *Audition can act like vision*

For supramodal accounts, there is a primacy of task over type of sensory input in the perceptual brain's organisation. This could imply that the senses sharing a task-based mechanism should support some degree of equivalence, or *crossmodal parity*, in the relevant perceptual behaviour. This is not to imply that perceivers will be equally good at performing a perceptual task using any form of sensory input. Instead, behavioural evidence should be observed suggesting that the necessary *supportive information* for a task is available in the modalities relevant to the task-based mechanism.

Take, for example, a supramodal mechanism that registers surface properties by detecting either reflected light or reflected sound. The crossmodal parity assumption would predict some degree of behavioural equivalence in the relevant tasks, between sight and sound. In fact, there is evidence for a surprising degree of parity between echolocating and seeing surface properties. It has long been known that both blind and sighted individuals can identify an object's position, size, shape, and material surface using echolocation (for a review, see Kolarik, Cirstea, Pardhan, & Moore, 2014). Performance on these tasks can be impressive. Recent research shows that some expert echolocators can identify the position of an object within 1.2 degrees azimuth. This accuracy rivals the resolution for *visual* periphery position judgments (Teng, Puri, & Whitney, 2012). Another recent study showed that expert echolocators can identify object shape (from a choice of four) with an average of 77% accuracy, with some experts performing almost perfectly (Milne, Goodale, & Thaler, 2014). While these levels of accuracy were limited to echolocation experts, they do show that the informational support exists for some aspects of echolocation to have parity with vision.

Another perspective on evaluating cross-sensory parity would be to determine whether the involved modalities show *functional* equivalence in supporting

successful behaviour. The critical test would be to determine whether echolocation, like vision, can safely guide actions (e.g. Gibson, 1979; see also Stoffregen & Pittenger, 1995). In this sense, the classic echolocation task of walking to a wall and stopping just before contact (Milton & Dallenbach, 1950; for a review, see Kolarik, Cirstea, et al., 2014) may be the more relevant test of parity than judging absolute size and distance. Numerous studies have shown that both blind and blindfolded, sighted listeners can perform this walking and stopping task with as little as 10 minutes of training (Rosenblum, Gordon, & Jarquin, 2000; for a review, see Kolarik, Cirstea, et al., 2014).

Other behaviour-relevant research shows that the blind can use sound reflections to walk straight along the length of a wall (e.g. Ashmead et al., 1998; Strelow & Brabyn, 1982). Echolocators can also determine the location of obstacles vs. doorways so as to guide themselves successfully (Carlson-Smith & Wiener, 1996). More generally, it is well documented that echolocation can be the main means of navigation for some blind individuals, allowing them to hike and bike on mountain trails (e.g. Rosenblum, 2010). Other evidence suggests that blind individuals who actively use echolocation may be able to better navigate new environments, which may even lead to better employment opportunities (Thaler, 2013). Finally, it is well known that the re-coded echolocation information available in some sensory substitution devices can be used for successful navigation (for a review, see Kolarik, Timmis, Cirstea, & Pardhan, 2014).

In sum, while visual perception is most always more accurate than echolocation when judging object position, size, shape, and surface composition (e.g. Kolarik, Cirstea, et al., 2014), there are instances in which echolocation rivals vision (e.g. Teng et al., 2012). This suggests that echolocation contains the necessary information to support a supramodal mechanism. Also, when evaluated in terms of its functional potential to guide safe behaviours, echolocation contains the necessary features to make this possible. Arguably, it is this type of parity that is needed to make a task-based perceptual architecture useful.

Outside of echolocation, there is research showing how general aspects of spatial sound can guide successful behaviour (for reviews, see Carello, Wagman, & Turvey, 2005; Rosenblum, 2010). Our laboratory has found that blindfolded listeners are quite

good at judging whether a sound source is within arm's reach (Rosenblum, Wuestefeld, & Anderson, 1996). Impressively, listeners in this study displayed accuracy comparable to when perceivers judge the reachability of a *seen* object (e.g. Carello, Groszofsky, Reichel, Solomon, & Turvey, 1989). We have also tested whether the spatial sound created by acoustic obstructions can support behaviour-relevant judgments. For these studies, we created a "wall of sound" by attaching multiple loudspeakers to a wall and having them play recorded crowd babble. By positioning two large foam insulation boards in front of these loudspeakers, we found that listeners could accurately judge whether they could walk through the space formed between the boards (Gordon & Rosenblum, 2004). A follow-up study revealed that listeners could also successfully judge whether they could walk under a sound-obstructing board without ducking. These studies show that perceivers can use spatial aspects of sound to guide behaviour, and in some cases show accuracy rivaling visual judgments. In this sense, this research is supportive of the crossmodal functional parity that may be considered a requisite of a task-based, cross-sensory spatial mechanism.

Returning to echolocation, there are two other notable findings consistent with a task-based cortical organisation for space perception. First, there is evidence that some of the operations used for visual and echolocated object perception are the same. This evidence comes from research testing the size-weight illusion during echolocation. It is well known that for vision, the perceived weight of a lifted object depends on its seen size (Ross, 1969; and see Buckingham, 2014 for a review). Recent research shows that the illusion works in nearly identical ways for blind individuals such that an object's perceived weight is dependent on its *echolocated* size (Buckingham, Milne, Byrne, & Goodale, 2015). This could mean that the brain treats size information extracted from vision and echolocation in the same way.

Finally, it seems that even very minor visual decrements can induce a compensatory enhancement in auditory spatial perception. Findings have shown that individuals with myopia (near-sightedness) have subtly enhanced echolocation and sound location skills compared to individuals who have no myopia (Després, Candas, & Dufour, 2005; Dufour & Gérard, 2000). The authors of these studies suggest that because glasses, usually worn at an early age (before some individuals choose to

wear contact lenses) do not correct peripheral vision, some of the visual spatial input to the brain is degraded. The brain may then compensate by enhancing sensitivity to those spatial aspects of sound. While this study did not involve a neurophysiological examination of the finding, it could be that a task-based part of the visual cortex was enlisted for this auditory enhancement. Regardless, the ease with which crossmodal plasticity can occur seems supportive of a mechanism that can work well with visual or auditory input.

### ***Vision can act like audition***

Arguably, there are more demonstrations of crossmodal parity in the field of speech perception than any other research area. Multiple examples show that visual, and even tactile speech can provide information comparable to that provided auditorily. Again, crossmodal parity does not require that performance be equivalent across the modalities. Instead, parity will show that, because of the availability of supportive information, the involved modalities have the *potential* to be used similarly by the cortical mechanism to support the same function. In fact, there is evidence that some deaf lipreaders can understand speech as well as hearing individuals listening to speech (e.g. Andersson & Lidestam, 2005; Auer & Bernstein, 2007; Rosenblum, 2010). Also, both deaf and hearing individuals can be trained to perceive speech by *touching faces* and achieve understanding comparable to when listening to speech in low-level noise (Reed et al., 1985).

Reports of impressive lipreading abilities are consistent with research showing that much more information is available in visible speech—and is used by perceivers—than was originally thought. The seemingly invisible articulatory characteristics of interoral pressure and tongue-back position are available in visible cheek and jaw movements (Munhall & Vatikiotis-Bateson, 2004). Even the pitch distinctions signalling Mandarin tone are visible in cheek and head motions (Burnham, Ciocca, Lauw, Lau, & Stokes, 2000; Chen & Massaro, 2008; Smith & Burnham, 2012).

Next, recall that a superior evaluation of parity may involve determining the degree to which a *behaviourally relevant* task can be supported by the modalities. In fact, there is evidence that perceiving speech through either audition or vision can influence the details of one's own articulatory behaviour.<sup>2</sup> *Phonetic convergence* (also known as speech

alignment and speech accommodation) describes the tendency for perceivers to imitate the characteristics of speech spoken by other talkers. This inadvertent imitation not only occurs during live conversations (e.g. Pardo, 2006) but also when responding to recorded speech in a laboratory setting (e.g. Goldinger, 1998). A majority of the research conducted on the topic has tested how imitation is influenced by heard speech. However, research conducted in our own lab demonstrates that visual speech can also influence phonetic convergence. Miller, Sanchez, and Rosenblum (2010) found that perceivers who shadow (say aloud) speech they perceive spoken by pre-recorded talkers converge not only to the speech they hear, but also the speech they lipread. In fact, when phonetic convergence is assessed using the perceptual judgments of naïve raters (e.g. Goldinger, 1998), these raters judge utterances of shadowed lipread speech as sounding more like the *auditory* utterances of the shadowed talker. The raters are even able to make crossmodal judgments, judging shadowed utterances they *hear* as more similar to the *visible* (lipread) utterances of the shadowed talker (Miller et al., 2010). Later, we will discuss how audiovisually *integrated* speech can influence convergence. However, the fact that both modalities can induce convergence on their own provides evidence for behaviourally relevant crossmodal parity for auditory and visual speech perception. This makes it consistent with a supramodal perspective of the speech brain.

Other behavioural support for parity comes from evidence that the speech mechanism uses the same informational dimensions from the visual and auditory signals (for reviews, see Rosenblum, 2005, 2008). For example, both modalities make use of the abstract, time-varying aspects of their signals. For audition, it has long been known that speech can still be perceived after removal of the more small-scale characteristics of the signal such as noise-bursts, formants, and transitions. What is left over in these cases is a set of sine waves that follow the centre frequencies and amplitudes of the first three formants of an utterance as it unfolds over time (e.g. Remez, Rubin, Pisoni, & Carrell, 1981). Despite their reduced nature, these sine-wave utterances can be perceived as speech, and display the many idiosyncratic phenomena known to occur with normal speech input (Williams, Verbrugge, & Studdert-Kennedy, 1983). In fact, similar to how visual speech information can

improve comprehension of natural speech heard in difficult listening conditions (e.g. Sumbly & Pollack, 1954), sine-wave comprehension is improved when paired with the visible articulating face associated with the transformed utterance (Remez, Fellowes, Pisoni, Goh, & Rubin, 1998).

Work in our laboratory has used an analogous method to isolate time-varying *visual* speech information. In the point-light speech method, reflective dots are placed on the face, lips, teeth, and tongue-tip of a talker (e.g. Rosenblum, Johnson, & Saldaña, 1996; Rosenblum & Saldaña, 1996; see also Jaekl, Pesquita, Alsius, Munhall, & Soto-Faraco, 2015). The talker is then videotaped articulating speech under fluorescent light conditions. The videos are contrast-adjusted so that the resultant displays show only dots moving against a black background. Importantly, when the videos are paused so that no movement is seen, participants are unable to recognise the image as a face. This suggests the images do not contain any recognisable pictorial information. However, once the videos are played, the images are recognisable, and perceivers are able to lipread from these stimuli nearly as well as when seeing a regular face (Rosenblum, Johnson, et al., 1996). Also, as with auditory sine-wave speech stimuli, point-light stimuli show some of the same idiosyncratic phenomena known to occur with regular visual speech (e.g. Rosenblum & Saldaña, 1996). Additional research on both modalities suggests that the time-varying aspects of the signal (transitions in and out of vowels) can be *more* salient than the steady-state, “canonical” portions (Strange, Jenkins, & Johnson, 1983; Yakel, 2000). Thus, the speech mechanism seems to prioritise extraction of time-varying dimensions from both modalities. The fact that a common characteristic of information is used for both modalities is consistent with a supramodal mechanism for speech.

### **Modality-neutral forms of information**

Elsewhere we (and others) have argued that the salient form of speech information itself may be modality-neutral (amodal) (e.g. Rosenblum, 2004, 2005, 2008; see also Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009; Fowler, 2004; Summerfield, 1987). From this perspective, the speech articulators structure the acoustic and optic signals in a similar way such that the macroscopic properties of the signals take on the same general form. As an example, a reversal of an articulatory



act (mouth opened to closed to opened as in “aba”) is accompanied by a reversal in the pattern of acoustic intensity and spectral structure, as well as a reversal in the resultant optic structure (Summerfield, 1987). In this sense, both media are structured similarly over time, and in a way specific to the articulatory act. While the optic and acoustic speech signals are obviously comprised of different types of energy, articulation can impose a similar dynamic structure (e.g. pattern of *reversal*) on both types of energy media. Modality-neutral information then, can be considered as high-order patterns of energetic change whose abstract form allows it to exist in multiple energy media.

Support for modality-neutral information comes from research showing close correspondences between the optic and acoustic signals (e.g. Chandrasekaran et al., 2009; Munhall & Vatikiotis-Bateson, 1998; Yehia, Kuratate, & Vatikiotis-Bateson, 2002; Yehia, Rubin, & Vatikiotis-Bateson, 1998). For example, there is now strong evidence for correlations between the temporal characteristics of visible mouth opening area and formant frequencies in the acoustic signal (Chandrasekaran et al., 2009). These correlations are strong enough so that interpretable auditory speech can actually be synthesised from parameters extracted from the visible speech measurements (Yehia et al., 1998). It could very well be that these strong crossmodal correlations underlie a phenomenon known as the *unity effect for speech perception*: Observers seem to have an especially large perceptual tolerance for asynchronous (but matching) audiovisual speech—vs. nonspeech—signals (e.g. Vatakis, Ghazanfar, & Spence, 2008).

We have also considered modality-neutral information in the context of audiovisual time-to-arrival information (Gordon & Rosenblum, 2005; Rosenblum, 2004; and see Lee, van der Weel, Hitchcock, Matejowsky, & Pettigrew, 1992). There is evidence that for both auditory and visual stimuli, time to arrival judgments can be made based on a parameter defined by a relative rate of change in the respective energy media (Lee et al., 1992). For vision, the parameter can be instantiated in the changing rate of optical dilation formed by the approaching object. For audition, the parameter is potentially available in the changing rate of the signal’s intensity, spectral structure, or binaural structure. If the relevant time-to-arrival information does take on the same form in both modalities, we reasoned that the perceptual mechanism can

behave in a way that is agnostic about sensory modality (Gordon & Rosenblum, 2005). If so, then performance for time-to-arrival judgments should be comparable whether both modalities continuously convey the information, or the modalities alternate (every second) during the approach event. In this latter case, each modality is alternately interrupted, but continuous time-to-arrival information is always present. In a series of experiments, we found that judgments based on audiovisually alternating stimuli were as accurate as judgments based on continuous audiovisual stimuli (see also Cappe, Thelen, Romei, Thut, & Murray, 2012; Cappe, Thut, Romei, & Murray, 2009). Potentially, the perceptual mechanism can make use of modality-neutral time-to-arrival information regardless of the alternating nature of the modality through which it is presented.

Clearly, an informational form that is modality-neutral is consistent with a task-based, supramodal perceptual mechanism. If such a mechanism were able to extract the same high-order pattern from different energy media, it would allow for the parity, primacy of integration, and crossmodal learning that, we argue, are hallmark properties of the architecture.

## Cross-sensory integration

### Speech events

Not only does a supramodal architecture imply a functional parity between the senses, it also implies an especially early and automatic process of integration. Supramodal architecture suggests that multisensory activations can occur in primary sensory cortices, or earlier. This would mean that the involved senses would likely have the chance to converge early and completely. This fact should be observable in the behavioural data.

In fact, there is evidence for early and widespread crossmodal integration of speech perception. Under normal speech conditions, the presence of visual speech can both enhance and speed the recognition of auditory speech. Most obviously, visual speech can help compensate for auditory speech heard in noisy environments (e.g. Bernstein, Auer, & Takayanagi, 2004; Sumby & Pollack, 1954). Visual speech can also help with the understanding of speech spoken with a heavy accent, or that contains particularly dense content (Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987). It is known that visual

speech also facilitates first and second language acquisition (e.g. Arnold & Hill, 2001; Navarra & Soto-Faraco, 2007; Reisberg et al., 1987; Teinonen, Aslin, Alku, & Csibra, 2008). This fact is evidenced by blind children who take longer than their sighted counterparts to acquire phonetic distinctions (e.g. /m/ vs. /n/) that are hard to hear but easy to see (Mills, 1987). These findings, along with the neurophysiological results discussed above, indicate that visual speech facilitates auditory speech perception in normal circumstances.

However, some of the most compelling examples of multimodal speech influences are laboratory demonstrations using *incongruent* crossmodal information (for reviews, see Dias, Cook, & Rosenblum, *in press*; Tiippana, 2014). The well-known McGurk effect demonstrates how visible speech articulation, inconsistent with concurrent auditory speech, can influence reports of “heard speech” (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). For example, when presenting an auditory utterance of /ba/ with a concurrent visual articulation of /ga/, perceivers will identify the utterance they “hear” as /da/. This visual influence on auditory speech perception can be observed even when the visual signal is reduced to isolated kinematic information (Rosenblum & Saldaña, 1996).

The McGurk effect demonstrates the profound breadth, resiliency, and automaticity of crossmodal integration. Many different consonant combinations (e.g. MacDonald & McGurk, 1978; McGurk & MacDonald, 1976) and, more subtly, vowel combinations (e.g. Massaro & Cohen, 1993; Summerfield & McGrath, 1984) have been shown to integrate. The integration of auditory and visual speech information occurs across multiple languages (e.g. Fuster-Duran, 1996; Massaro, Cohen, Gesi, & Heredia, 1993; Sams, Manninen, Surakka, Helin, & Kättö, 1998; Sekiyama & Tohkura, 1991, 1993). Perceivers demonstrate the McGurk effect at an early age (McGurk & MacDonald, 1976), even during infancy (e.g. Burnham & Dodd, 2004; Desjardins & Werker, 2004; Rosenblum, Schmuckler, & Johnson, 1997).<sup>3</sup> Auditory and visual speech information integrates even when the sensory sources are noticeably displaced spatially (Bertelson, Vroomen, Wiegand, & de Gelder, 1994) or temporally (Munhall, Gribble, Sacco, & Ward, 1996; Soto-Faraco & Alsius, 2009). In fact, even if perceivers are made aware of the discrepant nature of the auditory and visual components of an audiovisual stimulus, they will still report integrated percepts (Bertelson & De Gelder, 2004; Soto-

Faraco & Alsius, 2009). This could suggest that the supramodal nature of the speech mechanism prohibits access to the unimodal information once that information is integrated (but see Gentilucci & Cattaneo, 2005; Ostrand, Blumstein, & Morgan, 2011).

Elsewhere, we have described the behavioural evidence that speech integration occurs early in the perception and language process (Rosenblum, 2005; and see Campbell, 2011; Green, 1998; but see Bernstein, Auer, & Moore, 2004). There is evidence that integration occurs before speech features are identified (for a review, see Green, 1998). Also, there are examples for which the auditory and visual streams, on their own, are nearly undecipherable, but when together are perceived as coherent speech (e.g. Rosen, Fourcin, & Moore, 1981). This finding would suggest that at least some integration must occur before phonetic analysis (e.g. Pardo & Remez, 2006).

The behavioural evidence for early integration is consistent with the evidence (above) that auditory cortex and brainstem respond to visual speech (Calvert et al., 1997; Musacchia et al., 2006). There is also a substantial literature concerning multisensory interactions in the auditory cortex as measured by event related potential (ERP) components N1 and P2 (see Besle, Bertrand, & Giard, 2009, for a review). The general finding is that the amplitude and timing of the N1 and P2 components is attenuated for audiovisual stimulus presentations relative to audio-alone (e.g. Besle, Fort, Delpuech, & Giard, 2004; Stekelenburg & Vroomen, 2007, 2012; van Wassenhove, Grant, & Poeppel, 2005). These effects co-occur with behavioural measures of crossmodal integration (Besle et al., 2004). Interestingly, these effects depend on a listener’s interpretation of the audiovisual stimuli. Using sine-wave speech, a recent study found attenuation in P2 for incongruent audiovisual displays, if participants had been primed to hear sine-wave speech as speech (Baart, Stekelenburg, & Vroomen, 2014). In contrast, participants who did not hear sine-wave speech as speech showed no effect of audiovisual congruency on P2. This might suggest that N1 responds to general properties of audiovisual stimuli, while P2 responds to perceptual information at the level of phonetics. Regardless, these ERP findings support the notion that integration occurs early in the perceptual and linguistic process.

A supramodal speech mechanism also seems sensitive to haptic and kinaesthetic input. Perceiving speech by either means is known to influence

auditory speech to produce McGurk-type effects (Fowler & Dekle, 1991; Gick & Derrick, 2009; Ito, Tiede, & Ostry, 2009; Sams, Mottonen, & Sihvonen, 2005). For example, touching an articulating face as described above can influence heard speech (Fowler & Dekle, 1991). A tactile sensation of an air puff on the back of the neck (Gick & Derrick, 2009), and even to the ankle (Derrick & Gick, 2013), can provide articulatory aspiration information that can make an auditory utterance of “ba”, sound like a “pa”. Next, one’s own articulatory information provided by voluntary (Sams et al., 2005) and involuntary (Ito et al., 2009) mouth movements can influence auditory speech judgments. Importantly, all of the participants in these experiments had normal hearing and vision and had little, if any experience perceiving speech information from tactile sources. Still, these novel sources were able to integrate with and influence auditory speech. As mentioned above, the fact that cross-sensory effects can occur *without* prior association between the sensory components is fully consistent with a supramodal mechanism. The mechanism can, in principle, be reactive to the relevant (modality-neutral) information in any modality regardless of experience with the information conveyed in the modality.

Cross-sensory effects without prior association have additional implications. The Bayesian approach to cue integration has been very influential in recent years (Knill & Pouget, 2004; Ma, Beck, Latham, & Pouget, 2006). This approach suggests that the brain can optimally integrate multiple cues (along with prior expectations) based on probabilities derived from associative experience between the cues. The merits of the Bayesian approach to multi-sensory phenomena are not to be underestimated. The Bayesian approach has been applied to multi-sensory speech (e.g. Altieri, Pisoni, & Townsend, 2011; Massaro, 1998) and nonspeech (Battaglia, Jacobs, & Aslin, 2003; Deneve & Pouget, 2004; Rohe & Noppeney, 2015; Shams, 2011) perception and has been used to explain a number of the phenomenon discussed throughout this report (e.g. McGurk effect; visual enhancement of speech in noise; flash illusion).

However, a challenge for the Bayesian approach is to account for crossmodal influences and transfer of experience for which no practical opportunities are available for perceivers to learn the probabilistic relationships between separate sensory events. Take, for example, how an aero-tactile stimulus to

the ankle can cause perceivers to identify an acoustic utterance of /ba/ as “pa” (Derrick & Gick, 2013). Certainly, the typical adult perceiver has not spent meaningful amounts of time in conversation with others while speaking into each other’s ankles. Future work on the Bayesian approach will need to account for these demonstrations.

There is also evidence for influences of integration on behaviourally relevant responses. Recall the phenomenon of speech convergence described earlier. Convergence shows that the talker-specific information available in either the auditory or visual modality can influence a listener’s production responses to sound more like that of the talker. In a series of studies conducted in our lab, we found that perceivers converge to the *integrated* audiovisual speech they perceive. If perceivers are presented an audiovisual syllable with incongruent time parameters between the modalities, their production responses will be a compromise between those timings (Sanchez, Miller, & Rosenblum, 2010). We also observed that both signals influence production responses such that seeing a talker can increase the amount of convergence to heard speech (Dias & Rosenblum, 2011). In a follow-up test, we determined that this increase is dependent on seeing the articulating mouth (Dias & Rosenblum, 2015a). This suggests that crossmodal speech integration can modulate behaviourally relevant responses, thereby satisfying that aspect of supramodal architecture.

### Nonspeech events

A good portion of the literature on nonspeech cross-modal integration concerns the influences of vision and audition on *spatial* perception. This topic is relevant to the proposed supramodal spatial mechanism discussed above. The well-known *ventriloquism effect* is demonstrated when perceivers identify the location of a sound as shifted so that it seems closer to a displaced visual stimulus (e.g. Choe, Welch, Gilford, & Juola, 1975; Frissen, Vroomen, de Gelder, & Bertelson, 2003; Vroomen & de Gelder, 2004). Conversely, auditory information can shift the perceived location of visual stimuli, albeit often to a weaker degree (e.g. Arnott & Goodale, 2006; Teramoto et al., 2010). Importantly, for the original effect, the influencing visual information seems to have a direct influence on the auditory cortex hinting at a supramodal architecture (e.g. Bonath et al., 2007; for a review, see Chen & Vroomen, 2013).

Similar crossmodal effects have been observed for perceived motion. Freeman and Driver (2008) found that the apparent motion induced by the ordered visual presentation of three boxes could be influenced by either the inter-stimulus-interval between the three box-presentations, or by the timing of an accompanying sound (beep). Relatedly, acoustic information for depth can induce perception of looming motion in static visual stimuli (Sutherland, Thut, & Romei, 2014).

Recall that much of the recent neurophysiological support for a spatial supramodal mechanism concerns echolocation. We are unaware of behavioural examples showing crossmodal integration of echolocation in humans. However, there is relevant evidence for visual influences on echolocation in bats. Often echolocation and vision are portrayed as functionally separate, but complimentary systems in bats (e.g. Boonman, Bar-On, Cvikel, & Yovel, 2013). However, there is evidence that when Little Brown bats are presented with extraneous visible objects, their echolocation accuracy is reduced, often leading to collisions (e.g. Orbach & Fenton, 2010). Interestingly, this reduced accuracy only occurred when bats were presented with extraneous visible objects similar to those they were echolocating. Irrelevant visual stimuli such as overall changes in illumination did not have this effect. Relatedly, the simple distress calls of other bats also did not reduce accuracy. These facts led the researchers to conclude that the presence of visual objects does not act as a simple distraction, but actually integrates with the echolocated signal to influence performance (Orbach & Fenton, 2010). There is also evidence that Big Brown bats integrate echolocation, visual, and vestibular senses in orienting their heads during prey capture (Horowitz, Cheney, & Simmons, 2004).

There is relevant human research on an area closely related to echolocation. Typically, the term echolocation refers to *active* instances in which the listener produces their own emitted sound (e.g. bat chirps; human tongue clicks or footsteps). In fact, a majority of the neurophysiological research on the topic is based on active echolocation. However, it is also known that humans and other animals can use *passive* echolocation: instances in which perceived reflections are created from a sound source external to the perceiver (e.g. Ashmead et al., 1998). Passive echolocation (or use of reflected/reverberant structure) is what enables us to hear the size of the room we occupy, as well as its surface materials (e.g. a gym locker room vs.

walk-in closet). In these cases, the salient reflected sound could be created from sound sources such as the talking of other people or even a room's ventilation system. It is this type of reflected structure that is simulated when sound engineers add reverb, echo, and delay to make a sound, a music performance, or a movie scene location seem, say, larger. It is unclear whether the same perceptual mechanism is used for active and passive echolocation. However, it is noteworthy that some of the original echolocation research conducted in the 1950s used external sound sources such as white noise and clicks produced distally from the listener (for a review, see Kolarik, Cirstea, et al., 2014).

Human passive echolocation is often studied in the context of concert hall and virtual reality research. In these studies, some source (or simulated source) emits a sound and perceivers are asked to judge properties of the space (e.g. concert hall) based on how it reflects the emanating sound. Research has shown that audiovisual integration can occur in these contexts. For example, the visible size of a concert space can influence auditory distance judgments (Maempel & Jentsch, 2013) as well as more expert auditory judgments such as "source width" and "listener envelopment" (Valente & Braasch, 2010). In virtual reality systems, judgments of auditory source distance and source width can be influenced by the quality of the visual simulation (Larsson, Västfjäll, & Kleiner, 2001). Conversely, the perceived quality of the visual simulation of a room is strongly influenced by the quality of the accompanying auditory simulation (for a review, see Hulusic et al., 2012). Finally, while not integration as such, there is evidence that observers can use a room's sound reflections to match to photographs of those spaces (e.g. Robart & Rosenblum, 2005; see also Scharine, Letowski, Mermagen, & Henry, 2010).

Beyond spatial perception, there are examples of audiovisual *event* integration. We have shown that seeing a cello being plucked or bowed can influence auditory judgments of plucked vs. bowed notes (Saldaña & Rosenblum, 1993). Relatedly, auditory judgments of a marimba note's duration can be influenced by a video of a player "sustaining" a marimba note (Schutz & Lipscomb, 2007). Similar cross-sensory influences in music include singer facial expression and body movements influencing "heard" song emotion (e.g. Thompson, Russo, & Quinto, 2008; Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011; Vuoskoski, Thompson, Clarke, &

Spence, 2014). Visible facial emotion can also influence heard pitch relations (Thompson, Russo, & Livingstone, 2010).

Possibly the most studied nonspeech case of audiovisual event integration is the *sound-induced flash illusion* (Shams, Kamitani, & Shimojo, 2000, 2002). In this effect, auditory beeps can modulate perception of visual flash stimuli so that the number of reported flashes depends on the number of beeps heard. Typically, a single flash is made to look like a quick double flash when accompanied by two quick beeps. The reverse can also occur so that the number of heard beeps is influenced by the number of flashes presented. The original illusion is very robust, working with different types of flashes and sounds. Like the McGurk effect, the illusion is also resistant to outside knowledge and even feedback training (for a review, see Shams & Kim, 2010). This perceptual robustness is consistent with neurophysiological findings on the effect. Measures of ERP in primary visual cortex show that an illusory, sound-induced flash creates similar activity to when an actual flash is presented (Shams, Kamitani, Thompson, & Shimojo, 2001). Similar results have been observed in primary visual cortex using magnetoencephalography and functional magnetic resonance imaging measures (Shams, Iwaki, Chawla, & Bhattacharya, 2005; Watkins, Shams, Tanaka, Haynes, & Rees, 2006). This crossmodal cortical activation is consistent with a supramodal architecture.

In sum, the ubiquity and primacy of crossmodal integration is supportive of a supramodal architecture. These crossmodal effects seem especially prominent in functions that have been shown to have neurophysiological support for a supramodal mechanism (e.g. speech and spatial perception). In the next section, we will present research supporting another behavioural outcome of this architecture: crossmodal transfer of experience.

### Crossmodal transfer of experience

Another prediction of a supramodal architecture is that the senses involved in a mechanism should be able to share their experience. If a mechanism is working with a modality-neutral form of information, then learning to better use this information should benefit any sense in which the information is available. In other words, the benefits of perceptual experience should be transferable between the involved modalities.

In fact, there is evidence that the speech experience gained by audition can be shared with vision and vice versa (Rosenblum, Miller, & Sanchez, 2007; Sanchez, Dias, & Rosenblum, 2013). It is known that for both modalities, gaining experience with a specific talker can improve recognition of that talker's speech (for reviews, see Nygaard, 2005; Rosenblum, 2005). Perceivers are better at understanding the speech-in-noise of familiar talkers and are better at lipreading from talkers with whom they have experience (see also Lander & Davies, 2008). These *talker familiarity effects* are thought to be based on perceivers gaining experience with a talker's idiosyncratic manner of speaking—or *idiolect*. Importantly, idiolect is a property of articulatory style and not tied to any modality: It is an *amodal* parameter available in multiple modalities. In fact the idiolectic information available in auditory and visual speech can actually allow perceivers to match a talker's face to their voice and vice versa (e.g. Lachs & Pisoni, 2004a, 2004b, 2004c; Rosenblum, Smith, Nichols, Hale, & Lee, 2006).

Thus, there is *amodal* talker idiolect information available in both modalities that, when learned, can facilitate speech perception in both modalities. It stands to reason then that the talker-experience gained in one modality should transfer to help speech perception in the other. For example, lipreading a specific talker for some period of time should facilitate *auditory* perception of that talker's speech and vice versa. We tested this prediction in a series of projects. In a first project, we gave participants one hour of experience lipreading simple sentences from a single speaker (Rosenblum et al., 2007). Participants were asked to recognise as many words as they could from the sentences, and were given feedback after each response. After the lipreading task, participants were given an unexpected auditory speech-in-noise test. For this task, participants listened to a new set of sentences, and identified as many words as they could. Importantly, for half of the participants, the auditory sentences were produced by the same speaker they had just lipread. The other half of the participants heard speech from a talker different from the one they had lipread. Results showed that the participants who listened to the speech of the speaker they had previously lipread, performed better on the speech-in-noise test. The experience participants gained from lipreading a talker was transferred crossmodally, allowing them to better understand that talker's speech auditorily.



In a second study, we reversed the modalities so that participants became familiar with a talker auditorily, and were then tested on their lipreading of that talker, as well as a new talker (Sanchez et al., 2013). Results showed that participants were better at lipreading from the talker they had just heard. This study also showed that the same-talker advantage occurred regardless of whether the same words, or new words were used between the auditory and lipreading tasks. This result suggests that some of the idiolectic information transferred across modalities is not tied to particular words. Instead, the relevant information is likely maintained at a more fundamental gestural level. Together, these studies suggest that perceptual experience with a speaker is not simply relevant to the modality through which it is originally acquired. Instead, the experience can be shared across modalities allowing us to see or hear familiar talkers more easily. The relevant speech mechanism seems designed around the *task* of familiar talker perception rather than around a specific sensory modality.

Other examples of crossmodal sharing of perceptual experience exist outside the speech domain. Many of these examples involve perceptual aftereffects. Aftereffects typically involve repeated stimulation of one type (e.g. downward visual motion) inducing adaptation, so that subsequent perception becomes more sensitive to—or even spuriously experienced as—the *opposite* type of stimulation (e.g. upward motion). In crossmodal adaptation effects, adaptation to one sense induces aftereffects in another. For example, following adaptation to visual horizontal motion in one direction, perceivers spuriously identify stationary sounds as moving in the opposite direction (Ehrenstein & Reinhardt-Rutland, 1996). Similarly, adaptation to visibly approaching images induces opposite changes in auditory perception (erroneous perception of decreasing loudness) (Kitagawa & Ichihara, 2002). In another example, listening to a rapid auditory signal can make the perceived rate of a comparably paced visual signal seem slower and vice versa (Levitan, Ban, Stiles, & Shimojo, 2015).

Importantly, in all of these examples, crossmodal aftereffects occurred despite the auditory and visual components never being presented together. This is also true of our experiments showing crossmodal transfer of talker information: perceivers never experienced a combined audiovisual presentation of the talkers. The fact that bimodal associative experience is unnecessary for crossmodal learning

transfer is noteworthy. This suggests that a supramodal mechanism can improve its extraction of (potentially) modality-neutral information from a single modality, such that this improvement can be later applied to extraction of that information from other modalities. No bimodal experience is required for the modalities to share their experience. As intimated above, these findings would seem to pose a challenge to purely Bayesian accounts of multisensory perception. This being said, there are certainly instances for which having bimodal experience can improve crossmodal sharing of experience (for a review, see Shams, Wozny, Kim, & Seitz, 2011). However, a supramodal mechanism, which can use modality-neutral information, allows crossmodal sharing to occur without *requiring* bimodal experience.

Regarding echolocation, we are unaware of human research showing crossmodal transfer of learning. However, there are relevant findings on dolphins. Dolphins show an impressive ability to match a novel object they have seen to that object when echolocating, and vice versa (Pack & Herman, 1995; Pack, Herman, Hoffmann-Kuhnt, & Branstetter, 2002). Dolphins are able to make these matches with 95% accuracy and can even indicate when the correct match is omitted from the test set (Pack et al., 2002). There is also evidence that the learning dolphins obtain through vision can be transferred to improve echolocation performance (Kuczaj, Solangi, Hoffland, & Romagnoli, 2008). In one study, dolphins were taught to visually discriminate five novel hand-arm gestures performed by a trainer. The gestures communicated which trick the dolphins should perform. After visual training, the trainer performed the gestures behind a screen that blocked the dolphins' vision but allowed echolocation. The dolphins were able to use echolocation to successfully discriminate the gestures to perform the correct tricks. The dolphins transferred their visual learning to their echolocation performance.

There is also a sizable literature on selective adaptation of crossmodal speech (for a review, see Samuel & Lieblich, 2014). Interestingly, these studies are typically interpreted as failing to show significant crossmodal adaptation. In one example, a silent video of a talker articulating "ba" was repeatedly shown to perceivers (Roberts & Summerfield, 1981). Perceivers were then asked to identify a continuum series of nine *auditory* syllables that ranged from a clear "ba" to a clear "da" with more ambiguous "ba/da" syllables in between. Past research had shown that

hearing a repeated audio “ba” would make more syllables in the continuum sound like “da”: a *speech selective adaptation effect*. However, the visual “ba” failed to induce the same adaptation. In addition, a repeated *audiovisual* syllable comprised of an audio “ba” and visual “ga” (mostly perceived as “da”) induced adaptation based solely on the audio “ba” component. This and other studies (e.g. Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Liebllich, 2014; and for a review, see Dias, Cook, & Rosenblum, 2016), have been interpreted as evidence that selective adaptation in speech depends on the unisensory information in the signal, and not the bimodal, phonetic information. This interpretation is clearly antithetical to a supramodal account of speech perception.

However, recent research conducted in our lab suggests that crossmodal selective adaptation of speech can occur. We have recently conducted a meta-analysis incorporating the results of Roberts and Summerfield (1981) with data from our own tests (Dias & Rosenblum, 2015b). This meta-analysis suggests that though crossmodal adaptation effects are small, they are consistent across multiple studies. Based on this finding, we conducted a follow-up experiment using larger samples sizes. This experiment revealed that adaptation to phonetic information *can* induce significant crossmodal aftereffects. Adaptation of visual speech syllables can change categorisation of auditory speech and vice versa (Dias & Rosenblum, 2015b).

Clearly, more research is required to flesh-out crossmodal speech adaptation and why it may be so much smaller than unimodal adaptation. However, it should be noted that other examples of crossmodal speech adaptation do exist—involving kinaesthetic, rather than visual, information. There is evidence that silently mouthing a syllable (“ba”) to one’s self can change the way a continuum of related syllables (“ba” to “da”) is perceived (Cooper, 1975; Cooper, Billings, & Cole, 1976). Conversely, hearing a repeated syllable can affect how one then articulates a related syllable (i.e. in a direction away from the heard syllable) (Cooper & Lauritsen, 1974). Together with our recent meta-analysis and large participant-number experiment, these results suggest that selective adaptation can modulate speech perception across sensory modalities, providing another demonstration of how the modalities can share their perceptual experience.<sup>4</sup>

Finally, there are crossmodal adaptation effects that do not involve audition. For example,

adaptation to visual motion in one vertical direction transfers to perceived tactile motion in the opposite direction, and vice versa (Konkle, Wang, Hayward, & Moore, 2009). Also, visually adapting to a smiling face can make a touched face seem less happy and vice versa (Matsumiya, 2013). With regard to this latter example, there is other research supportive of a supramodal task-based mechanism for face perception (for a review, see Matsumiya, 2013).

In sum, there is mounting evidence that the senses can share their experience. Perceptual learning can be transferred across sensory systems, and selective adaptation of one sense can influence another. This sharing of experience seems especially prevalent in sensory functions and tasks for which a supramodal mechanism seems viable. The fact that crossmodal sharing of experience can occur without prior bimodal exposure to that experience (a possible problem for Bayesian accounts) could be based on the workings of a supramodal mechanism that makes use of information that takes a modality-neutral form.

## Conclusions

Certainly, further neurophysiological and behavioural research is needed to test the viability of a supramodal architecture as it pertains to audition. For example, much of the supportive neurophysiological research has tested individuals with sensory impairments. This strategy can reveal the cortical potential resulting from compensatory-based perceptual expertise. However, more research needs to examine the viability of supramodal architecture in sensory-intact individuals. This research can test for the architecture under conditions of normal perception, as well as when the senses are *temporarily* deprived or distorted. Supramodal organisation should be a general strategy, and not simply reserved for just sensory loss. For instance, much of the neurophysiological support for crossmodal cortical activation for echolocation has been based on imaging of early blind individuals. However, it could very well be that this crossmodal activation will occur comparably in sighted individuals given substantial practice in echolocating, or when visually deprived for a few hours. Further neurophysiological work on sensory-intact individuals will help address these possibilities.

Another area for future research is testing *behaviourally relevant* auditory perception. Arguably, the most sensible way to evaluate crossmodal parity is

to determine if the involved senses can guide successful behaviour in a comparable manner. Unfortunately, relatively little auditory research has been designed to test behaviourally relevant performance. Likewise, there has been little research designed to directly compare behaviourally relevant performance across modalities. Future research along these lines would address the crossmodal parity implication of the supramodal architecture.

Next, while substantial research has been conducted on crossmodal integration of speech events, less has been conducted on integration of nonspeech events. Arguably, one reason that multisensory speech has been such a fruitful research area is that its core phenomenon—The McGurk Effect—is based on integration of complex information sources for a real-world distal event. There is currently no analogous effect for nonspeech perception. While the multisensory research on the ventriloquist and sound-induced flash illusions has been invaluable, establishing a *true* nonspeech analogue to a McGurk effect could allow for a more detailed understanding of event integration.

Other research should explore how auditory speech integrates with *tactile* speech perception. While certainly not an everyday skill for most of us, auditory-tactile integration provides an interesting tool with which to study the issue of how prior bimodal experience, or lack thereof, bears on multisensory integration. Relatedly, understanding how the senses can share their experience should be an important area of future research. Multisensory learning has become an important research area. Examining whether this sharing can occur without bimodal experience, as the supramodal architecture implies, could add critical information to this literature. Conversely, determining how bimodal experience can enhance this learning is also an interesting question.

In sum, the new conception of a supramodal brain architecture poses important questions about audition, the other senses, and their integration. Consequently, our understanding of how audition is used by the brain to support behaviour may need to broaden.

## Notes

1. Here the term *task* is given its most functional definition. The term refers to the recovery of distal event properties (spatial layout of objects and surfaces; articulatory acts of a speaker) that can be

used to guide successful behaviour (navigating an environment; guiding a speech production response). No other assumptions about sub-components of a task are made.

2. There is vast literature on the link between speech perception and production which has inspired at least two theories of speech perception (e.g. Fowler, 1986, 2004; Galantucci, Fowler, & Turvey, 2006; Liberman & Mattingly, 1985). While some of this literature is relevant to multisensory perception – and is cited here (e.g. Dias & Rosenblum, 2015a; Sanchez et al., 2010) – readers can refer to these other sources for non-related findings on perception-production links.
3. Developmental research shows that there are changes in the strength of the McGurk effect throughout childhood (Dupont, Aubin, & Menard, 2005; McGurk & MacDonald, 1976), with a notable decrease when children are learning to read (Jerger, Damian, Spence, Tye-Murray, & Abdi, 2009).
4. It should be noted that besides instances for which unimodal speech experience can influence speech perception in another modality, there are many instances for which bimodal speech experience can later influence perception in one of the involved modalities. For example, it is well known that adding a visual speech component to train perceivers to better hear degraded auditory speech improves later perception of degraded auditory speech alone (e.g. Eberhardt, Auer, & Bernstein, 2014). Relatedly, adding visual speech to an ambiguous auditory speech syllable will act to *recalibrate* the auditory syllable so that it is unambiguously interpreted as that indicated by the visual stimulus (for a review, see Vroomen & Baart, 2012). For example, an ambiguous auditory token as likely to be heard as a “ba” or “da”, when dubbed with a video of a face clearly articulating “ba”, will lead the auditory syllable to later be heard as an unambiguous “ba”. While these examples certainly show how bimodal experience can fine-tune perception of unimodal speech, they do not serve to demonstrate how experience can be shared across individual modalities—which is the point of this section of the paper.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Abboud, S., Maidenbaum, S., Dehaene, S., & Amedi, A. (2015). A number-form area in the blind. *Nature Communications*, 6. doi:10.1038/ncomms7026
- Allman, B. L., Keniston, L. P., & Meredith, M. A. (2009). Adult deafness induces somatosensory conversion of ferret auditory cortex. *Proceedings of the National Academy*

- of *Sciences*, 106(14), 5925–5930. doi:10.1073/pnas.0809483106
- Altieri, N., Pisoni, D. B., & Townsend, J. T. (2011). Some behavioral and neurobiological constraints on theories of audiovisual speech integration: A review and suggestions for new directions. *Seeing and Perceiving*, 24(6), 513–539. doi:10.1163/187847611X595864
- Andersson, U., & Lidestam, B. (2005). Bottom-Up driven speechreading in a speechreading expert: The case of AA (JK023). *Ear and Hearing*, 26(2), 214–224. doi:10.1097/00003446-200504000-00008
- Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, 92(2), 339–355. doi:10.1348/000712601162220
- Arnott, S. R., & Goodale, M. A. (2006). Distorting visual space with sound. *Vision Research*, 46(10), 1553–1558. doi:10.1016/j.visres.2005.11.020
- Arnott, S. R., Thaler, L., Milne, J. L., Kish, D., & Goodale, M. A. (2013). Shape-specific activation of occipital cortex in an early blind echolocation expert. *Neuropsychologia*, 51(5), 938–949. doi:10.1016/j.neuropsychologia.2013.01.024
- Ashmead, D. H., Wall, R. S., Eaton, S. B., Ebinger, K. A., Snook-Hill, M. M., Guth, D. A., & Yang, X. (1998). Echolocation reconsidered: Using spatial variations in the ambient sound field to guide locomotion. *Journal of Visual Impairment & Blindness*, 92(9), 615–632. <http://www.afb.org/jvib/jvibabstractNew.asp?articleid=jvib920905>
- Auer, E. T., & Bernstein, L. E. (2007). Enhanced visual speech perception in individuals with early-onset hearing impairment. *Journal of Speech, Language, and Hearing Research*, 50(5), 1157–1165. doi:10.1044/1092-4388(2007)080
- Auer, E. T., Bernstein, L. E., Sungkarat, W., & Singh, M. (2007). Vibrotactile activation of the auditory cortices in deaf versus hearing adults. *NeuroReport*, 18(7), 645–648. doi:10.1097/WNR.0b013e3280d943b9
- Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia*, 53, 115–121. doi:10.1016/j.neuropsychologia.2013.11.011
- 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*, 20(7), 1391–1397. doi:10.1364/JOSAA.20.001391
- Bavelier, D., Dye, M. W. G., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in Cognitive Sciences*, 10(11), 512–518. doi:10.1016/j.tics.2006.09.006
- Bernstein, L. E., Auer, E. T. J., & Moore, J. K. (2004). Audiovisual speech binding: Convergence or association. In G. A. Calvert (Ed.), *The handbook of multisensory processes* (pp. 203–223). Cambridge, MA: MIT Press.
- Bernstein, L. E., Auer Jr, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lip-reading. *Speech Communication*, 44(1–4), 5–18. doi:10.1016/j.specom.2004.10.011
- Bertelson, P., & De Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141–179). Oxford: Oxford University Press.
- Bertelson, P., Vroomen, J., Wiegand, G., & de Gelder, B. (1994). Exploring the relation between McGurk interference and ventriloquism. In *International conference on spoken language processing* (Vol. 2, pp. 559–562). Yokohama: Acoustical Society of Japan.
- Besle, J., Bertrand, O., & Giard, M.-H. (2009). Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hearing Research*, 258(1–2), 143–151. doi:10.1016/j.heares.2009.06.016
- Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20(8), 2225–2234. doi:10.1111/j.1460-9568.2004.03670.x
- Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., & King, A. J. (2007). Physiological and anatomical evidence for multi-sensory interactions in auditory cortex. *Cerebral Cortex*, 17(9), 2172–2189. doi:10.1093/cercor/bhl128
- Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology*, 17(19), 1697–1703. doi:10.1016/j.cub.2007.08.050
- Boonman, A., Bar-On, Y., Cvikel, N., & Yovel, Y. (2013). It's not black or white – on the range of vision and echolocation in echolocating bats. *Frontiers in Physiology*, 4, 248. doi:10.3389/fphys.2013.00248
- Bottari, D., Heimler, B., Caclin, A., Dalmolin, A., Giard, M.-H., & Pavani, F. (2014). Visual change detection recruits auditory cortices in early deafness. *NeuroImage*, 94, 172–184. doi:10.1016/j.neuroimage.2014.02.031
- Buckingham, G. (2014). Getting a grip on heaviness perception: A review of weight illusions and their probable causes. *Experimental Brain Research*, 232(6), 1623–1629. doi:10.1007/s00221-014-3926-9
- Buckingham, G., Milne, J. L., Byrne, C. M., & Goodale, M. A. (2015). The size-weight illusion induced through human echolocation. *Psychological Science*, 26(2), 237–242. doi:10.1177/0956797614561267
- Burnham, D. K., Ciocca, V., Lauw, C., Lau, S., & Stokes, S. (2000). Perception of visual information for Cantonese tones. *Proceedings of the Eighth Australian International Conference on Speech Science and Technology*, pp. 86–91.
- Burnham, D. K., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 45(4), 204–220. doi:10.1002/dev.20032
- Butler, B. E., & Lomber, S. G. (2013). Functional and structural changes throughout the auditory system following congenital and early-onset deafness: Implications for hearing restoration. *Frontiers in Systems Neuroscience*, 7, Article 92. doi:10.3389/fnsys.2013.00092
- Callan, D. E., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG case study. *Cognitive Brain Research*, 10, 349–353. doi:10.1016/S0926-6410(00)00054-9
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., ... David, A. S. (1997).



- Activation of auditory cortex during silent lipreading. *Science*, 276, 593–596. doi:10.1126/science.276.5312.593
- Campbell, C. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 1001–1010. doi:10.1098/rstb.2007.2155
- Campbell, R. (2011). Speechreading: What's MISS-ing? In A. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.), *Oxford handbook of face perception* (pp. 605–630). Oxford: Oxford University Press.
- Cappe, C., Thelen, A., Romei, V., Thut, G., & Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration. *Journal of Neuroscience*, 32(4), 1171–1182. doi:10.1523/JNEUROSCI.5517-11.2012
- Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia*, 47(4), 1045–1052. doi:10.1016/j.neuropsychologia.2008.11.003
- Carello, C., Groszofsky, A., Reichel, F. D., Solomon, H. Y., & Turvey, M. T. (1989). Visually perceiving what is reachable. *Ecological Psychology*, 1, 27–54. doi:10.1207/s15326969eco0101\_3
- Carello, C., Wagman, J. B., & Turvey, M. T. (2005). Acoustic specification of object properties. In J. D. Anderson & B. F. Anderson (Eds.), *Moving image theory: Ecological considerations* (pp. 79–104). Carbondale, IL: Southern Illinois University Press.
- Carlson-Smith, C., & Wiener, W. R. (1996). The auditory skills necessary for echolocation: A new explanation. *Journal of Visual Impairment & Blindness*, 90, 21–35. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=508580609&site=ehost-live>
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audio-visual speech. *PLoS Computational Biology*, 5(7), 1–18. doi:10.1371/journal.pcbi.1000436
- Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics*, 75(5), 790–811. doi:10.3758/s13414-013-0475-4
- Chen, T. H., & Massaro, D. W. (2008). Seeing pitch: Visual information for lexical tones of Mandarin-Chinese. *The Journal of the Acoustical Society of America*, 123(4), 2356–2366. doi:10.1121/1.2839004
- Cheung, S. H., Fang, F., & He, S. (2009). Retinotopically specific reorganization of visual cortex for tactile pattern recognition. *Current Biology*, 19, 596–601. doi:10.1016/j.cub.2009.02.063
- Choe, C. S., Welch, R. B., Gilford, R. M., & Juola, J. F. (1975). The “ventriloquism effect”: Visual dominance or response bias. *Perception & Psychophysics*, 18(1), 55–60. doi:10.3758/BF03199367
- Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, 113, 495–506. doi:10.1016/S1388-2457(02)00024-X
- Collignon, O., Lassonde, M., Lepore, F., Bastien, D., & Veraart, C. (2007). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cerebral Cortex*, 17(2), 457–465. doi:10.1093/cercor/bhj162
- Collignon, O., Voss, P., Lassonde, M., & Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Experimental Brain Research*, 192(3), 343–358. doi:10.1007/s00221-008-1553-z
- Cooper, W. E., Blumstein, S. E., & Nigro, G. (1975). Articulatory effects on speech perception: A preliminary report. *Journal of Phonetics*, 3, 87–98.
- Cooper, W. E., Billings, D., & Cole, R. A. (1976). Articulatory effects on speech perception: A second report. *Journal of Phonetics*, 4(3), 219–232.
- Cooper, W. E., & Lauritsen, M. R. (1974). Feature processing in the perception and production of speech. *Nature*, 252, 121–123. doi:10.1038/252121a0
- Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology-Paris*, 98(1–3), 249–258. doi:10.1016/j.jphysparis.2004.03.011
- Derrick, D., & Gick, B. (2013). Aerotactile integration from distal skin stimuli. *Multisensory Research*, 26(5), 405–416. doi:10.1163/22134808-00002427
- Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, 45(4), 187–203. doi:10.1002/dev.20033
- Després, O., Candas, V., & Dufour, A. (2005). Spatial auditory compensation in early-blind humans: Involvement of eye movements and/or attention orienting? *Neuropsychologia*, 43(13), 1955–1962. doi:10.1016/j.neuropsychologia.2005.03.002
- Dias, J. W., Cook, T. C., & Rosenblum, L. D. (2016). Influences of selective adaptation on perception of audiovisual speech. *Journal of Phonetics*, 56, 75–84. doi:10.1016/j.wocn.2016.02.004
- Dias, J. W., Cook, T. C., & Rosenblum, L. D. (in press). The McGurk effect and the primacy of multisensory perception. In A. G. Shapiro & D. Todorovic (Eds.), *Oxford compendium of visual illusions*. Oxford University Press.
- Dias, J. W., & Rosenblum, L. D. (2011). Visual influences on interactive speech alignment. *Perception*, 40, 1457–1466. doi:10.1068/p7071
- Dias, J. W., & Rosenblum, L. D. (2015a). Visibility of speech articulation enhances auditory phonetic convergence. *Attention, Perception, & Psychophysics*, 77(6). doi:10.3758/s13414-015-0982-6
- Dias, J. W., & Rosenblum, L. D. (2015b). Selective adaptation of crossmodal speech information: A case of small but significant effects. *Abstracts of the 56th Meeting of the Psychonomics Society*. 202–203.
- Dufour, A., & Gérard, Y. (2000). Improved auditory spatial sensitivity in near-sighted subjects. *Cognitive Brain Research*, 10(1–2), 159–165. doi:10.1016/S0926-6410(00)00014-8
- Dupont, S., Aubin, J., & Menard, L. (2005). The study of the McGurk effect in 4 and 5-year-old French Canadian children. *ZAS Papers in Linguistics*, 40, 1–17.
- Eberhardt, S. P., Auer Jr, E. T., & Bernstein, L. E. (2014). Multisensory training can promote or impede visual



- perceptual learning of speech stimuli: Visual-tactile vs. visual-auditory training. *Frontiers in Human Neuroscience*, 8, 185. doi:10.3389/fnhum.2014.00829
- Ehrenstein, W. H., & Reinhardt-Rutland, A. H. (1996). A cross-modal aftereffect: Auditory displacement following adaptation to visual motion. *Perceptual and Motor Skills*, 82, 23–26. doi:10.2466/pms.1996.82.1.23
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3–28.
- Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 189–202). Cambridge, MA: MIT Press.
- Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3), 816–828. doi:10.1037/0096-1523.17.3.816
- Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology*, 18(16), 1262–1266. doi:10.1016/j.cub.2008.07.066
- Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2003). The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychologica*, 113, 315–327. doi:10.1016/S0001-6918(03)00043-X
- Fuster-Duran, A. (1996). Perception of conflicting audio-visual speech: An examination across Spanish and German. *Speechreading by Humans and Machines*, 150, 135–143. doi:10.1007/978-3-662-13015-5\_9
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13, 361–377. doi:10.3758/BF03193857
- Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Experimental Brain Research*, 167, 66–75. doi:10.1007/s00221-005-0008-z
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462, 502–504. doi:10.1038/nature08572
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. doi:10.1037/0033-295X.105.2.251
- Gordon, M. S., & Rosenblum, L. D. (2004). Perception of sound-obstructing surfaces using body-scaled judgments. *Ecological Psychology*, 16(2), 87–113. doi:10.1207/s15326969eco1602\_1
- Gordon, M. S., & Rosenblum, L. D. (2005). Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. *Perception & Psychophysics*, 67(4), 580–594. doi:10.3758/BF03193516
- Green, K. P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell & B. Dodd (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and audiovisual speech* (pp. 3–25). Hove: Psychology Press.
- Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *The Journal of Comparative Neurology*, 502(6), 924–952. doi:10.1002/cne.21326
- Horowitz, S. S., Cheney, C. A., & Simmons, J. A. (2004). Interaction of vestibular, echolocation, and visual modalities guiding flight by the big brown bat, *Eptesicus fuscus*. *Journal of Vestibular Research: Equilibrium & Orientation*, 14(1), 17–32. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&db=a2h&AN=12968600&site=ehost-live>
- Hulusic, V., Harvey, C., Debattista, K., Tsingos, N., Walker, S., Howard, D., & Chalmers, A. (2012). Acoustic rendering and auditory-visual cross-modal perception and interaction. *Computer Graphics Forum*, 31(1), 102–131. doi:10.1111/j.1467-8659.2011.02086.x
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245–1248. doi:10.1073/pnas.0810063106
- Jaekl, P., Pesquita, A., Alsius, A., Munhall, K., & Soto-Faraco, S. (2015). The contribution of dynamic visual cues to audiovisual speech perception. *Neuropsychologia*, 75, 402–410. doi:10.1016/j.neuropsychologia.2015.06.025
- Jerger, S., Damian, M. F., Spence, M. J., Tye-Murray, N., & Abdi, H. (2009). Developmental shifts in children's sensitivity to visual speech: A new multimodal picture-word task. *Journal of Experimental Child Psychology*, 102, 40–59. doi:10.1016/j.jecp.2008.08.002
- Karns, C. M., Dow, M. W., & Neville, H. J. (2012). Altered cross-modal processing in the primary auditory cortex of congenitally deaf adults: A visual-somatosensory fMRI study with a double-flash illusion. *Journal of Neuroscience*, 32(28), 9626–9638. doi:10.1523/JNEUROSCI.6488-11.2012
- Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, 416, 172–174. doi:10.1038/416172a
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712–719. doi:10.1016/j.tins.2004.10.007
- Kolarik, A. J., Cirstea, S., Pardhan, S., & Moore, B. C. J. (2014). A summary of research investigating echolocation abilities of blind and sighted humans. *Hearing Research*, 310, 60–68. doi:10.1016/j.heares.2014.01.010
- Kolarik, A. J., Timmis, M. A., Cirstea, S., & Pardhan, S. (2014). Sensory substitution information informs locomotor adjustments when walking through apertures. *Experimental Brain Research*, 232(3), 975–984. doi:10.1007/s00221-013-3809-5
- Konkle, T., Wang, Q., Hayward, V., & Moore, C. I. (2009). Motion aftereffects transfer between touch and vision. *Current Biology*, 19(9), 745–750. doi:10.1016/j.cub.2009.03.035
- Kral, A., Schröder, J.-H., Klinke, R., & Engel, A. K. (2003). Absence of cross-modal reorganization in the primary auditory cortex of congenitally deaf cats. *Experimental Brain Research*, 153(4), 605–613. doi:10.1007/s00221-003-1609-z
- Kuczaj, S., Solangi, M., Hoffland, T., & Romagnoli, M. (2008). Recognition and discrimination of human

- actions across the senses of echolocation and vision in the Bottlenose Dolphin: Evidence for Dolphin cross-modal integration of dynamic information. *International Journal of Comparative Psychology*, 21(2). Retrieved from <http://www.escholarship.org/uc/item/8805892k>
- Kupers, R., Chebat, D. R., Madsen, K. H., Paulson, O. B., & Ptito, M. (2010). Neural correlates of virtual route recognition in congenital blindness. *Proceedings of the National Academy of Sciences*, 107(28), 12716–12721. doi:10.1073/pnas.1006199107
- Lachs, L., & Pisoni, D. B. (2004a). Cross-modal source information and spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 378–396. doi:10.1037/0096-1523.30.2.378
- Lachs, L., & Pisoni, D. B. (2004b). Crossmodal source identification in speech perception. *Ecological Psychology*, 16(3), 159–187. doi:10.1207/s15326969eco1603\_1
- Lachs, L., & Pisoni, D. B. (2004c). Specification of cross-modal source information in isolated kinematic displays of speech. *The Journal of the Acoustical Society of America*, 116(1), 507–518. doi:10.1121/1.1757454
- Lander, K., & Davies, R. (2008). Does face familiarity influence speechreadability? *The Quarterly Journal of Experimental Psychology*, 61(7), 961–967. doi:10.1080/17470210801908476
- Larsson, P., Västfjäll, D., & Kleiner, M. (2001). *Better presence and performance in virtual environments by improved binaural sound rendering*. Paper presented at the 22nd International Conference: Virtual, Synthetic, and Entertainment Audio.
- Lee, D., van der Weel, F. R., Hitchcock, T., Matejowsky, E., & Pettigrew, J. (1992). Common principle of guidance by echolocation and vision. *Journal of Comparative Physiology A*, 171(5), 563–571. doi:10.1007/BF00194105
- Levitan, C. A., Ban, Y.-H. A., Stiles, N. R. B., & Shimojo, S. (2015). Rate perception adapts across the senses: Evidence for a unified timing mechanism. *Scientific Reports*, 5. doi:10.1038/srep08857
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36. doi:10.1016/0010-0277(85)90021-6
- Lomber, S. G., Meredith, M. A., & Kral, A. (2010). Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nature Neuroscience*, 13(11), 1421–1427. doi:10.1038/nn.2653
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11), 1432–1438. doi:10.1038/nn1790
- MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, 24(3), 253–257. doi:10.3758/BF03206096
- Maempel, H., & Jentsch, M. (2013). *Audio-visual interaction of size and distance perception in concert halls – a preliminary study*. Paper presented at the International Symposium on Room Acoustics, Toronto, Canada.
- Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-Specific organization in the human brain does not require visual experience. *Neuron*, 63(3), 397–405. doi:10.1016/j.neuron.2009.07.012
- Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- Massaro, D. W., & Cohen, M. M. (1993). Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Communication*, 13, 127–134. doi:10.1016/0167-6393(93)90064-R
- Massaro, D. W., Cohen, M. M., Gesi, A., & Heredia, R. (1993). Bimodal speech perception: An examination across languages. *Journal of Phonetics*, 21, 445–478.
- Matsumiya, K. (2013). Seeing a haptically explored face: Visual facial-expression aftereffect from haptic adaptation to a face. *Psychological Science*. doi:10.1177/0956797613486981
- Matteau, I., Kupers, R., Ricciardi, E., Pietrini, P., & Ptito, M. (2010). Beyond visual, aural and haptic movement perception: hMT+ is activated by electrotactile motion stimulation of the tongue in sighted and in congenitally blind individuals. *Brain Research Bulletin*, 82(5–6), 264–270. doi:10.1016/j.brainresbull.2010.05.001
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748. doi:10.1038/264746a0
- Merabet, L. B., & Pascual-Leone, A. (2010). Neural reorganization following sensory loss: The opportunity of change. *Nature Reviews Neuroscience*, 11, 44–52. doi:10.1038/nrn2758
- Meredith, M. A., & Allman, B. L. (2012). Early hearing-impairment results in crossmodal reorganization of ferret core auditory cortex. *Neural Plasticity*, 2012, 6, 1–13. doi:10.1155/2012/601591
- Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72(6), 1614–1625. doi:10.3758/APP.72.6.1614
- Mills, A. E. (1987). The development of phonology in the blind child. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip reading* (pp. 145–162). Hillsdale, NJ: Lawrence Erlbaum.
- Milne, J. L., Arnott, S. R., Kish, D., Goodale, M. A., & Thaler, L. (2015). Parahippocampal cortex is involved in material processing via echoes in blind echolocation experts. *Vision Research*, 109(Part B), 139–148. doi:10.1016/j.visres.2014.07.004
- Milne, J. L., Goodale, M. A., & Thaler, L. (2014). The role of head movements in the discrimination of 2-D shape by blind echolocation experts. *Attention, Perception, & Psychophysics*, 76(6), 1828–1837. doi:10.3758/s13414-014-0695-2
- Milton, C., & Dallenbach, K. M. (1950). "Facial vision." The rôle of pitch and loudness in the perception of obstacles by the blind. *The American Journal of Psychology*, 63(4), 485–515. doi:10.2307/1418868
- Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, 58(3), 351–362.
- Munhall, K. G., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances*

- in the psychology of speechreading and audio-visual speech (pp. 123–139). East Sussex: Psychology Press.
- Munhall, K. G., & Vatikiotis-Bateson, E. (2004). Spatial and temporal constraints on audiovisual speech perception. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 177–187). Cambridge, MA: MIT Press.
- Musacchia, G., Sams, M., Nicol, T., & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research*, 168(1–2), 1–10. doi:10.1007/s00221-005-0071-5
- Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, 71, 4–12. doi:10.1007/s00426-005-0031-5
- Nygaard, L. C. (2005). Perceptual integration of linguistic and non-linguistic properties of speech. In D. Pisoni & R. Remez (Eds.), *Handbook of speech perception* (pp. 390–419). Malden, MA: Blackwell.
- Orbach, D. N., & Fenton, B. (2010). Vision impairs the abilities of bats to avoid colliding with stationary obstacles. *PLoS One*, 5(11), e13912. doi:10.1371/journal.pone.0013912
- Ostrand, R., Blumstein, S. E., & Morgan, J. L. (2011). When hearing lips and seeing voices becomes perceiving speech: Auditory-visual integration in lexical access. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33, 1376–1381.
- Pack, A. A., & Herman, L. M. (1995). Sensory integration in the bottlenosed dolphin: Immediate recognition of complex shapes across the senses of echolocation and vision. *The Journal of the Acoustical Society of America*, 98(2), 722–733. doi:10.1121/1.413566
- Pack, A. A., Herman, L. M., Hoffmann-Kuhnt, M., & Branstetter, B. K. (2002). The object behind the echo: dolphins (*Tursiops truncatus*) perceive object shape globally through echolocation. *Behavioural Processes*, 58(1–2), 1–26. doi:10.1016/S0376-6357(01)00200-5
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393. doi:10.1121/1.2178720
- Pardo, J. S., & Remez, R. E. (2006). The perception of speech. In M. Traxler & M. A. Gernsbacher (Eds.), *The handbook of psycholinguistics* (2nd ed., pp. 201–248). New York, NY: Academic Press.
- Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research*, 134(Chapter 27), 1–19.
- Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottonen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3T. *Auditory and Vestibular Systems*, 16(2), 125–128.
- Rebillard, G., Carlier, E., Rebillard, M., & Pujol, R. (1977). Enhancement of visual responses on the primary auditory cortex of the cat after an early destruction of cochlear receptors. *Brain Research*, 129(1), 162–164. doi:10.1016/0006-8993(77)90980-5
- Reed, C. M., Rabinowitz, W. M., Durlach, N. I., Braida, L. D., Conway-Fithian, S., & Schultz, M. C. (1985). Research on the Tadoma method of speech communication. *The Journal of the Acoustical Society of America*, 77(1), 247–257. doi:10.1121/1.392266
- Reich, L., Maidenbaum, S., & Amedi, A. (2012). The brain as a flexible task machine: Implications for visual rehabilitation using noninvasive vs. invasive approaches. *Current Opinion in Neurology*, 25(1), 86–95. doi:10.1097/WCO.0b013e32834ed723
- Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Remez, R. E., Fellowes, J. M., Pisoni, D. B., Goh, W. D., & Rubin, P. E. (1998). Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. *Speech Communication*, 26, 65–73. doi:10.1016/S0167-6393(98)00050-8
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212, 947–949. doi:10.1126/science.7233191
- Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture? *Neuroscience & Biobehavioral Reviews*, 41, 64–77. doi:10.1016/j.neubiorev.2013.10.006
- Robart, R. L., & Rosenblum, L. D. (2005). Hearing space: Identifying rooms by reflected sound. In H. Heft & K. L. Marsh (Eds.), *Studies in perception and action XIII* (pp. 153–161). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics*, 30(4), 309–314. doi:10.3758/BF03206144
- Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLOS Biology*, 13(2), e1002073. doi:10.1371/journal.pbio.1002073
- Rosen, S. M., Fourcin, A. J., & Moore, B. C. J. (1981). Voice pitch as an aid to lipreading. *Nature*, 291(5811), 150–152. doi:10.1038/291150a0
- Rosenblum, L. D. (2004). Perceiving articulatory events: Lessons for an ecological psychoacoustics. In J. G. Neuhoﬀ (Ed.), *Ecological psychoacoustics* (pp. 219–248). San Diego, CA: Elsevier Academic Press.
- Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. Pisoni & R. Remez (Eds.), *Handbook of speech perception* (pp. 51–78). Malden, MA: Blackwell.
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, 17(6), 405–409. doi:10.1111/j.1467-8721.2008.00615.x
- Rosenblum, L. D. (2010). *See what i'm saying: The extraordinary powers of our five senses*. New York, NY: W. W. Norton & Company.
- Rosenblum, L. D. (2013). A confederacy of senses. *Scientific American*, 308, 72–75. doi:10.1038/scientificamerican0113-72
- Rosenblum, L. D., Gordon, M. S., & Jarquin, L. (2000). Echolocating distance by moving and stationary



- listeners. *Ecological Psychology*, 12(3), 181–206. doi:10.1207/S15326969ECO1203\_1
- Rosenblum, L. D., Johnson, J. A., & Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech Language and Hearing Research*, 39(6), 1159–1170. doi:10.1044/jshr.3906.1159
- Rosenblum, L. D., Miller, R. M., & Sanchez, K. (2007). Lip-read me now, hear me better later: Cross-modal transfer of talker-familiarity effects. *Psychological Science*, 18(5), 392–396. doi:10.1111/j.1467-9280.2007.01911.x
- Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), 318–331. doi:10.1037/0096-1523.22.2.318
- Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, 59(3), 347–357. doi:10.3758/BF03211902
- Rosenblum, L. D., Smith, N. M., Nichols, S. M., Hale, S., & Lee, J. (2006). Hearing a face: Cross-modal speaker matching using isolated visible speech. *Perception & Psychophysics*, 68(1), 84–93. doi:10.3758/BF03193658
- Rosenblum, L. D., Wuestefeld, A. P., & Anderson, K. L. (1996). Auditory reachability: An affordance approach to the perception of sound source distance. *Ecological Psychology*, 8(1), 1–24. doi:10.1207/s15326969eco0801\_1
- Ross, H. E. (1969). When is a weight not illusory? *Quarterly Journal of Experimental Psychology*, 21(4), 346–355. doi:10.1080/14640746908400230
- Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, 54(3), 406–416. doi:10.3758/BF03205276
- Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *The Journal of the Acoustical Society of America*, 95(6), 3658–3661. doi:10.1121/1.409935
- Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127, 141–145. doi:10.1016/0304-3940(91)90914-F
- Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö, R. (1998). McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. *Speech Communication*, 26(1–2), 75–87. doi:10.1016/S0167-6393(98)00051-X
- Sams, M., Mottonen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23, 429–435. doi:10.1016/j.cogbrainres.2004.11.006
- Samuel, A. G., & Lieblich, J. (2014). Visual speech acts differently than lexical context in supporting speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1479–1490. doi:10.1037/a0036656
- Sanchez, K., Dias, J. W., & Rosenblum, L. D. (2013). Experience with a talker can transfer across modalities to facilitate lipreading. *Attention, Perception, & Psychophysics*, 75(7), 1359–1365. doi:10.3758/s13414-013-0534-x
- Sanchez, K., Miller, R. M., & Rosenblum, L. D. (2010). Visual influences on alignment to voice onset time. *Journal of Speech Language and Hearing Research*, 53, 262–272. doi:10.1044/1092-4388(2009/08-0247)
- Sandmann, P., Dillier, N., Eichele, T., Meyer, M., Kegel, A., Pascual-Marqui, R. D., ... Debener, S. (2012). Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. *Brain*, 135, 555–568. doi:10.1093/brain/awr329
- Scharine, A. A., Letowski, T. R., Mermagen, T., & Henry, P. P. (2010). Learning to detect and identify acoustic environments from reflected sound. *Military Psychology*, 22(1), 24–40. doi:10.1080/08995600903206461
- Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*, 14(1), 187–198. doi:10.1016/S0926-6410(02)00073-3
- Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the Macaque monkey. *Journal of Neurophysiology*, 85(3), 1322–1327.
- Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36, 888–897. doi:10.1068/p5635
- Scott, G. D., Karns, C. M., Dow, M. W., Stevens, C., & Neville, H. J. (2014). Enhanced peripheral visual processing in congenitally deaf humans is supported by multiple brain regions, including primary auditory cortex. *Frontiers in Human Neuroscience*, 8, 5925. doi:10.3389/fnhum.2014.00177
- Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, 90(4), 1797–1805. doi:10.1121/1.401660
- Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21(4), 427–444.
- Shams, L. (2011). Early integration and Bayesian causal inference in multisensory perception. In M. M. Murray & M. T. Wallace, *The neural bases of multisensory processes* (pp. 217–232). Boca Raton, FL: CRC Press/Tay & Francis.
- Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: An MEG study. *Neuroscience Letters*, 378, 76–81. doi:10.1016/j.neulet.2004.12.035
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, 408, 788. doi:10.1038/35048669
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14, 147–152. doi:10.1016/S0926-6410(02)00069-1
- Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *NeuroReport*, 12, 3849–3852. doi:10.1097/00001756-200112040-00049
- Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, 7, 269–284. doi:10.1016/j.plrev.2010.04.006

- Shams, L., Wozny, D. R., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology*, 2, 264. doi:10.3389/fpsyg.2011.00264
- Smiley, J. F., & Falchier, A. (2009). Multisensory connections of monkey auditory cerebral cortex. *Hearing Research*, 258(1–2), 37–46. doi:10.1016/j.heares.2009.06.019
- Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *The Journal of Comparative Neurology*, 502(6), 894–923. doi:10.1002/cne.21325
- Smith, D., & Burnham, D. K. (2012). Facilitation of Mandarin tone perception by visual speech in clear and degraded audio: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, 131(2), 1480–1489. doi:10.1121/1.3672703
- Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 580–587. doi:10.1037/a0013483
- Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, 19(12), 1964–1973. doi:10.1162/jocn.2007.19.12.1964
- Stekelenburg, J., & Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Frontiers in Integrative Neuroscience*, 6. doi:10.3389/fnint.2012.00026
- Stewart, D. L., & Starr, A. (1970). Absence of visually influenced cells in auditory cortex of normal and congenitally deaf cats. *Experimental Neurology*, 28(3), 525–528. doi:10.1016/0014-4886(70)90190-1
- Stoffregen, T. A., & Pittenger, J. B. (1995). Human echolocation as a basic form of perception and action. *Ecological Psychology*, 7(3), 181–216. doi:10.1207/s15326969ec0703\_2
- Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America*, 74(3), 695–705. doi:10.1121/1.389855
- Strelow, E. R., & Brabyn, J. A. (1982). Locomotion of the blind controlled by natural sound cues. *Perception*, 11(6), 635–640. doi:10.1068/p110635
- Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people “seeing” by using sounds. *Current Biology*, 24(6), 687–692. doi:10.1016/j.cub.2014.02.010
- Striem-Amit, E., Dakwar, O., Hertz, U., Meijer, P., Stern, W., Pascual-Leone, A., & Amedi, A. (2011). The neural network of sensory-substitution object shape recognition. *Functional Neurology, Rehabilitation, and Ergonomics*, 1(2), 271–278.
- Striem-Amit, E., Ovadia-Caro, S., Caramazza, A., Margulies, D. S., Villringer, A., & Amedi, A. (2015). Functional connectivity of visual cortex in the blind follows retinotopic organization principles. *Brain*, 138, 1679–1695. doi:10.1093/brain/awv083
- Stropahl, M., Plotz, K., Schönfeld, R., Lenarz, T., Sandmann, P., Yovel, G., ... Debener, S. (2015). Cross-modal reorganization in cochlear implant users: Auditory cortex contributes to visual face processing. *NeuroImage*, 121, 159–170. doi:10.1016/j.neuroimage.2015.07.062
- Sumby, W. H., & Pollack, I. (1954). Visual contribution of speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26, 212–215. doi:10.1121/1.1907309
- Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In I. Barbara & C. Ruth (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3–51). Hillsdale, NJ: Erlbaum.
- Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *The Quarterly Journal of Experimental Psychology Section A*, 36, 51–74. doi:10.1080/14640748408401503
- Sutherland, C. A. M., Thut, G., & Romei, V. (2014). Hearing brighter: Changing in-depth visual perception through looming sounds. *Cognition*, 132(3), 312–323. doi:10.1016/j.cognition.2014.04.011
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, 108(3), 850–855. doi:10.1016/j.cognition.2008.05.009
- Teng, S., Puri, A., & Whitney, D. (2012). Ultrafine spatial acuity of blind expert human echolocators. *Experimental Brain Research*, 216, 483–488. doi:10.1007/s00221-011-2951-1
- Teramoto, W., Manaka, Y., Hidaka, S., Sugita, Y., Miyauchi, R., Sakamoto, S., ... Suzuki, Y. (2010). Visual motion perception induced by sounds in vertical plane. *Neuroscience Letters*, 479, 221–225. doi:10.1016/j.neulet.2010.05.065
- Thaler, L. (2013). Echolocation may have real-life advantages for blind people: An analysis of survey data. *Frontiers in Physiology*, 4, 1–8. doi:10.3389/fphys.2013.00098
- Thaler, L., Arnott, S. R., & Goodale, M. A. (2011). Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS One*, 6(5), e20162. doi:10.1371/journal.pone.0020162
- Thompson, W. F., Russo, F. A., & Livingstone, S. R. (2010). Facial expressions of singers influence perceived pitch relations. *Psychonomic Bulletin & Review*, 17(3), 317–322. doi:10.3758/PBR.17.3.317
- Thompson, W. F., Russo, F. A., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition and Emotion*, 22(8), 1457–1470. doi:10.1080/02699930701813974
- Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, 5, 839. doi:10.3389/fpsyg.2014.00725
- Treille, A., Cordeboeuf, C., Vilain, C., & Sato, M. (2014). Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions. *Neuropsychologia*, 57, 71–77. doi:10.1016/j.neuropsychologia.2014.02.004
- Treille, A., Vilain, C., & Sato, M. (2014). The sound of your lips: Electrophysiological cross-modal interactions during hand-to-face and face-to-face speech perception. *Frontiers in Psychology*, 5, 195. doi:10.3389/fpsyg.2014.00420



- Valente, D. L., & Braasch, J. (2010). Subjective scaling of spatial room acoustic parameters influenced by visual environmental cues. *The Journal of the Acoustical Society of America*, 128(4), 1952–1964. doi:10.1121/1.3478797
- Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the “unity effect” reveals that speech is special. *Journal of Vision*, 8(9), 14–14. doi:10.1167/8.9.14
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, 118(2), 157–170. doi:10.1016/j.cognition.2010.11.010
- Vroomen, J., & Baart, M. (2012). Phonetic recalibration in audiovisual speech. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 363–379). Boca Raton, FL: CRC Press.
- Vroomen, J., & de Gelder, B. (2004). Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 581–596). Cambridge, MA: MIT Press.
- Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics*, 76(2), 591–604. doi:10.3758/s13414-013-0582-2
- van Wassenhove, V. (2013). Speech through ears and eyes: Interfacing the senses with the supramodal brain. *Frontiers in Psychology*, 4, 1–17. doi:10.3389/fpsyg.2013.00388
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4), 1181–1186. doi:10.1073/pnas.0408949102
- Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 31(3), 1247–1256. doi:10.1016/j.neuroimage.2006.01.016
- Weisser, V., Stilla, R., Peltier, S., Hu, X., & Sathian, K. (2005). Short-term visual deprivation alters neural processing of tactile form. *Experimental Brain Research*, 166(3–4), 572–582. doi:10.1007/s00221-005-2397-4
- Williams, D. R., Verbrugge, R., & Studdert-Kennedy, M. (1983). Judging sine wave stimuli as speech and as non-speech. *The Journal of the Acoustical Society of America*, 74(S1), S66. doi:10.1121/1.2021086
- Yakel, D. A. (2000). *Effects of time-varying information on vowel identification accuracy in visual speech perception* (Ph.D.), University of California, Riverside.
- Yehia, H. C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion, and speech acoustics. *Journal of Phonetics*, 30(3), 555–568. doi:10.1006/jpho.2002.0165
- Yehia, H. C., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, 26(1–2), 23–43. doi:10.1016/S0167-6393(98)00048-X