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, Melean, & 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” (Alsìus, Navarra, Campbell, & Soto-Faraco, 2005; Brancacio, Best, & Fowler, 2006; Brancacio & Miller, 2005; Green, 1994; MacDonald & McGurk, 1978; Rosenblum, 2008; Soto-Faraco & Alsìus, 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). 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 magnetoencephalography 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, Mencel, 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; Szycki, 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 with ASD, who may have a diminished or disrupted AV processing system (Brancazio et al., 2015).

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.
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


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).


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