JASA FORUM

Audiovisual speech perception: Moving beyond McGurk[a)]

Kristin J. Van Engen,[1,b)] Avanti Dey,[2] Mitchell S. Sommers,[1] and Jonathan E. Peelle[3,c)]
[1]*Department of Psychological and Brain Sciences, Washington University, St. Louis, Missouri 63130, USA*
[2]*PLOS ONE, 1265 Battery Street, San Francisco, California 94111, USA*
[3]*Department of Otolaryngology, Washington University, St. Louis, Missouri 63130, USA*

ABSTRACT:
Although it is clear that sighted listeners use both auditory and visual cues during speech perception, the manner in which multisensory information is combined is a matter of debate. One approach to measuring multisensory integration is to use variants of the McGurk illusion, in which discrepant auditory and visual cues produce auditory percepts that differ from those based on unimodal input. Not all listeners show the same degree of susceptibility to the McGurk illusion, and these individual differences are frequently used as a measure of audiovisual integration ability. However, despite their popularity, we join the voices of others in the field to argue that McGurk tasks are ill-suited for studying real-life multisensory speech perception: McGurk stimuli are often based on isolated syllables (which are rare in conversations) and necessarily rely on audiovisual incongruence that does not occur naturally. Furthermore, recent data show that susceptibility to McGurk tasks does not correlate with performance during natural audiovisual speech perception. Although the McGurk effect is a fascinating illusion, truly understanding the combined use of auditory and visual information during speech perception requires tasks that more closely resemble everyday communication: namely, words, sentences, and narratives with congruent auditory and visual speech cues.
 https://doi.org/10.1121/10.0015262
(Received 28 March 2022; revised 26 October 2022; accepted 5 November 2022; published online 2 December 2022)
[Editor: Matthew B. Winn] Pages: 3216–3225

I. INTRODUCTION

Speech perception in face-to-face conversations is a prime example of multisensory integration: listeners have access not only to a speaker's voice, but to visual cues from their face, gestures, and body posture. More than 45 years ago, McGurk and MacDonald (1976) published a remarkable (and now famous) example of visual influence on auditory speech perception: when an auditory stimulus (e.g., /ba/) was presented with the face of a talker articulating a different syllable (e.g., /ga/), listeners often experienced an illusory percept distinct from both sources (e.g., /da/).[1] Since that time, McGurk stimuli have been used in countless studies of audiovisual integration in humans (not to mention the multitude of classroom demonstrations on multisensory processing) (Marques *et al.*, 2016). At the same time, the stimuli typically used to elicit a McGurk effect differ substantially from what we usually encounter in conversation. In this paper, we consider how to best investigate the benefits listeners receive from being able to see a speaker's face while listening to their speech during natural communication. We start by reviewing behavioral findings regarding audiovisual speech perception and some theoretical constraints on our understanding of multisensory integration. With this background, we examine the McGurk effect to assess its usefulness for furthering our understanding of audiovisual speech perception, joining the voices of other speech scientists who argue that it is time to move beyond McGurk (Alsius *et al.*, 2017; Getz and Toscano, 2021; Massaro, 2017).

II. BENEFITS OF AUDIOVISUAL SPEECH COMPARED TO AUDITORY-ONLY SPEECH

A great deal of research in audiovisual speech processing has focused on the finding that speech signals are consistently more intelligible when both auditory and visual information is available compared to auditory-only (or visual-only) perception. Advantages in multisensory processing most obviously come from complementarity between auditory and visual speech signals. For example, /m/ and /n/ may sound very similar, but are visually distinct; similarly, /p/ and /b/ are nearly identical visually but can be distinguished more easily with the auditory signal. Having two modalities of speech information makes it more likely a listener will perceive the intended item. However, advantages may also come in cases when information is redundant across two modalities, particularly in the context of background noise. That is, auditory-visual (AV) (as opposed to unimodal) presentations provide listeners with two opportunities to perceive the intended information.

In experiments that present speech to listeners in noisy backgrounds, recognition of AV speech is significantly better

[a)]This paper is part of a special issue on Reconsidering Classic Ideas in Speech Communication.
[b)]Electronic mail: kvanengen@wustl.edu
[c)]Also at: Center for Cognitive and Brain Health, Department of Communication Sciences and Disorders, and Department of Psychology, Northeastern University, Boston, MA 02115, USA.

than recognition of auditory-only speech (Erber, 1975; Sommers *et al.*, 2005; Tye-Murray *et al.*, 2007b; Van Engen *et al.*, 2014; Van Engen *et al.*, 2017), and listeners are able to reach predetermined performance levels at more difficult signal-to-noise ratios (SNRs) (Grant and Seitz, 2000; Macleod and Summerfield, 1987; Sumby and Pollack, 1954). In addition, AV speech has been shown to speed performance in shadowing tasks (where listeners repeat a spoken passage in real time) (Reisberg *et al.*, 1987) and improve comprehension of short stories (Arnold and Hill, 2001). Furthermore, given that acoustically challenging speech is also associated with increased cognitive challenge (Peelle, 2018), visual information may reduce the cognitive demand associated with speech-in-noise processing (Gosselin and Gagné, 2011) (but see Brown and Strand, 2019). Below we review two key mechanisms by which audiovisual speech benefits listeners relative to unimodal speech.

A. Phonetic discrimination

Visual speech conveys information about the position of a speaker's articulators, which can help a listener identify speech sounds by distinguishing their place of articulation. For example, labial articulations (such as bilabial consonants or rounded vowels) are visually salient, so words like "pack" and "tack"—which may be difficult to distinguish in a noisy, auditory-only situation—can be readily distinguished if a listener can see the speaker's lips.

In addition to details about specific articulators, visual speech can provide information regarding durations of speech sounds. In English, for example, the difference between a voiceless /p/ and a voiced /b/ (phonemes that share a place of articulation) at the end of a word is largely instantiated as a difference in the duration of the preceding vowel (say the words "tap" and "tab" aloud and take note of the vowel duration in each—"tab" almost certainly contains a longer vowel). Visual speech can provide a critical durational cue for such phonemes, even if the auditory signal is completely masked by the acoustic environment.

Electroencephalography (EEG) and magnetoencephalography (MEG) studies, which can measure the degree to which evoked auditory responses are modulated by visual information, provide ample evidence in support of the rapid and predictive nature of visual speech. Davis *et al.* (2008), for example, used MEG to measure evoked responses to phonemes (/pi/, /ti/, /vi/) that were presented either auditory-only or audiovisually. They found that the magnitude of the N100m, an early marker of auditory processing (~100 ms post onset), was significantly reduced in the audiovisual condition relative to the auditory-only condition. These and other studies showing modulations of early evoked responses to speech are consistent with visual facilitatory effects on phonetic processing occurring within the first 100 ms of speech (Arnal *et al.*, 2009; van Wassenhove *et al.*, 2005).

One way to quantify the effects of such facilitation is in the context of a lexical competition framework. Lexical competition frameworks start from the assumption that listeners must select the appropriate target word from among a set of similar-sounding words (phonological neighbors), which act as competitors. Words with a relatively high number of phonological neighbors (such as "cat") thus have higher levels of lexical competition and may thus rely more on cognitive processes of inhibition or selection compared to words with relatively few phonological neighbors (such as "orange"). Originally developed for auditory-only speech perception (Luce and Pisoni, 1998; Marslen-Wilson and Tyler, 1980), lexical competition frameworks have since been extended to consider AV speech perception (Feld and Sommers, 2011; Strand, 2014). Tye-Murray *et al.* (2007a) introduced the concept of intersection density, suggesting that it is only the words that overlap in both auditory and visual features that compete during word recognition. Specifically, they investigated whether auditory-only, visual-only, or AV neighborhoods predicted listeners' performance in AV speech perception tasks, and found that it was the combined AV neighborhood—reflecting the intersection of auditory and visual cues—that was most closely related to a listener's speech perception accuracy.

In summary, visual speech can provide informative cues to place of articulation and timing that are important for phonetic identification. Visual information can therefore reduce the number of lexical competitors during comprehension and increase accuracy of perception.

B. Temporal prediction

At a basic level, temporal information in the visual signal can also help listeners by simply serving as a cue that a person has begun speaking. In a noisy restaurant, for example, seeing a talker's mouth can help listeners direct their attention to the talker's speech and better segregate it from other interfering sound sources.

However, in addition to alerting listeners to the start of speech, mouth movements also provide ongoing information about the temporal envelope of the acoustic signal during continuous speech, with louder amplitudes being associated with opening of the mouth (Chandrasekaran *et al.*, 2009; Grant and Seitz, 2000; Summerfield, 1987). The visual signal thus provides clues about the rhythmic structure of connected speech, which helps listeners form expectations about incoming information and therefore supports linguistic processing. As discussed in more detail below, visual information regarding the acoustic envelope of sentences may aid in how ongoing brain oscillations track the incoming acoustic speech signal.

III. MULTISENSORY INTEGRATION

Although it is clear that visual information significantly aids auditory speech recognition, understanding speech perception requires determining *how* (or if) listeners combine the simultaneously-available information from visual and auditory modalities during communication. The issue is critical because if there is only a single mechanism for audiovisual integration, then any task (such as one using McGurk

stimuli) that involves integrative processes necessarily relates to other audiovisual tasks (such as everyday speech perception). On the other hand, if there are multiple ways in which information can be combined across modalities, it is critical to determine whether two tasks rely on the same integrative mechanism.

One way to classify different integrative mechanisms is by whether integration occurs at an early or late stage of processing (Peelle and Sommers, 2015): Early integration refers to modulations of activity in primary sensory cortex—that is, "early" in terms of both processing hierarchy and time. Late integration occurs in regions of heteromodal cortex [for example, posterior superior temporal sulcus (STS)], incorporating auditory and visual information into a unified percept only after unimodal processing has occurred. The possibility of complementary mechanisms for multisensory integration has implications for how we interpret the McGurk effect.

A. Late integration models of audiovisual integration

Many classical frameworks for audiovisual speech processing have focused on late integration (Grant *et al.*, 1998; Oden and Massaro, 1978), which makes sense under the intuitive assumption that auditory and visual cortex receive largely unimodal inputs. In other words, perceptual information is necessarily segregated at the level of sensory receptors, setting the stage for independent parallel processing of auditory and visual information. Under this view, auditory and visual information are processed separately and combined later in a region of the brain that has anatomical and functional connections to the relevant sensory cortices. For example, if unimodal auditory processing occurs in the auditory cortex, and visual processing in the visual cortex, then the posterior superior temporal sulcus (STS) is well-positioned to receive input from both modalities and combine this information into a unified percept.

If integration indeed occurs after unimodal processing, it is straightforward to postulate independent abilities for auditory-only speech processing, visual-only speech processing, and audiovisual integration. Separating audiovisual integrative ability is of theoretical benefit in explaining the performance of listeners who are matched in unimodal processing, but differ in their success with audiovisual tasks. For example, Tye-Murray *et al.* (2016) matched individuals on auditory-only and visual-only performance by using individually determined levels of babble noise and visual blur to obtain 30% accuracy in both unimodal conditions. Principal component analysis identified separate auditory and visual factors, but no additional factor related to integration: that is, after accounting for individual differences in auditory-only and visual-only performance, AV performance was consistent from 22 to 92 years of age. Subsequent work showed that if unimodal performance is taken into account, knowing the age of participants does not explain additional variance (Myerson *et al.*, 2021)—again, arguing against the need for an "integration" ability to explain age-related differences in AV speech perception.

The posterior STS is a region of heteromodal cortex that shows responses to both auditory and visual inputs, making it an appealing candidate to underlie late audiovisual integration (Beauchamp, 2005). We consider the posterior STS to be involved in late integration, as it receives inputs from primary sensory regions. Consistent with an integrative role, regions of middle and posterior STS show activity for AV speech relative to unimodal speech (Venezia *et al.*, 2017). Nath and Beauchamp (2011) offered a particularly nice demonstration of audiovisual integration by investigating functional connectivity between auditory regions, visual regions, and posterior STS while participants were presented with syllables and words. Importantly, they varied the accuracy of unimodal information by either blurring the visual speech signal ("auditory reliable") or putting the speech in noise ("visual reliable"). They found that functional connectivity between sensory cortices and posterior STS changed as a function of unimodal clarity such that when auditory information was clear, connectivity between auditory regions and posterior STS was stronger than when visual information was clear. These functional imaging results are consistent with the performance of patients with brain damage, who are most strongly impaired when regions of STS are damaged (Hickok *et al.*, 2018), and the finding that transcranial magnetic stimulation to posterior STS interferes with the McGurk illusion (Beauchamp *et al.*, 2010). Together, these findings support the posterior STS playing an important role in audiovisual integration and suggest that it may weigh modalities as a function of their informativeness.

B. Early integration models of audiovisual integration

Despite their intuitive appeal, significant challenges for strict late integration models come from a variety of perspectives. For example, electrophysiological research in nonhuman primates shows that neural activity in the primary auditory cortex is modulated by multisensory cues, including from motor/haptic (Lakatos *et al.*, 2007) and visual (Lakatos *et al.*, 2008) inputs. In humans, regions of posterior STS also represent visual amplitude envelope information (Micheli *et al.*, 2018). Non-auditory information may arrive at primary auditory cortex through nonspecific thalamic pathways (Schroeder *et al.*, 2008) and/or cortico-cortical connections from visual to auditory cortex (Arnal and Giraud, 2012). Multisensory effects in primary auditory regions argue against late integration models because they suggest that, at least by the time information reaches cortex, there are no longer pure unisensory representations.

Research on the role of brain oscillations in speech perception also supports an early integration account for audiovisual speech (Giraud and Poeppel, 2012; Peelle and Davis, 2012). Cortical oscillations entrain to acoustic information in connected speech (Luo and Poeppel, 2007), an effect that is enhanced when speech is intelligible (Peelle *et al.*, 2013).

https://doi.org/10.1121/10.0015262

This cortical entrainment may be thought of as encoding a sensory prediction because it relates to the times at which incoming information is likely to occur. Interestingly, visual information increases the entrainment of oscillatory activity in the auditory cortex to the amplitude envelope of connected speech in both quiet and noisy listening environments (Crosse *et al.*, 2015; Luo *et al.*, 2010; Zion Golumbic *et al.*, 2013). Visual influences on auditory processing do not simply reflect feedback from later multisensory regions but affect processing in real time (Atilgan *et al.*, 2018; Maddox *et al.*, 2015; Schroeder and Foxe, 2005). These findings suggest a role for audiovisual integration *within the* auditory cortex. Because oscillatory entrainment is a phenomenon associated with connected speech (that is, sentences or stories) rather than isolated words or syllables, these findings raise the possibility that shorter stimuli, such as the single syllables typically used in McGurk tasks, might fail to engage this type of early integrative process.

IV. THE MCGURK EFFECT

On the surface, the McGurk effect appears to be an ideal measure of auditory-visual integration: discrepant inputs from two modalities (auditory and visual) are combined to produce a percept that can be distinct from either of the unimodal inputs. Video demonstrations of the McGurk effect often ask viewers to open and close their eyes, with the percept changing instantaneously depending on the availability of visual input. In addition, the effect is robust: auditory judgments are affected by visual information even when listeners are aware of the illusion, when male voices are dubbed onto female faces (and vice versa) (Green *et al.*, 1991) when auditory information lags behind the visual signal by as much as 300 ms (Munhall *et al.*, 1996; Soto-Faraco and Alsius, 2009; Venezia *et al.*, 2016), and for point light displays (Rosenblum and Saldana, 1996). As illustrated in Fig. 1(a), individual observers differ widely in their susceptibility to the McGurk illusion: some hardly ever experience it, whereas others experience it often.

In addition to demonstrating a combination of auditory and visual information, the McGurk effect is a useful tool in that it is observable across a range of populations: prelinguistic infants (Rosenblum *et al.*, 1997), young children (Massaro *et al.*, 1986), speakers of different languages (Magnotti *et al.*, 2015; Sekiyama, 1997), and individuals with Alzheimer's disease (Delbeuck *et al.*, 2007), among others, have all been shown to be susceptible to the illusion. Because susceptibility to the McGurk effect varies considerably across individuals even within these groups, it has been used as an index of integrative ability: individuals who show greater susceptibility to the McGurk effect are typically considered to be better integrators (Magnotti and Beauchamp, 2015; Stevenson *et al.*, 2012). Finally, the McGurk effect appears to be quite reliable: Strand *et al.* (2014) and Basu Mallick *et al.* (2015) also both showed high test-retest reliability in the McGurk effect [even, in the case of Basu Mallick *et al.*, 2015, with a year between measurements, Fig. 1(b)], suggesting that these tasks measure a stable trait within individual listeners.

Neuroanatomically, as noted above, functional brain imaging studies frequently find activity in left posterior STS associated with McGurk stimuli. Given the heteromodal anatomical connectivity of these temporal regions and their response to a variety of multisensory stimuli, this brain region makes intuitive sense as playing a key role in audiovisual integration.

V. CONCERNS ABOUT THE MCGURK EFFECT

There is no question that the McGurk effect demonstrates a combination of auditory and visual information, is a reliable within-listener measure, and reveals considerable

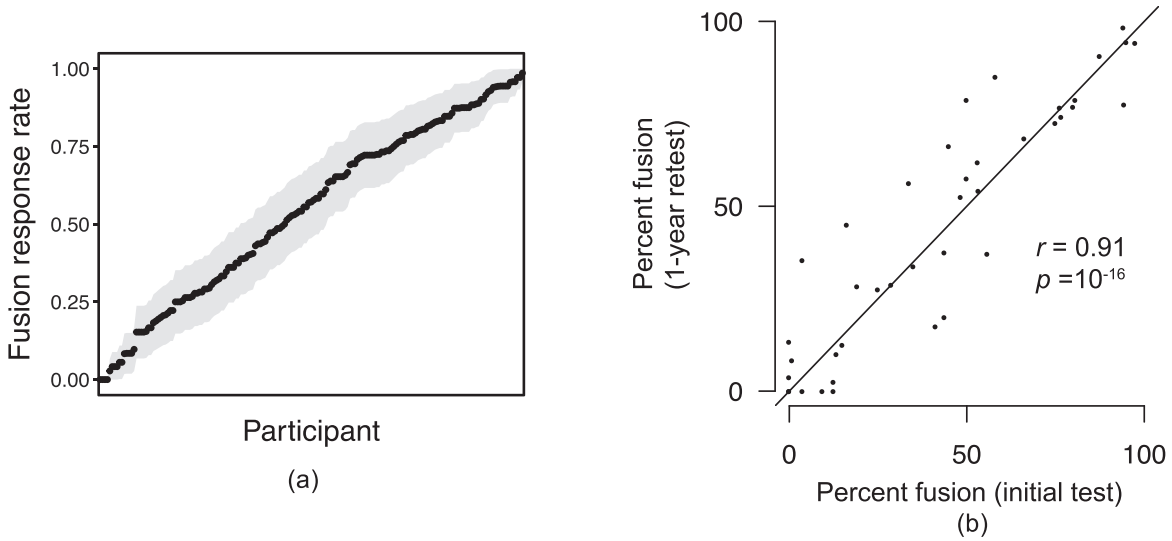

FIG. 1. (Color online) (a) Individual variability in susceptibility to the McGurk effect (Brown *et al.*, 2018). Shaded regions represent two standard errors for each participant's fusion rate. (b) Reliability of individual susceptibility to the McGurk effect on separate testing occasions separated by one year (Basu Mallick *et al.*, 2015).

TABLE I. Characteristics of McGurk stimuli and natural speech.

	McGurk stimuli	Natural speech
Audiovisual	x	x
Phonetic content	x	x
Audiovisual congruence		x
Phonological constraints		x
Lexical constraints		x
Semantic context		x
Syntactic context		x

variability among listeners.[2] However, the imperfect correspondence between McGurk stimuli and natural speech raises significant concerns about the face validity of comparing the two. Table I lists several characteristics of speech and whether these are present in either natural speech or in typical McGurk stimuli (i.e., isolated syllables presented with incongruent visual information and without additional linguistic context). Although there are certainly commonalities, there are also significant differences. Importantly, most standard McGurk stimuli lack the phonological, lexical, syntactic, and semantic context central to natural speech communication. Of course, the incongruent auditory and visual cues of the McGurk effect never occur in natural face-to-face speech (a speaker cannot produce speech incongruent with their vocal apparatus). This incongruence necessarily disrupts the phonetic and temporal correspondence between the visual signal and the auditory signal, putting inputs that are normally complementary and/or redundant in conflict with one another. Given the artificial nature of the stimuli typically used to elicit the McGurk effect, we move on to consider experimental evidence that can speak to how the McGurk effect relates to natural speech perception, arguing that there are good reasons to think the mechanisms of integration for McGurk stimuli and audiovisual speech processing may differ.

A. Differences between processing natural audiovisual speech and typical McGurk stimuli

Perhaps the most important difference between McGurk stimuli and everyday audiovisual speech signals is that the auditory and visual inputs in McGurk tasks are, by definition, incongruent. Resolving discrepant inputs is decidedly *not* what listeners are doing when they use congruent auditory and visual information to better understand spoken utterances. In an functional magnetic resonance imaging (fMRI) study, Erickson *et al.* (2014) showed distinct brain regions involved in the processing of congruent AV speech and incongruent AV speech when compared to unimodal speech (acoustic-only and visual-only). Left posterior STS was recruited during congruent bimodal AV speech, whereas left posterior superior temporal gyrus [posterior superior temporal gyrus (STG)] was recruited when processing McGurk speech, suggesting that left posterior STG may be especially important when there is a discrepancy between auditory and visual cues. Moris Fernandez *et al.* (2017) further showed regions of the brain associated with conflict detection and resolution (anterior cingulate cortex and left inferior frontal gyrus) were more active for incongruent than congruent speech stimuli. Using EEG, Crosse *et al.* (2015) have shown that neural entrainment to the speech envelope is inhibited when auditory and visual information streams are incongruent compared to congruent. On balance these studies suggest the processes supporting the perception of incongruent speech (including McGurk stimuli) differ from those engaged in the perception of congruent speech.[3]

Second, the near-exclusive use of single syllables for demonstrating the McGurk effect (with some exceptions; e.g., Dekle *et al.*, 1992) does not reflect the nature of AV integration in the processing of more natural linguistic stimuli such as words and sentences (Sams *et al.*, 1998; Windmann, 2004). Sommers *et al.* (2005), for example, found no correlation between AV perception of syllables and AV perception of words or sentences. Grant and Seitz (1998) similarly showed no correlation between integration measures derived from consonant vs sentence recognition. Although additional research is required, these results suggest that correlations between McGurk susceptibility for syllables, words, and sentences is likely to be similarly low, making it questionable whether integration based on syllable perception is closely related to integration in the processing of words or sentences. Indeed, Windmann (2004) showed that the clarity and likelihood of the McGurk effect was dependent on listeners' semantic and lexical expectations.

B. McGurk susceptibility and audiovisual speech perception

Perhaps the most compelling evidence against the use of the McGurk effect for understanding audiovisual integration in natural speech comprehension comes from studies that have investigated whether an individual listener's susceptibility to McGurk stimuli correlates with their audiovisual speech benefit (for example, the intelligibility of sentences in noise with and without access to the visual signal). Grant and Seitz (1998) showed equivocal results in this regard. For older adults with acquired hearing loss, McGurk susceptibility was correlated with visual enhancement (i.e., the proportion of available improvement listeners made relative to their audio-only performance) for sentence recognition, but McGurk susceptibility did not contribute significantly to a regression model predicting visual enhancement, suggesting McGurk susceptibility did not provide information above and beyond other, non-McGurk factors.

Van Engen *et al.* (2017) measured participants' susceptibility to the McGurk effect and their ability to identify sentences in noise (Fig. 2). Listeners were tested in both speech-shaped noise and in two-talker babble across a range of signal-to-noise ratios. For each condition, listeners were tested in both audio-only and audiovisual modalities. McGurk susceptibility did not predict performance overall, nor did it interact with noise level or modality. In fact, in speech-shaped noise there was a trend reflecting a negative association between McGurk susceptibility and speech

https://doi.org/10.1121/10.0015262

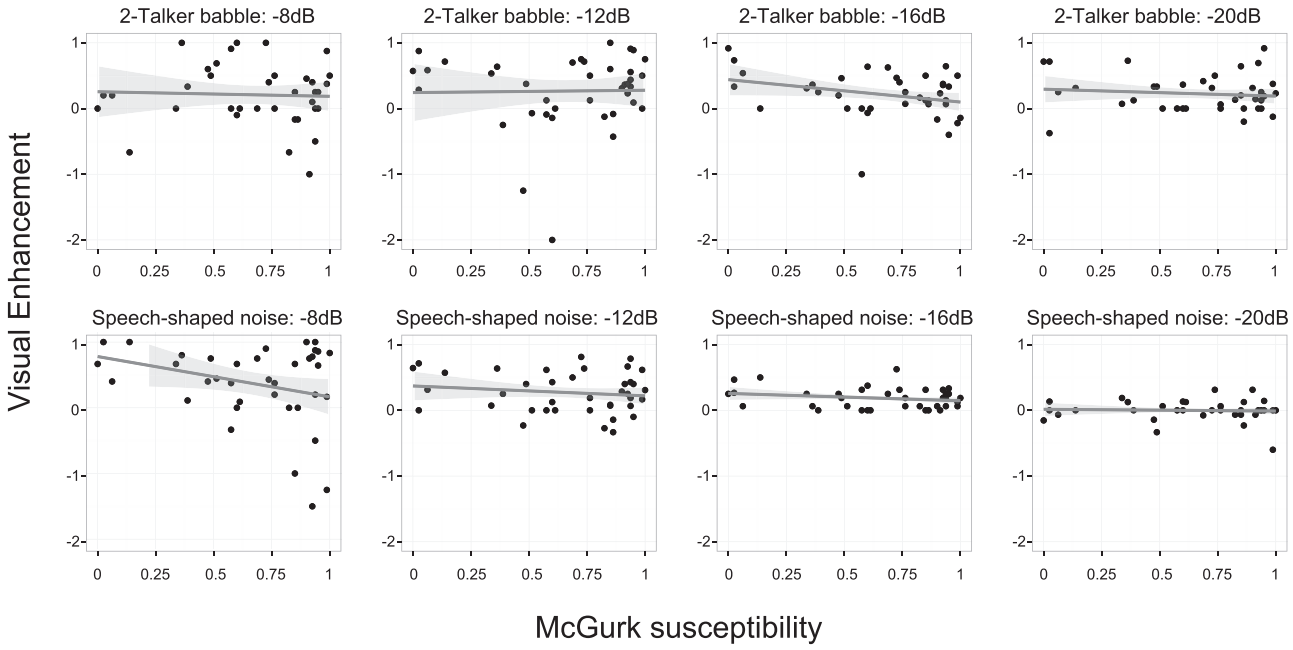

FIG. 2. (Color online) Visual enhancement (improved intelligibility for AV speech compared to auditory-only speech) during sentence comprehension does not depend on susceptibility to the McGurk effect across two types of background noise and several signal-to-noise ratios (Van Engen *et al.*, 2017).

recognition, particularly in the AV conditions. If anything, then, McGurk susceptibility may predict poorer performance on some AV speech tasks. The same study also failed to find any significant correlations between McGurk susceptibility and visual enhancement in any of the listening conditions. (Notably, the two conditions which showed marginally significant relationships again indicated that, if anything, greater susceptibility to the McGurk illusion was associated with lower rates of visual enhancement.) Similar results were reported by Hickok *et al.* (2018), who found no relationship between McGurk susceptibility and AV enhancement for phoneme perception. These empirical findings indicate that McGurk susceptibility is measuring something different from congruent audiovisual processing.

C. General problems for the McGurk task as a measure of audiovisual integration

In addition to concerns about whether McGurk tasks tap into the same mechanisms of integration that are relevant for speech communication, there are also more general concerns about the McGurk effect as a measure of integration (see also Alsius *et al.*, 2017). One recent study, for example, argues that McGurk-like responses may arise as default responses to ambiguous inputs as opposed to audiovisual fusions (Gonzales *et al.*, 2021). Getz and Toscano (2021), after demonstrating significant top-down influences on the illusion, argued that the McGurk effect should not be viewed as a robust perceptual illusion, but instead as an effect that exists only for some participants under specific task and stimulus conditions.

One factor complicating interpretation of group and individual differences in McGurk susceptibility is that such differences can also arise from variability in unimodal (auditory-only or visual-only) performance. Group and individual differences in unimodal encoding are particularly important given the extensive variability in visual-only (lip reading) abilities, even within homogeneous groups of participants. Therefore, even for normal-hearing young adults, individual differences in McGurk susceptibility could be attributed to differences in auditory-visual integration, visual-only encoding, or some combination of both. Strand *et al.* (2014) showed, for example, that McGurk susceptibility depends in part on differences in lipreading skill and detection of incongruity. Issues of individual differences in unimodal performance will be magnified in populations, such as older adults, where both auditory-only and visual-only encoding can vary substantially across individuals, further complicating interpretation of group differences in McGurk susceptibility as arising exclusively from differences in auditory-visual integration.

Given that individual differences in auditory-visual integration and unimodal encoding can contribute to measures of McGurk susceptibility, it is perhaps not surprising that individuals vary tremendously in how likely they are to experience the McGurk illusion: susceptibility to the McGurk effect has been shown to vary from 0% to 100% across individuals viewing the identical stimuli (Basu Mallick *et al.*, 2015; Strand *et al.*, 2014). Of perhaps even greater concern is that the distribution of responses for McGurk stimuli is often bimodal—most individuals will perceive the fused response from a given McGurk stimulus either less than 10% of the time or greater than 90% of the time. Basu Mallick *et al.* (2015), for example, found that across 20 different pairs of McGurk stimuli, 77% of participants either showed the fused response less than 10% of the time or greater than 90% of the time. The bimodal nature of

McGurk susceptibility is at odds with more graded responses to audiovisual speech perception—people vary in how much they benefit from visual speech information, but the distribution is not bimodal (e.g., Grant *et al.*, 1998; Grant and Seitz, 1998; Sommers *et al.*, 2005; Van Engen *et al.*, 2017). One likely reason for the extensive individual variability is that McGurk susceptibility reflects a combination of differences in auditory-visual integration and unimodal encoding.

Brown *et al.* (2018) investigated whether individual variability in McGurk susceptibility might be explained by differences in sensory or cognitive ability. Although lipreading ability was related to McGurk susceptibility, the authors failed to find any influence of attentional control (using a flanker task), lexical processing speed (using a lexical decision task), or working memory capacity (using an operation span task), even though these abilities are frequently related to auditory speech perception. The lack of any clear perceptual or cognitive correlates (outside of lipreading) of McGurk susceptibility suggests the integration operating during McGurk tasks is cognitively isolated from many factors involved in speech perception.

Studies of age differences in susceptibility to the McGurk effect also raise questions as to whether the illusion reflects the operation of mechanisms typically engaged in speech perception. Specifically, as noted, both age and individual differences in visual enhancement can be accounted for by differences in visual-only speech perception – not surprisingly, individuals who are better lipreaders typically exhibit greater visual enhancement (Tye-Murray *et al.*, 2016). Although older adults typically have reduced visual-only performance (Sommers *et al.*, 2005), however, studies have shown either equivalent (Cienkowski and Carney, 2002) or enhanced (Sekiyama *et al.*, 2014) susceptibility to the McGurk effect relative to young adults. In other words, the relationship between lipreading ability and audiovisual enhancement from congruent visual input appears to differ from the relationship between lipreading and McGurk susceptibility in this population. In keeping with this discrepancy, the Jones and Noppeney (2021) review article on aging and multisensory integration points out that older adults tend to benefit from congruent multisensory signals to a similar degree as younger adults, but that they consistently experience greater conflict from incongruent stimuli.

VI. MOVING BEYOND MCGURK

McGurk stimuli differ in a number of important properties from natural speech, and—perhaps as a result—we are not aware of empirical evidence relating McGurk susceptibility to audiovisual perception of natural speech. For these reasons we argue that McGurk stimuli are not well-suited for improving our understanding of audiovisual speech perception. What, then, is a better alternative?

If our goal is to better understand natural speech, we suggest that currently the best option is to use natural speech stimuli: in other words, stimuli that contain all of the speech cues listed in Table I. There are good precedents for using actual speech stimuli across multiple levels of language representation. For example, studies of word perception use real words that can vary in psycholinguistic attributes, including frequency, phonological neighborhood density, or imageability. Similarly, countless studies of sentence processing use sentences that vary in attributes including syntactic complexity, prosodic information, or semantic predictability. By using coherent stimuli that vary in their perceptual and linguistic properties, it has proven possible to learn a great deal about the dimensions that influence speech understanding. Studies of audiovisual speech perception would benefit from a similar approach: using natural speech samples in auditory-only, visual-only, and audiovisual modalities, all of which are encountered in real life conversation. By manipulating specific properties of these signals—for example, the degree to which auditory and visual information are complementary vs redundant—it is still possible to vary the audiovisual integration demands.

One way to differentially engage multisensory processing is to alter the clarity of the visual data. Tye-Murray *et al.* (2016) presented AV sentences; the video portion of the stimuli was blurred to manipulate the fidelity of the visual signal, resulting in poorer visual-only and AV performance. Critically, the authors found that auditory-only and visual-only abilities related to audiovisual performance, but age *per se* did not, and neither was there evidence for an "integration" ability. Many other examples of AV paradigms that avoid McGurk stimuli are available (Crosse *et al.*, 2015; Luo *et al.*, 2010; McLaughlin *et al.*, 2022; Okada *et al.*, 2013; Park *et al.*, 2016; Park *et al.*, 2018; Yi *et al.*, 2013; Yi *et al.*, 2014).

Of course, the clarity of the auditory signal can also be manipulated, for instance using background noise. One example from the neuroimaging literature is Peelle *et al.* (2022), who presented single AV words in different levels of background noise. They found that, compared to auditory-only speech, AV speech resulted in increased functional connectivity between auditory and visual cortex. These findings suggest that the temporal coordination of activity across brain regions may play a key role in multisensory speech processing. Support for the role of posterior superior temporal sulcus was somewhat equivocal, which highlights the importance of using ecologically-valid stimuli.

Finally, in addition to using stimuli that are real speech, computational modeling may also provide a fruitful way of understanding natural speech. Magnotti *et al.* (2020) use a model framework to analyze responses to McGurk stimuli on both a classic McGurk fusion task and a speech-in-noise task, as well as sentences in noise taken from Van Engen *et al.* (2017). They conclude that McGurk tasks and speech-in-noise understanding share some processes, but that speech-in-noise has different task requirements, making the lack of correlation of these two abilities across listeners not unexpected. Although we find the prospect of unifying responses to different stimuli within a computational framework promising, the lack of agreement of behavioral scores between McGurk and speech-in-noise tasks remains.

It is worth highlighting that the implications of how we assess audiovisual speech processing also have clinical implications. As reviewed by Irwin and DiBlasi (2017), AV processing (frequently assessed using McGurk stimuli) has been reported to differ in a number of clinical populations, including individuals with autism spectrum disorders, developmental language disorders, or hearing difficulty. However, whether McGurk-based differences are sufficient to identify possible interventions remains to be seen. It may be that stimuli closer to what we encounter in everyday conversation will be more useful in clinical settings.

VII. CONCLUSIONS

There is no question that the McGurk effect is a robust and compelling demonstration of multisensory integration: auditory and visual cues combine to form a fused percept. At the same time, there are many reasons to think it may not be an effective means to study the processes occurring during natural speech perception. Speech communication is centered around real words, usually involves connected speech, and involves auditory and visual inputs that are congruent, meaning they can provide complementary and/or redundant (rather than conflicting) information. To understand what occurs during speech perception, we join the voices of others (Alsius *et al.*, 2017; Getz and Toscano, 2021; Massaro, 2017; Rosenblum, 2019) in advocating that we move beyond McGurk. A more productive line of research requires that we focus on audiovisual speech comprehension in situations that are closer to real-life communication. If these results are consistent with those provided by McGurk stimuli, we can be more certain that both are indexing the same phenomena. However, if they disagree, we will need to consider the real possibility that the McGurk effect, although fascinating, does not inform our understanding of everyday speech perception.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. R01 DC014281, R01 DC019507, and R01 DC016594 from the National Institutes of Health. We are grateful to Violet Brown for helpful comments on prior drafts of this manuscript.

[1]Interested readers can find the story of the original discovery of the McGurk effect (which was an accident!) in the MacDonald (2018) article in Multisensory Research.

[2]Note that Getz and Toscano (2021) have recently argued that the McGurk effect should not be viewed as a robust perceptual illusion, but instead as an effect that exists only for some participants under specific task and stimulus conditions, and one that is subject to many top-down influences.

[3]Studies of spatial effects (i.e., the influence of the relative locations of the talker's face and their voice) and eye-position effects (i.e., whether or not people are looking at the visual part of an AV object) also reveal differences between typical AV speech and McGurk processing (Fleming *et al.*, 2021; Jones and Munhall, 1997; Kim and Davis, 2011; Siddig *et al.*, 2019).

Alsius, A., Paré, M., and Munhall, K. G. (**2017**). "Forty years after hearing lips and seeing voices: The McGurk effect revisited," Multisens. Res. **31**, 111–144.

Arnal, L. H., and Giraud, A.-L. (**2012**). "Cortical oscillations and sensory predictions," Trends Cogn. Sci. **16**, 390–398.

Arnal, L. H., Morillon, B., Kell, C. A., and Giraud, A.-L. (**2009**). "Dual neural routing of visual facilitation in speech processing," J. Neurosci. **29**(43), 13445–13453.

Arnold, P., and Hill, F. (**2001**). "Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact," Br. J. Audiol. **92**, 339–355.

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., and Bizley, J. K. (**2018**). "Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding," Neuron **97**(3), 640–655..

Basu Mallick, D., Magnotti, J. F., and Beauchamp, M. S. (**2015**). "Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type," Psychon. Bull. Rev. **22**(5), 1299–1307.

Beauchamp, M. S. (**2005**). "See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex," Curr. Opin. Neurobiol. **15**, 145–153.

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

Brown, V. A., Hedayati, M., Zanger, A., Mayn, S., Ray, L., Dillman-Hasso, N., and Strand, J. F. (**2018**). "What accounts for individual differences in susceptibility to the McGurk effect?," PLoS ONE **13**(11), e0207160.

Brown, V. A., and Strand, J. F. (**2019**). "About face: Seeing the talker improves spoken word recognition but increases listening effort," J. Cogn. **2**(1), 44.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (**2009**). "The natural statistics of audiovisual speech," PLoS Comput. Biol. ***5***, e1000436.

Cienkowski, K. M., and Carney, A. E. (**2002**). "Auditory-visual speech perception and aging," Ear Hear. **23**(5), 439–449.

Crosse, M. J., Butler, J. S., and Lalor, E. C. (**2015**). "Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions," J. Neurosci. **35**(42), 14195–14204.

Davis, C., Kislyuk, D., Kim, J., and Sams, M. (**2008**). "The effect of viewing speech on auditory speech processing is different in the left and right hemispheres," Brain Res. **1242**, 151–161.

Dekle, D. J., Fowler, C. A., and Funnell, M. G. (**1992**). "Audiovisual integration in perception of real words," Percept. Psychophys. **51**(4), 355–362.

Delbeuck, X., Collette, F., and Van der Linden, M. (**2007**). "Is Alzheimer's disease a disconnection syndrome? Evidence from a crossmodal audiovisual illusory experiment," Neuropsychologia **45**(14), 3315–3323.

Erber, N. P. (**1975**). "Auditory-visual perception of speech," J. Speech Hear. Disord. **40**, 481–492.

Erickson, L. C., Zielinski, B. A., Zielinski, J. E., Liu, G., Turkeltaub, P. E., Leaver, A. M., and Rauschecker, J. P. (**2014**). "Distinct cortical locations for integration of audiovisual speech and the McGurk effect," Front. Psychol. **5**, 534.

Feld, J., and Sommers, M. S. (**2011**). "There goes the neighborhood: Lipreading and the structure of the mental lexicon," Speech Commun. **53**, 220–228.

Fleming, J. T., Maddox, R. K., and Shinn-Cunningham, B. G. (**2021**). "Spatial alignment between faces and voices improves selective attention to audio-visual speech," J. Acoust. Soc. Am. **150**(4), 3085.

Getz, L. M., and Toscano, J. C. (**2021**). "Rethinking the McGurk effect as a perceptual illusion," Atten. Percept. Psychophys. **83**(6), 2583–2598.

Giraud, A.-L., and Poeppel, D. (**2012**). "Cortical oscillations and speech processing: Emerging computational principles and operations," Nat. Neurosci. **15**, 511–517.

Gonzales, M. G., Backer, K. C., Mandujano, B., and Shahin, A. J. (**2021**). "Rethinking the Mechanisms Underlying the McGurk Illusion," Front. Hum. Neurosci. ***15***, 616049.

Gosselin, P. A., and Gagné, J.-P. (**2011**). "Older adults expend more listening effort than younger adults recognizing audiovisual speech in noise," Int. J. Audiol. **50**, 786–792.

Grant, K. W., and Seitz, P. F. (**1998**). "Measures of auditory–visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am. **104**, 2438–2450.

Grant, K. W., and Seitz, P.-F. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am. **108**, 1197–1208.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.
Green, K. P., Kuhl, P. K., Meltzoff, A. N., and Stevens, E. B. (**1991**). "Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect," Percept. Psychophys. **50**(6), 524–536.
Hickok, G., Rogalsky, C., Matchin, W., Basilakos, A., Cai, J., Pillay, S., Ferrill, M., Mickelsen, S., Anderson, S. W., Love, T., Binder, J., and Fridriksson, J. (**2018**). "Neural networks supporting audiovisual integration for speech: A large-scale lesion study," Cortex **103**, 360–371.
Irwin, J., and DiBlasi, L. (**2017**). "Audiovisual speech perception: A new approach and implications for clinical populations," Lang. Linguist. Compass. **11**(3), 77–91.
Jones, J. A., and Munhall, K. G. (**1997**). "Effects of separating auditory and visual sources on audiovisual integration of speech," Can. Acoust. **25**(4), 13–19.
Jones, S. A., and Noppeney, U. (**2021**). "Ageing and multisensory integration: A review of the evidence, and a computational perspective," Cortex **138**, 1–23.
Kim, J., and Davis, C. (**2011**). "Audiovisual speech processing in visual speech noise," in *Proceedings of the AVSP 2011*, September 1–2, Volterra, Italy.
Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., and Schroeder, C. E. (**2007**). "Neuronal oscillations and multisensory interaction in primary auditory cortex," Neuron **53**, 279–292.
Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (**2008**). "Entrainment of neuronal oscillations as a mechanism of attentional selection," Science **320**, 110–113.
Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36.
Luo, H., Liu, Z., and Poeppel, D. (**2010**). "Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation," PLoS Biol. **8**, e1000445.
Luo, H., and Poeppel, D. (**2007**). "Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex," Neuron **54**, 1001–1010.
MacDonald, J. (**2018**). "Hearing lips and seeing voices: The origins and development of the 'McGurk effect' and reflections on audio–visual speech perception over the last 40 years," Multisens. Res. **31**(1–2), 7–18.
Macleod, A., and Summerfield, Q. (**1987**). "Quantifying the contribution of vision to speech perception in noise," Br. J. Audiol. **21**, 131–141.
Maddox, R. K., Atilgan, H., Bizley, J. K., and Lee, A. K. C. (**2015**). "Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners," eLife **4**, e04995.
Magnotti, J. F., Basu Mallick, D., Feng, G., Zhou, B., Zhou, W., and Beauchamp, M. S. (**2015**). "Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers," Exp. Brain Res. ***233***(9), 2581–2586.
Magnotti, J. F., and Beauchamp, M. S. (**2015**). "The noisy encoding of disparity model of the McGurk effect," Psychon. Bull. Rev. ***22***(3), 701–709.
Magnotti, J. F., Dzeda, K. B., Wegner-Clemens, K., Rennig, J., and Beauchamp, M. S. (**2020**). "Weak observer-level correlation and strong stimulus-level correlation between the McGurk effect and audiovisual speech-in-noise: A causal inference explanation," Cortex **133**, 371–383.
Marques, L. M., Lapenta, O. M., Costa, T. L., and Boggio, P. S. (**2016**). "Multisensory integration processes underlying speech perception as revealed by the McGurk illusion," Lang. Cogn. Neurosci. **31**, 1115–1129.
Marslen-Wilson, W. D., and Tyler, L. K. (**1980**). "The temporal structure of spoken language processing," Cognition **8**, 1–71.
Massaro, D. W. (**2017**). "The McGurk effect: Auditory visual speech perception's piltdown man," in *Proceedings of the 14th International Conference on Auditory-Visual Speech Processing (AVSP2017)*, August 25–26, Stockholm, Sweden.
Massaro, D. W., Thompson, L. A., Barron, B., and Laren, E. (**1986**). "Developmental changes in visual and auditory contributions to speech perception," J. Exp. Child Psychol. **41**(1), 93–113.
McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature **264**, 746–748.
McLaughlin, D. J., Brown, V. A., Carraturo, S., and Van Engen, K. J. (**2022**). "Revisiting the relationship between implicit racial bias and audiovisual benefit for nonnative-accented speech," Atten. Percept. Psychophys. **84**, 2074–2086.
Micheli, C., Schepers, I. M., Ozker, M., Yoshor, D., Beauchamp, M. S., and Rieger, J. W. (**2018**). "Electrocorticography reveals continuous auditory and visual speech tracking in temporal and occipital cortex," Eur. J. Neurosci. **51**, 1364–1376.
Moris Fernandez, L., Macaluso, E., and Soto-Faraco, S. (**2017**). "Audiovisual integration as conflict resolution: The conflict of the McGurk illusion," Hum. Brain Mapp. **38**(11), 5691–5705.
Munhall, K. G., Gribble, P., Sacco, L., and Ward, M. (**1996**). "Temporal constraints on the McGurk effect," Percept. Psychophys. **58**(3), 351–362.
Myerson, J., Tye-Murray, N., Spehar, B., Hale, S., and Sommers, M. (**2021**). "Predicting audiovisual word recognition in noisy situations: Toward precision audiology," Ear Hear. **42**(6), 1656–1667.
Nath, A. R., and Beauchamp, M. S. (**2011**). "Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech," J. Neurosci. **31**(5), 1704–1714.
Oden, G. C., and Massaro, D. W. (**1978**). "Integration of featural information in speech perception," Psychol. Rev. **85**, 172–191.
Okada, K., Venezia, J. H., Matchin, W., Saberi, K., and Hickok, G. (**2013**). "An fMRI study of audiovisual speech perception reveals multisensory interactions in auditory cortex," PloS One **8**, e68959.
Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (**2018**). "Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex," PLoS Biol. **16**(8), e2006558.
Park, H., Kayser, C., Thut, G., and Gross, J. (**2016**). "Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility," eLife ***5***, e14521.
Peelle, J. E. (**2018**). "Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior," Ear Hear. **39**(2), 204–214.
Peelle, J. E., and Davis, M. H. (**2012**). "Neural oscillations carry speech rhythm through to comprehension," Front. Psychol. **3**, 320.
Peelle, J. E., Gross, J., and Davis, M. H. (**2013**). "Phase-locked responses to speech in human auditory cortex are enhanced during comprehension," Cerebral Cortex **23**(6), 1378–1387.
Peelle, J. E., and Sommers, M. S. (**2015**). "Prediction and constraint in audiovisual speech perception," Cortex **68**, 169–181.
Peelle, J. E., Spehar, B., Jones, M. S., McConkey, S., Myerson, J., Hale, S., Sommers, M. S., and Tye-Murray, N. (**2022**). "Increased connectivity among sensory and motor regions during visual and audiovisual speech perception," J. Neurosci. **42**(3), 435–442.
Reisberg, D., McLean, J., and Goldfield, A. (**1987**). "Easy to hear but hard to understand: A speechreading advantage with intact stimuli," in *Hearing by Eye: The Psychology of Lip-Reading*, edited by R. Campbell and B. Dodd (Erlbaum, Mahwah, NJ), pp. 97–113.
Rosenblum, L. (**2019**). "Audiovisual speech perception and the McGurk effect," in Oxford Research Encyclopedia, Linguistics, https://par.nsf.gov/servlets/purl/10190134 (Last viewed November 21, 2022).
Rosenblum, L. D., and Saldana, H. M. (**1996**). "An audiovisual test of kinematic primitives for visual speech perception," J. Exp. Psychol. Human Percept. Perform. **22**(2), 318–331.
Rosenblum, L. D., Schmuckler, M. A., and Johnson, J. A. (**1997**). "The McGurk effect in infants," Percept. Psychophys. **59**(3), 347–357.
Sams, M., Manninenb, P., Surakkab, V., Helinb, P., and Kättöb, R. (**1998**). "McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context," Speech Commun. **26**, 75–87.
Schroeder, C. E., and Foxe, J. (**2005**). "Multisensory contributions to low-level, 'unisensory' processing," Curr. Opin. Neurobiol. **15**, 454–458.
Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (**2008**). "Neuronal oscillations and visual amplification of speech," Trends Cogn. Sci. **12**, 106–113.
Sekiyama, K. (**1997**). "Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects," Percept. Psychophys. **59**(1), 73–80.
Sekiyama, K., Soshi, T., and Sakamoto, S. (**2014**). "Enhanced audiovisual integration with aging in speech perception: A heightened McGurk effect in older adults," Front. Psychol. ***5***, 323.
Siddig, A., Sun, P. W., Parker, M., and Hines, A. (**2019**). "Perception deception: Audio-visual mismatch in virtual reality using the mcgurk effect," AICS **2019**, 176–187.

Sommers, M. S., Tye-Murray, N., and Spehar, B. (**2005**). "Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults," Ear Hear. **26**, 263–275.
Soto-Faraco, S., and Alsius, A. (**2009**). "Deconstructing the McGurk-MacDonald illusion," J. Exp. Psychol. Hum. Percept. Perform. **35**(2), 580–587.
Stevenson, R. A., Zemtsov, R. K., and Wallace, M. T. (**2012**). "Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions," J. Exp. Psychol. Human Percept. Perform. **38**(6), 1517–1529.
Strand, J. F. (**2014**). "Phi-square lexical competition database (Phi-Lex): An online tool for quantifying auditory and visual lexical competition," Behav. Res. **46**, 148–158.
Strand, J. F., Cooperman, A., Rowe, J., and Simenstad, A. (**2014**). "Individual differences in susceptibility to the McGurk effect: Links with lipreading and detecting audiovisual incongruity," J. Speech. Lang. Hear. Res. **57**(6), 2322–2331.
Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Summerfield, A. Q. (**1987**). "Some preliminaries to a comprehensive account of audio-visual speech perception," in *Hearing by Eye: The Psychology of Lip Reading*, edited by B. Dodd and R. Campbell (Erlbaum, Mahwah, NJ), pp. 3–87.
Tye-Murray, N., Sommers, M. S., and Spehar, B. (**2007a**). "Auditory and visual lexical neighborhoods in audiovisual speech perception," Trends Amplif. **11**, 233–241.
Tye-Murray, N., Sommers, M. S., and Spehar, B. (**2007b**). "The effects of age and gender on lipreading abilities," J. Am. Acad. Audiol. **18**, 883–892.
Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., and Sommers, M. (**2016**). "Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration," Psychology Aging **31**(4), 380–389.
Van Engen, K. J., Phelps, J. E. B., Smiljanic, R., and Chandrasekaran, B. (**2014**). "Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker," J. Speech. Lang. Hear. Res. **57**, 1908–1918.
Van Engen, K. J., Xie, Z., and Chandrasekaran, B. (**2017**). "Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect," Atten. Percept. Psychophys. **79**, 396–403.
van Wassenhove, V., Grant, K. W., and Poeppel, D. (**2005**). "Visual speech speeds up the neural processing of auditory speech," Proc. Nat. Acad. Sci. U.S.A. **102**(4), 1181–1186.
Venezia, J. H., Thurman, S. M., Matchin, W., George, S. E., and Hickok, G. (**2016**). "Timing in audiovisual speech perception: A mini review and new psychophysical data," Atten. Percept. Psychophys. **78**(2), 583–601.
Venezia, J. H., Vaden, K. I., Jr., Rong, F., Maddox, D., Saberi, K., and Hickok, G. (**2017**). "Auditory, visual and audiovisual speech processing streams in superior temporal sulcus," Front. Hum. Neurosci. **11**, 174.
Windmann, S. (**2004**). "Effects of sentence context and expectation on the McGurk illusion," J. Mem. Lang. **50**, 212–230.
Yi, H.-G., Phelps, J. E. B., Smiljanic, R., and Chandrasekaran, B. (**2013**). "Reduced efficiency of audiovisual integration for nonnative speech," J. Acoust. Soc. Am. **134**(5), EL387–EL393.
Yi, H.-G., Smiljanic, R., and Chandrasekaran, B. (**2014**). "The neural processing of foreign-accented speech and its relationship to listener bias," Front. Hum. Neurosci. **8**, 768.
Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., and Poeppel, D. (**2013**). "Visual input enhances selective speech envelope tracking in auditory cortex at a 'cocktail party,' " J. Neurosci. **33**, 1417–1426.