Received: 18 December 2015 | Revised: 30 November 2016 | Accepted: 25 January 2017
DOI: 10.1111/lnc3.12237

ARTICLE

WILEY

Audiovisual speech perception: A new approach and implications for clinical populations

Julia Irwin | Lori DiBlasi

LEARN Center, Haskins Laboratories Inc., USA

Correspondence
Julia Irwin, LEARN Center, Haskins Laboratories Inc., 300 George Street, New Haven, CT 06511, USA.
Email: irwin@haskins.yale.edu

Funding Information
NIH, Grant/Award Number: DC013864. National Institutes of Health, Grant/Award Number: DC013864.

Abstract
This selected overview of audiovisual (AV) speech perception examines the influence of visible articulatory information on what is heard. Thought to be a cross-cultural phenomenon that emerges early in typical language development, variables that influence AV speech perception include properties of the visual and the auditory signal, attentional demands, and individual differences. A brief review of the existing neurobiological evidence on how visual information influences heard speech indicates potential loci, timing, and facilitatory effects of AV over auditory only speech. The current literature on AV speech in certain clinical populations (individuals with an autism spectrum disorder, developmental language disorder, or hearing loss) reveals differences in processing that may inform interventions. Finally, a new method of assessing AV speech that does not require obvious cross-category mismatch or auditory noise was presented as a novel approach for investigators.

1 | SPEECH IS MORE THAN A SOUND

Speech is generally thought to consist of sound. An overview of speech research reflects this, with many influential papers on speech perception reporting effects in the auditory domain (e.g., DeCasper & Spence, 1986; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Hickok & Poeppel, 2007; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; McClelland & Elman, 1986; Saffran, Aslin, & Newport, 1996; Werker & Tees, 1984). Certainly, the study of the speech sound is of great import and influence in a range of fields, including cognitive and developmental psychology, linguistics, and speech and hearing sciences. However, during speech production, the entire articulatory apparatus is implicated, some of which is visible on the face of the speaker. For example, lip rounding (as in the case of /u/) or lip closure (for /b/) can be seen on the speaker's face and, if seen, can provide information for the listener about what was said. This visible articulatory information facilitates language processing (Desjardins, Rogers, & Werker, 1997; Lachs & Pisoni, 2004; MacDonald & McGurk, 1978; MacDonald, Andersen, & Bachmann, 2000; McGurk & MacDonald, 1976; Reisberg, Mclean, & Goldfield, 1987). Moreover, typical speech and language development is thought to take place in this audiovisual (AV) context, fostering native language acquisition (Bergeson & Pisoni, 2004; Desjardins et al., 1997; Lachs, Pisoni, & Kirk, 2001; Lewkowicz & Hansen-Tift, 2012; Legerstee, 1990; Meltzoff & Kuhl, 1994). Sighted speakers produce vowels further apart in articulatory space than those of blind speakers, further suggesting that the development of speech perception and production is influenced by experience with the speaking face (Ménard, Dupont, Baum, & Aubin, 2009).

2 | AUDIOVISUAL SPEECH PERCEPTION: VISUAL GAIN AND THE McGURK EFFECT

Over 60 years ago, Sumby and Pollack (1954) reported that visible articulatory information assists listeners in the identification of speech in auditory noise, creating a "visual gain" over auditory speech alone (also see Erber, 1975; Grant & Seitz, 2000; MacLeod & Summerfield, 1987; Payton, Uchanski, & Braida, 1994; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Schwartz, Bertommier, & Savariaux, 2004). It is clear that in noisy listening conditions, the face can assist the listener in understanding the speaker's message. Significantly, visible articulatory information influences what a listener hears even when the auditory signal is in clear listening conditions. One powerful demonstration of the influence of visual information on what is heard is during perception of mismatched AV speech. McGurk and MacDonald (1976) presented mismatched audio and video consonant vowel consonant vowel (CVCV) tokens to perceivers. That is, the speaker was recorded producing a visual (for example) /gaga/, while an auditory /baba/ stimulus produced by the same speaker was dubbed over the face. Listeners often respond to this mismatch by reporting that they hear /dada/. In this demonstration, perceptual information about visible place of articulation integrates with acoustic voicing and manner (Goldstein & Fowler, 2003). Perceivers watching these dubbed productions sometimes reported hearing consonants that combined the places of articulation of the visual and auditory tokens (e.g., a visual /ba/ + an auditory /ga/ would be heard as /bga/), "fused" the two places (e.g., visual /ga/ + auditory /ba/ heard as /da/), or reflected the visual place information alone (visual /va/ + auditory /ba/ heard as /va/). Based on McGurk and MacDonald's now classic 1976 paper, this phenomenon is frequently referred to as the "McGurk effect" or "McGurk illusion" (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Brancazio, Best, & Fowler, 2006; Brancazio & Miller, 2005; Green, 1994; MacDonald & McGurk, 1978; Rosenblum, 2008; Soto-Faraco & Alsius, 2009; Walker, Bruce, & O'Malley, 1995; Windmann, 2004).

The McGurk Effect is often described as robust, occurring even if perceivers are aware of the manipulation (Rosenblum, & Saldaña, 1996), in the case when female face and male voice are dubbed (e.g., Green, Kuhl, Meltzoff, & Stevens, 1991; Johnson, Strand, & D'Imperio, 1999) and if the audio and visual signals are not temporally aligned (e.g., Munhall, Gribble, Sacco, & Ward, 1996).[1] Additional variables that have been explored using McGurk type stimuli include sex of the listener, where women are more visually influenced than men with very brief visual stimuli (e.g., Irwin, Whalen, and Fowler, 2006), gaze to the speaker's face (where direct gaze on the face of the speaker need not be present for the effect to occur; Paré, Richler, ten Hove, & Munhall 2003) and the role of attention, where visual influence is attenuated when visual attention is drawn to another stimulus placed on the speaking face (e.g., Alsius et al., 2005; Tiippana, Andersen, & Sams, 2004). Quality of the visual signal has also been directly manipulated in the study of AV speech. Even in degraded visual signals (MacDonald, Anderson, & Bachmann, 2000; Munhall, Kroos, & Vatikiotis-Bateson, 2002) and point-light displays of the face (Rosenblum & Saldaña, 1996), Callan et al. (2004) yield a visual influence on heard speech. Further, Munhall, Jones, Callan, Kuratate, and Vatikiotis-Bateson (2004) and Davis and Kim (2006) show an increase in intelligibility when a speaker's head movements are visible to the perceiver, even for some stimuli where only the upper part of the face is available (Davis and Kim, 2006).

However, there are a number of clear constraints on the McGurk effect as a method. Significant individual variation exists in response to McGurk stimuli, with some individuals more influenced than others (Nath & Beauchamp, 2012; Schwartz, 2010). Variability in response has been reported dependent on the listener's native language (Kuhl, Tsuzaki, Tohkura, & Meltzoff, 1994; Sekiyama, 1997; Sekiyama & Burnham, 2008; Sekiyama & Tokura, 1993). Sekiyama and colleagues report an attenuated McGurk effect for native adult speakers of Japanese as compared to native speakers of English, potentially due to differential cultural patterns of gaze, where Japanese speakers may gaze less directly to another's face. (However, note that Magnotti et al. (2015) report no differences in frequency of McGurk effect within language for native Mandarin Chinese and American English speakers.) Further, the original study by McGurk and MacDonald (1976) used CVCV clusters and many studies have used CVs (e.g., Burnham & Dodd, 1996; Green et al., 1991; Irwin, Tornatore, Brancazio, & Whalen, 2011) or VCVs (e.g. Jones & Callan, 2003; Munhall et al., 1996), all low-level stimuli that carry little meaning for the average listener (Note, however, that word- or lexical-level McGurk effects have also been reported: Brancazio, 2004; Tye-Murray, Sommers, & Spehar, 2007; Windmann, 2004).

Theoretical accounts of the McGurk effect include an associative pairing between the auditory and visual speech signal in memory, due to their co-occurrence in natural speech (e.g., Diehl, & Kluender, 1989; Massaro, 1998; Stephens & Holt, 2010). In contrast, the motor theory of speech perception (e.g., Liberman et al., 1967) indicates that speech perception and action is based on articulation or motor movements of the vocal apparatus. Goldstein and Fowler (2003) propose there is a "common currency" between the seen and heard speech signal with the underlying currency the articulatory motor gestures that produce speech, whether visual or acoustic. In this manner, AV speech is detected and integrated as both vision and audition specify the underlying motor movements used to produce speech.

In summary, visual gain and the McGurk effect are the primary methods by which we have come to understand AV speech perception. Given that speech is noise is a specific perceptual condition and the constraints associated with the McGurk effect, there is a need for novel approaches that do not leverage noise or mismatch in AV speech.

3 | PATTERNS OF GAZE TO THE FACE: UPTAKE OF VISUAL INFORMATION

With respect to gaze behavior to a speaking face, adult perceivers have been reported to exhibit reduced gaze on the eyes and increased on the nose or mouth when background auditory noise is present (Buchan, Paré, & Munhall, 2008; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998; Yi, Wong, & Eizenman, 2013). In clear auditory listening conditions, Lansing and McConkie (2003) found that before and after a sentence is spoken, perceivers gaze to the eyes of a speaker, while gaze is largely toward the mouth of the speaker during the production of the sentence. Lewkowicz and colleagues have assessed patterns of gaze to a speaking face producing native and nonnative speech in infants (Lewkowicz & Hansen-Tift, 2012) and report a shift in focus toward the mouth from the eyes of a speaker that corresponds to the onset of speech production in the infant. In adults, gaze to the speaker's mouth is greater for an unfamiliar language in monolinguals, but not in bilinguals, and only in a speech task (Barenholtz, Mavica, & Lewkowicz, 2016).

4 | THE DEVELOPMENT OF AUDIOVISUAL SPEECH PERCEPTION

Sensitivity to visual speech has been demonstrated as early as 5 months of age (Burnham & Dodd, 1998; Burnham & Dodd, 2004; Desjardins & Werker, 2004; Kushnerenko, Teinonen, Volein, & Csibra, 2008; Meltzoff & Kuhl, 1994; Lewkowicz, 2000; Rosenblum, Schmuckler, & Johnson, 1997; Yeung & Werker, 2013) but is not consistently reported in infancy (Desjardins & Werker, 2004). Young children appear to be less visually influenced than older children and adults (Barutchu, Crewther, Kiely, Murphy, & Crewther, 2008; Desjardins et al., 1997; Hockley & Polka 1994; Lalonde & Holt, 2014; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Ross et al., 2011; Tremblay, Champoux, Bacon, Lepore, & Théoret, 2007). This developmental effect of increased visual influence may not hold throughout the life span, however, as older adults are reported to be poorer at lipreading than younger adults (Spehar, Tye-Murray, & Sommers, 2004; Sommers, Tye-Murray, & Spehar, 2005). Missing from the current literature is a mechanistic account of developmental differences in sensitivity to AV speech as other factors, such as differences in maturation, cognition, and attention across the lifespan might predict the observed U-shaped function in visual influence on what is heard.

5 | THE NEUROBIOLOGY OF AUDIOVISUAL SPEECH

Neural circuits for AV speech are largely overlapping to those associated with auditory speech processing in the brain (for a detailed review of auditory speech, see Blumstein & Myers, 2013). These areas include temporal superior temporal gyrus, superior temporal sulcus (STS), middle temporal gyrus and Heschl's gyrus, parietal (supramarginal and angular gyri), and frontal lobe structures (inferior frontal gyrus). Functional magnetic resonance imaging (fMRI) and magnetoencepholography studies of AV speech reveal activation primarily in the STS (most recently in the posterior STS, Okada, Venezia, Matchin, Saberi, & Hickok, 2013) and superior temporal gyrus, with some activation in the inferior frontal gyrus and middle temporal gyrus for mismatched or McGurk-type percepts (Belin, Zatorre, & Ahad 2002; Binder, Liebenthal, Possing, Medler, & Ward, 2004; Callan et al., 2003; Calvert, 2001; Calvert et al. 1997; Irwin, Frost, Mencl, Chen, & Fowler, 2011; Jones & Callan, 2003; MacSweeney et al., 2000; Miller & D'esposito, 2005; Möttönen, Schürmann, & Sams, 2004; Olson, Gatenby, & Gore, 2002; Sams et al., 1991; Skipper, van Wassenhove, Nusbaum, & Small, 2007; Skipper et al., 2007). With respect to a site for processing of AV speech, the STS has been implicated in a large number of studies of AV speech perception both in adults (e.g., Beauchamp, Nath, & Pasalar, 2010; Callan et al., 2003; Calvert et al., 2000; Jones & Callan, 2003, Nath & Beauchamp, 2012; Okada et al., 2013; Szycik, Tausche, & Münte, 2008) and children (Nath, Fava, & Beauchamp, 2011).

More recently, electrophysiological measures have been employed to examine AV speech, such as electroencephalography (EEG) and event-related potential (ERP; e.g., Pilling, 2009; Klucharev, Möttönen, & Sams, 2003; Molholm et al., 2002; Saint-Amour, De Sanctis, Molholm, Ritter, & Foxe, 2007; Van Wassenhove, Grant, & Poeppel, 2005). These techniques provide excellent temporal resolution, allowing for sensitive assessment of timing in response to AV stimuli, which can provide additional information beyond the location of activation as in fMRI. For example, Bernstein et al. (2008) propose a potential spatiotemporal AV speech processing circuit based on current density reconstructions of ERP data. Specifically, Bernstein et al. (2008) reported very early (less than 100 ms) simultaneous activation of the supramarginal gyrus, the angular gyrus, the intraparietal sulcus, the inferior frontal gyrus and the dorsolateral prefrontal cortex. Further, supramarginal gyrus activation in the left hemisphere was detected at 160 to 220 ms (Bernstein et al., 2008) for the AV condition as proposed to be a potential site for AV speech integration. Bernstein et al. (2008) note that the STS, frequently considered the primary location for processing of AV speech identified with fMRI methodology was not the earliest or the most prominent site of activation using ERP. This suggests that these more sensitive temporal techniques can reveal new, earlier activated circuits from those identified with fMRI.

In addition, there are a number of studies that look at components, which are sensitive to early auditory and visual features in the auditory N1 and P2. The auditory N1/P2 complex is known to be highly responsive to sounds, including auditory speech (e.g., Pilling, 2009; Tremblay, Kraus, McGee, Ponton, & Otis, 2001; Van Wassenhove et al., 2005). Van Wassenhove et al. (2005) and Pilling (2009) both report that congruent visual speech information presented with the auditory signal attenuates the amplitude of the N1/P2 auditory ERP response, resulting in lower peak amplitude and a shortening of their latency. Knowland et al. (2014) employed ERP to examine the development of AV speech perception using auditory-only (A), visual only (V), and AV (both matched and mismatched A + V) stimuli in children ranging in age from 6–11. Amplitude modulation increased over age, suggesting a developmental trend in the N1/P2 complex, while reduced latency was stable over age.

In general, EEG/ERP studies reveal that the combination of auditory and visual speech appears to dampen amplitude and speed processing of the speech signal.

6 | AV SPEECH PERCEPTION IN CLINICAL POPULATIONS

A comparison of auditory, visual, and AV speech stimuli allow us a noninvasive method of assessing unimodal (i.e., voice alone or face alone) and multimodal (face and voice) perception. Given that there may be impairments in processing of one or either modalities or deficits in processing of the two modalities, this may be a particularly important way to understand atypical speech and language development, such as in those individuals with autism spectrum disorders (ASDs), developmental language disorders, or with hearing impairments.

Individuals with an ASD, have been shown in a number of behavioral studies to be less influenced by visible articulation on what is heard, both in clear (De Gelder, Vroomen, & van der Heide, 1991; Mongillo et al. 2008; Taylor, Isaac, & Milne, 2010; Williams, Massaro, Peel, Bosseler & Suddendorf, 2004) and noisy listening conditions (Magnée, de Gelder, van Engeland, & Kemner, 2008; Smith & Bennetto, 2007). In addition, electrophysiological studies indicate weaker integration of the face and voice of the speaker (although this may vary with development; Brandwein et al., 2012; Russo et al. 2010). However, these findings are complicated by the characteristic avoidance of gaze to faces in individuals with an ASD (e.g., Irwin & Brancazio, 2014). Because of this, it may be that these listeners are not influenced by the visible speech signal simply because they are not looking at the face of the speaker. To control for this, Irwin, Tornatore, et al. (2011) tested children with ASD on a set of AV speech perception tasks, including an AV speech-in-noise (visual gain) and a McGurk task while concurrently recording eye fixation patterns. Significantly, Irwin, Tornatore, et al. (2011) excluded all trials where the participant did not fixate on the speaker's face. Even when fixated on the speaker's face, children with ASD were less influenced by visible articulatory information than their typically developing (TD) peers, both in the speech-in-noise tasks and with AV mismatched (McGurk) stimuli. Findings of Irwin, Frost, et al. (2011) indicate that fixation on the face is not sufficient to support efficient AV speech perception (this may not be the case for typical listeners, however, see Paré, et al., 2003). One possibility is that different gaze patterns on a face exhibited by individuals with ASD underlie perceptual differences in this population. To answer this question, an analysis of eye gaze data (Irwin & Brancazio, 2014) revealed that instead of gazing at the mouth during a speech in noise task, the children with ASD tended to spend more time directing their gaze to nonfocal areas of the face (also see Pelphrey et al., 2002), such as the ears, cheeks, and forehead, which carry little, if any, articulatory information. In particular, for speech in noise, as the speaker began to produce the articulatory signal, the TD children looked more to the mouth than did the children with ASD, who continued to gaze at nonfocal regions. Because the mouth is the primary source of phonetically relevant articulatory information available on the face (Thomas & Jordan, 2004), the findings may help account for the language and communication difficulties exhibited by children with ASD.

Like children on the autism spectrum, children who have been identified with developmental language disorders (Meronen, Tiippana, Westerholm, & Ahonen, 2013), specific language impairment (Norrix, Plante, Vance, & Boliek, 2007) evidence a weaker McGurk effect than their typically developing peers. Further, children with language learning impairments were less accurate at speechreading and speech in noise tasks when compared to TD peers (Knowland, Evans, Snell, & Rosen, 2016). These findings indicate that language problems may have implications for speech perception in general—impacting perception of both visual and auditory speech.

Because of the multimodal nature of AV speech, a natural extension is to examine visual and AV speech perception in children and adults with hearing loss, who have impairments in the auditory modality. Bernstein, Tucker and DeMorest (2000) report that individuals who have reduced hearing are more sensitive to visual phonetic information than hearing controls (also see Auer & Bernstein, 2007; Grant, Walden, & Seitz, 1998). AV speech perception has also been assessed in children who were prelingually deaf and regained the ability to hear with a cochlear implant (Bernstein & Grant, 2009; e.g., Bergeson, Pisoni & Davis, 2003; Kaiser, Kirk, Lachs, & Pisoni, 2003; Lachs et al., 2001; Moody-Antonio, Takayanagi, Masuda, & Auer, 2005; Schorr, Fox, van Wassenhove, & Knudsen, 2005). Bergeson et al. (2003) report that AV speech is particularly useful for children with cochlear implants in identification of speech as they gain experience with sound post-implantation (also see Lachs et al., 2001), although this effect may vary with experience, as age of implantation influenced AV fusion (Schorr et al., 2005).

Thus, the extant literature suggests that clinical populations may benefit from specific intervention that includes training on visual speech to support heard speech, because of difficulties processing the unimodal (auditory or visual) signals or because of weak integration (Irwin, Preston, Brancazio, D'Angelo, & Turcios, 2014).

7 | A NEW APPROACH: BEYOND VISUAL GAIN AND THE McGURK EFFECT

The selected review above illustrates the ubiquity of speech in noise and McGurk type stimuli in studies of AV speech perception. While much has been learned from these techniques, both noisy and mismatched stimuli have some potential drawbacks. For example, auditory noise may be especially disruptive for individuals with developmental disabilities (Alcántara, Weisblatt, Moore, & Bolton, 2004). Moreover, the McGurk effect creates a percept where what is heard is a separate syllable (or syllables) from either the visual or auditory signal, which generates conflict between the two modalities (Brancazio, 2004). These "illusory" percepts are rated as less good exemplars of the category (e.g., a poorer example of a "ba") than tokens where the visual and auditory stimuli specify the same syllable (Brancazio & Miller, 2005). Poorer exemplars of a category could lead to decision-level difficulties in executive functioning (potentially problematic for perceivers, and an established area of weakness for those with ASD; Eigsti, 2011).

To overcome these limitations, we have begun to assess the influence of visible articulatory information on heard speech with a novel measure that involves neither noise nor auditory and visual category conflict (also see Jerger, Damian, Tye-Murray, & Abdi, 2014). This new paradigm uses restoration of weakened auditory tokens with visual stimuli. There are two types of stimuli presented to the listener: clear exemplars of an auditory token (e.g., /ba/) and reduced tokens in which the auditory cues for the consonant are substantially weakened so that the consonant is not detected (reduced /ba/, heard as /a/). The auditory stimuli are created by synthesizing speech based on a natural production of /ba/ and systematically flattening the formant transitions to create the reduced /ba/. Video of the speaker's face does not change (always producing /ba/), but the auditory stimuli (/ba/ or "/a/") vary. Thus, in this example, when the "/a/" stimulus is dubbed with the visual /ba/, a visual influence will result in effectively "restoring" the weakened auditory cues so that the stimulus is perceived as a /ba/, akin to a visual phonemic restoration effect (Kashino, 2006; Samuel, 1981; Warren, 1970). Notably, in this design, the visual information for the same phoneme /ba/ supplements insufficient auditory information to assess the influence of visual information on what is heard without overt intersensory conflict (Brancazio et al., 2015). While currently used with CV stimuli, this method could be adapted to lexical or word-level stimuli. This method has been used to collect behavioral and ERP data to assess visual influence on what is heard (Brancazio et al., 2015; Irwin, Brancazio, Turcios, Avery, & Landi, 2015). Behavioral data using the new method with typically developing adults demonstrated visual influence for the reduced /ba/ signal (Brancazio et al., 2015), with significantly stronger ratings of /ba/ category exemplars in the presence of visual information, what Brancazio et al. (2015) call a within-category strengthening effect. Further, a more robust visual effect for the reduced acoustic signal than with no acoustic signal (visual only) using an AX same–different discrimination task was observed.

To examine neural signatures (ERP) of AV processing, an oddball paradigm is being used to examine mismatch negativity response to the infrequently presented reduced /ba/ (deviant) embedded within the more frequently occurring intact /ba's/ (standard). As with the behavioral design, each token is paired with a face producing /ba/. In this paradigm, we expect a larger mismatch negativity response for the reduced /ba/ in individuals who do not have a strong visual influence on what is heard. This passive design can be used to assess sensitivity to visual influence in children who cannot or do not reliably respond verbally, including lower functioning children with ASD (Irwin et al., 2015; Sorcinelli et al., 2013; Turcios, Gumkowski, Brancazio, Landi, & Irwin, 2014).

A goal of this new method is to leverage AV speech to reveal processing of this critical signal in both typically developing perceivers and in clinical populations. This may have important implications for everyday life, such as understanding speech in noisy environments like classrooms, cafeterias, and playgrounds. Further, results from well-controlled empirical studies may guide specific interventions.

8 | SUMMARY

This selected overview of AV speech perception examined the influence of visible articulatory information on heard speech. Thought to be a cross-cultural phenomenon that emerges early in typical language development, variables that influence AV speech perception include properties of the visual and the auditory signal, attention and individual differences. A review of the existing neurobiological evidence on how visual information influences heard speech reveals potential loci, timing, and facilitatory effects. Finally, a new method of assessing AV speech that does not require obvious cross-category mismatch or auditory noise was presented as a novel approach for investigators.

ENDNOTE

[1] Synchrony judgments for AV stimuli can occur within a window of audio and visual synchrony, with an asymmetry such that visual leading auditory stimuli are more likely to be detected as asynchronous than visual lead, Conrey and Pisoni (2006).

REFERENCES

Alcántara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noiseperception in high-functioning individuals with autism or Asperger's syndrome. *The Journal of Child Psychology and Psychiatry*, *45*(6), 1107–1114.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839–843.

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.

Barenholtz, E., Mavica, L., & Lewkowicz, D. J. (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition*, 100–105.

Barutchu, A., Crewther, S. G., Kiely, P., Murphy, M. J., & Crewther, D. P. (2008). When/b/ill with/g/ill becomes/d/ill: Evidence for a lexical effect in audiovisual speech perception. *European Journal of Cognitive Psychology*, *20*(1), 1–11.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *The Journal of Neuroscience*, *30*(7), 2414–2417.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26.

Bergeson, T. R., & Pisoni, D. B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. *The handbook of multisensory processes*, 749–772.

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. (2003). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *The Volta Review*, *103*(4), 347.

Bernstein, J. G., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *125*(5), 3358–3372.

Bernstein, L. E., Tucker, P. E., & Demorest, M. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, *62*(2), 233–252.

Bernstein, L. E., Auer, E. T., Wagner, M., & Ponton, C. W. (2008). Spatiotemporal dynamics of audiovisual speech processing. *NeuroImage*, *39*(1), 423–435.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, *7*(3), 295–301.

Blumstein, S. E., & Myers, E. B. (2013). Neural systems underlying speech perception. K. Ochsner and S. Kosslyn (Eds.), *Oxford Handbook of Cognitive Neuroscience, Volume I*. New York: Oxford University Press.

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 445.

Brancazio, L., & Miller, J. L. (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. *Perception & Psychophysics*, *67*(5), 759–769.

Brancazio, L., Best, C. T., & Fowler, C. A. (2006). Visual influences on perception of speech and nonspeech vocal-tract events. *Language and Speech*, *49*(1), 21–53.

Brancazio, L., Moore, D., Tyska, K., Burke, S., Cosgrove, D., & Irwin, J. (2015, May). McGurk-like effects of subtle audiovisual mismatch in speech perception. Presented at the 27th annual convention of the Association for Psychological Science, New York, May 23, 2015.

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2012). The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral Cortex*, *23*, 1329–1341. doi:10.1093/cercor/bhs109

Buchan, J. N., Paré, M., & Munhall, K. G. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research*, *1242*, 162–171.

Burnham, D., & Dodd, B. (1996). Auditory-visual speech perception as a direct process: The McGurk effect in infants and across languages. In *Speechreading by humans and machines* (pp. 103–114). Berlin Heidelberg: Springer.

Burnham, D., & Dodd, B. (1998). Familiarity and novelty in infant cross-language studies: Factors, problems, and a possible solution. *Advances in Infancy Research*, *12*, 170–187.

Burnham, D., & 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.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport*, *14*(17), 2213–2218.

Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience*, *16*(5), 805–816.

Calvert, G. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, 1110–1123. doi:10.1093/cercor/11.12.1110

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., … David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649–657.

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and non-speech signals. *The Journal of the Acoustical Society of America*, *119*(6), 4065–4073.

Davis, C., & Kim, J. (2006). Audio-visual speech perception off the top of the head. *Cognition*, *100*(3), B21–B31.

De Gelder, B. D., Vroomen, J., & Van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*(1), 69–86.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, *9*(2), 133–150.

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45*(4), 187–203.

Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, *66*(1), 85–110.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121–144.

Eigsti, I. (2011). Executive functions in ASD. *The Neuropsychology of Autism*, 185–203.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Erber, N. P. (1975). Auditory-visual perception of speech. *The Journal of Speech and Hearing Disorders*, *40*, 481–492.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207).

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3), 1197–1208.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America*, *103*(5), 2677–2690.

Green, K. (1994). The influence of an inverted face on the McGurk effect. *The Journal of the Acoustical Society of America*, 3014 .https://doi.org/10.1121/1.408802

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics*, *50*(6), 524–536.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Hockley, N. S., & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *The Journal of the Acoustical Society of America*, *96*(5), 3309–3309.

Irwin, J. R., & Brancazio, L. (2014). Seeing to hear? Patterns of gaze to speaking faces in children with autism spectrum disorders. *Frontiers, Language Sciences*. doi:10.3389/fpsyg.2014.00397

Irwin, J. R., Whalen, D. H., & Fowler, C. A. (2006). A sex difference in visual influence on heard speech. *Perception & Psychophysics*, *68*(4), 582–592.

Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011). Can children with autism spectrum disorders "hear" a speaking face? *Child Development*, *82*(5), 1397–1403.

Irwin, J. R., Frost, S. J., Mencl, W. E., Chen, H., & Fowler, C. A. (2011). Functional activation for imitation of seen and heard speech. *Journal of Neurolinguistics*, *24*(6), 611–618.

Irwin, J.R., Preston, J.L., Brancazio, L., D'Angelo, M., and Turcios, J. (2014). Development of an audiovisual speech perception app for children with autism spectrum disorders. *Clinical Linguistics & Phonetics* doi:10.3109/02699206.2014.966395

Irwin, J., Brancazio, L., Turcios, J., Avery, T., & Landi, N. (2015, October). A method for assessing audiovisual speech integration: Implications for children withASD. Presented at the meeting of the Society for Neurobiology of Language, Chicago, Illinois.

Jerger, S., Damian, M. F., Tye-Murray, N., & Abdi, H. (2014). Children use visual speech to compensate for non-intact auditory speech. *Journal of Experimental Child Psychology*, *126*, 295–312.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport*, *14*(8), 1129–1133.

Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *46*(2), 390–404.

Kashino, M. (2006). Phonemic restoration: The brain creates missing speech sounds. *Acoustical Science and Technology*, *27*(6), 318–321.

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophisiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, *18*, 65–75.

Knowland, V. C. P., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. C. (2014). Audio-visual speech perception: A developmental ERP investigation. *Developmental Science*, *17*(1), 110–124.

Knowland, V. C. P., Evans, S., Snell, C., & Rosen, S. (2016). Visual speech perception in children with language learning impairments. *Journal of Speech, Language, and Hearing Research*, *59*, 1–14. doi:10.1044/2015_JSLHR-S-14-0269

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Kuhl, P. K., Tsuzaki, M., Tohkura, Y., & Meltzoff, A. N. (1994). Human processing of auditory-visual information in speech perception: potential for multimodal human–machine interfaces. In Acoustical Society of Japan (Ed.), *Proceedings of the International Conference of Spoken Language Processing* (pp. 539–542). Tokyo: the Acoustical Society of Japan.

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences*, *105*(32), 11222–11445.

Lachs, L., & Pisoni, D. B. (2004). Specification of cross-modal source information in isolated kinematic displays of speech. *The Journal of the Acoustical Society of America*, *116*(1), 507–518.

Lachs, L., Pisoni, D. B., & Kirk, K. I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear and Hearing*, *22*(3), 236.

Lalonde, K., & Holt, R. F. (2014). Audiovisual speech integration development at varying levels of perceptual processing. *The Journal of the Acoustical Society of America*, *136*(4), 2263–2263.

Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, *65*(4), 536–552.

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior & Development*, *13*(3), 343–354.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychological Bulletin*, *126*(2), 281.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*(3), 253–257.

MacDonald, J., Andersen, S., & Bachmann, T. (2000). Hearing by eye: How much spatial degradation can be tolerated? *Perception*, *29*(10), 1155–1168.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131–141.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., … Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: An event related fMRI study. *Neuroreport*, *11*(8), 1729–1733.

Magnée, M. J., De Gelder, B., Van Engeland, H., & Kemner, C. (2008). Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *Journal of Child Psychology and Psychiatry*, *49*(9), 995–1000.

Magnotti, J. F., Mallick, D. B., Feng, G., Zhou, B., Zhou, W., & Beauchamp, M. S. (2015). Similar frequency of the McGurk effect in large samples of Mandarin Chinese and American English speakers. *Experimental Brain Research*, *233*(9), 2581–2586.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development*, 1777–1788.

Massaro, D. W. (1998). *Perceiving Talking Faces*. Cambridge, MA: MIT Press.

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, *41*(1), 93–113.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

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

Meltzoff, A. N., & Kuhl, P. K. (1994). Faces and speech: Intermodal processing of biologically relevant signals in infants and adults. In *The development of intersensory perception: Comparative perspectives* (pp. 335–369).

Ménard, L., Dupont, S., Baum, S. R., & Aubin, J. (2009). Production and perception of French vowels by congenitally blind adults and sighted adults. *The Journal of the Acoustical Society of America*, *126*(3), 1406–1414.

Meronen, A., Tiippana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*, *56*(1), 211–221.

Miller, L. M., & D'esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *The Journal of Neuroscience*, *25*(25), 5884–5893.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*, *14*(1), 115–128.

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(7), 1349–1358.

Moody-Antonio, S., Takayanagi, S., Masuda, A., Auer, E. T. Jr., Fisher, L., & Bernstein, L. E. (2005). Improved speech perception in adult congenitally deafened cochlear implant recipients. *Otology & Neurotology*, *26*(4), 649–654.

Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: A magnetoencephalographic study. *Neuroscience Letters*, *363*(2), 112–115.

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., Kroos, C., & Vatikiotis-Bateson, E. (2002). Audiovisual perception of band-pass filtered faces. *Journal of the Acoustical Society of Japan*, *21*, 519–520.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility head movement improves auditory speech perception. *Psychological Science*, *15*(2), 133–137.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage*, *59*(1), 781–787.

Nath, A., Fava, E. E., & Beauchamp, M. S. (2011). Neural correlates of interindividual differences in children's audiovisual speech perception. *The Journal of Neuroscience*, *31*(39), 13963–13971.

Norrix, L. W., Plante, E., Vance, R., & Boliek, C. A. (2007). Auditory-visual integration for speech by children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1639–1651.

Okada, K., Venezia, J. H., Matchin, W., Saberi, K., & Hickok, G. (2013). An fMRI study of audiovisual speech perception reveals multisensory interactions in auditory cortex. *PloS One*, *8*(6) e68959.

Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio–visual information processing in the human cerebral cortex. *Cognitive Brain Research*, *14*(1), 129–138.

Paré, M., Richler, R. C., ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics*, *65*(4), 553–567.

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *95*(3), 1581–1592.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, *32*(4), 249–261.

Pilling, M. (2009). Auditory event related potentials (ERP's) in audiovisual speech perception. *Journal of Speech, Language, and Hearing Research*, *52*, 1073–1081.

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. 97–114).

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*(6), 405–409.

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.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*(3), 347–357.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*, 1147–1153.

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, *33*(12), 2329–2337.

Russo, N., Foxe, J. J., Brandwein, A., Altschuler, T., Gomes, H., & Molholm, S. (2010). Multisensory processing in children with autism: High-density electrical mapping of auditory-somatosensory integration. *Autism Research*, *3*(5), 253–267.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*(3), 587–597.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *127*(1), 141–145.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474–494.

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18748–18750.

Schwartz, J. L. (2010). A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. *The Journal of the Acoustical Society of America*, *127*(3), 1584–1594.

Schwartz, J. L., Bertommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, *93*, B69–B78.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, *59*(1), 73–80.

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, *11*(2), 306–320.

Sekiyama, K., & Tokura, Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*, 427–444.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*(10), 2387–2399.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, *48*(8), 813–821.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*, *26*(3), 263–275.

Sorcinelli, A., Irwin, J., Gumkowski, N., Brancazio, L., Preston, J., & Landi, N. (2013, May). Diminished audiovisual speech integration for children with autism spectrum disorders. Poster presented at the 25th Association for Psychological Science Annual Convention, Washington, D.C., USA.

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

Spehar, B., Tye-Murray, N., & Sommers, M. (2004). Time-compressed visual speech and age: A first report. *Ear and Hearing*, *25*(6), 565–572.

Stephens, J. D., & Holt, L. L. (2010). Learning to use an artificial visual cue in speech identifications. *The Journal of the Acoustical Society of America*, *128*(4), 2138–2149.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215.

Szycik, G. R., Tausche, P., & Münte, T. F. (2008). A novel approach to study audiovisual integration in speech perception: Localizer fMRI and sparse sampling. *Brain Research*, *1220*, 142–149.

Taylor, N., Isaac, C., & Milne, E. (2010). A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. *Journal of Autism and Developmental Disorders*, *40*(11), 1403–1411.

Thomas, S. M., & Jordan, T. R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 873.

Tiippana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, *16*(3), 457–472.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*(2), 79–90.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and non-speech audiovisual illusions: A developmental study. *PLoS One*, *2*(8), 742.

Turcios, J., Gumkowski, N., Brancazio, L., Landi, N., & Irwin, J. (2014, April). Patterns of gaze to speaking faces in individuals with autism spectrum disorders and typical development. Poster presented at the 5th Annual University of Connecticut Language Festival, Storrs, CT.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear and Hearing*, *28*(5), 656–668.

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.

Vatikiotis-Bateson, E., Eigsti, I. M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, *60*(6), 926–940.

Walker, S., Bruce, V., and O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics*, *57*(8), 1124-1133.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*, 49–63.

Williams, J. H., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual–auditory integration during speech imitation in autism. *Research in Developmental Disabilities*, *25*(6), 559–575.

Windmann, S. (2004). Effects of sentence context and expectation on the McGurk illusion. *Journal of Memory and Language*, *50*(2), 212–230.

Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *24*(5), 603–612.

Yi, A., Wong, W., & Eizenman, M. (2013). Gaze patterns and audiovisual speech enhancement. *Journal of Speech, Language, and Hearing Research*, *56*(2), 471–480.

AUTHOR BIOGRAPHIES

Dr. Julia Irwin is an Associate Professor at Southern Connecticut State University and Director of the Language and Early Assessment Research Network (LEARN) Center at Haskins Laboratories. Her research focuses on audiovisual speech processing in typical and clinical populations using multiple methodologies, including eye tracking and EEG/ERP.

Ms. Lori DiBlasi is a Masters student in Psychology at Southern Connecticut State University and a student researcher at Haskins Laboratories. Her research focuses on promoting prereading skills in children in kindergarten using a phonological awareness–based approach.

How to cite this article: Irwin J, DiBlasi L. Audiovisual speech perception: A new approach and implications for clinical populations. *Lang Linguist Compass*. 2017;11:e12237. https://doi.org/10.1111/lnc3.12237