DOI: 10.1111/lnc3.12238

TEACHING AND LEARNING GUIDE

WILEY

1828

Teaching and learning guide for audiovisual speech perception: A new approach and implications for clinical populations

Julia Irwin[1,2] | Jacqueline Turcios[1,2]

[1] Haskins Laboratories, New Haven, CT, USA

[2] Southern Connecticut State University, New Haven, CT, USA

This guide accompanies the following article(s):

Audiovisual Speech Perception: A New Approach and Implications for Clinical Populations. Language and Linguistics Compass (2017), pp. XXX-XXX. DOI: 10.1111/Inc3.12237

1 | AUTHOR'S INTRODUCTION

When a speaker talks, the visible consequences of what they are saying can be seen. This auditory (the speech sound) and visual (movements of the lips and other articulators), or AV speech influences what listeners hear both in noisy listening environments and when auditory speech can easily be heard. Thought to be a cross-cultural phenomenon that emerges early in typical language development, variables that influence audiovisual speech perception include properties of the visual and the auditory signal, attentional demands, and individual differences. Further, the existing neurobiological evidence suggests facilitatory effects of audiovisual over auditory only speech. Studies of audiovisual speech perception in certain clinical populations (e.g., individuals with an autism spectrum disorder, developmental language disorder or hearing loss) have revealed differences in processing that may inform future interventions. Finally, a new method of assessing AV speech that does not require obvious cross-category mismatch or auditory noise is proposed as an alternative approach to traditional measures of AV speech perception for investigators.

2 | AUTHOR RECOMMENDS

Please add 5–10 annotated readings to help situate readers in the key relevant works in this field. If referencing an online publication, please provide a full citation, link, and a digital object identifier (DOI) if possible. See examples below:

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

The classic report that that visible articulation on a speaker's face assists listeners in the identification of speech in auditory noise, creating a “visual gain” over auditory speech alone.

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

The original paper that reports the McGurk Effect or Illusion, the most widely used measure of AV speech perception. In this study, child and adult perceivers show a visual influence on heard speech in clear (non-noisy) listening conditions. In addition, McGurk and MacDonald report developmental effects, with more visual influence in older participants.

3. Irwin, J., Brancazio, L. & Volpe, N. (2017). (resubmit with minor revisions) The Development of Gaze to a Speaking Face.

Visible influence of seen on heard speech has been reported to increase from early to late childhood (e.g., see reading #2, above), but little is known about the mechanism that underlies this developmental trend. One possibility is that patterns of gaze change with age. Eye tracking data from child and adult participants to a speaking face indicated an increase in gaze on the face, and specifically, to the mouth of a speaker between the ages of 5 and 10. These findings suggest an increasing focus on the source of speech with development, which may underlie the observed developmental trend.

4. 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.

This study assessed sensitivity to visible speech in children with an autism spectrum disorder (ASD) in comparison to chronological- and verbal mental age-matched typically developing (TD) controls. A hallmark behavior associated with an ASD is poorly modulated gaze to the faces of others. However, even when gaze was fixated on the speaker's face, children with ASD were less influenced by visible articulatory information than their TD peers, in speechreading, speech-in-noise tasks and with audiovisual mismatched (McGurk) stimuli.

5. Irwin, J., Avery, T., Brancazio, L., Ryherd, K., Turcios, J. & Landi, N. (in press). Electrophysiological Indices of Audiovisual Speech Perception: Beyond the McGurk Effect and Speech in Noise

Commonly used approaches to study AV speech include mismatched audiovisual stimuli (e.g., McGurk type stimuli, see citation 2, above) or visual speech in auditory noise (see citation 1, above). This paper introduces a novel visual phonemic restoration method that does not require obvious cross-category mismatch or auditory noise is proposed as an alternative approach to traditional measures of AV speech perception. If listeners “hear” /ba/ when looking at a face when the auditory signal is /a/, the listener has restored the missing consonant, indicating that integration has occurred. Behavioral and ERP findings reflect this phonemic restoration in typically developing adults; specifically, we observed reduced accuracy and P300 response in the presence of visual speech. This novel method has potential utility for listeners who cannot respond actively, such as infants and individuals with developmental disabilities.

3 | ONLINE MATERIALS

Please add up to 5 links to other external sites. If available, sites/blogs with an emphasis on regular updates and multimedia content (images, audio, video, etc.) will add depth and visual interest to your guide. Please provide files for any images which you own or which have no rights issues (including acknowledgement of any sources). Any images should be submitted in either JPEG or GIF format. A dpi of 120 is recommended.

1. http://www.faculty.ucr.edu/~rosenblu/lab-index.html

Dr. Larry Rosenblum's AV Speech and Audition Lab, with demos of a range of AV stimuli and effects.

2. http://haskinslabs.org/people/julia-irwin

Dr. Julia Irwin's AV perception lab at Haskins Laboratories, New Haven, CT.

Clip of McGurk stimuli, Visual Ga and Auditory Ma, often heard as /na/ (also for use for seminar activity, below).

3. https://auditoryneuroscience.com/McGurkEffect

Dr. Pat Kuhl's Demonstration of the effect.

4. https://www.nytimes.com/2017/02/21/science/lip-reading-mcgurk-effect.html?_r=0

Article in the New York Times on the McGurk Effect.

4 | SAMPLE SYLLABUS

Please add a syllabus for a full or partial syllabus (i.e., covering 2–5 weeks of teaching) based around your article and the material it covers. Alternatively, you may include a syllabus for an entire course if appropriate. See examples below:

This sample syllabus includes readings appropriate to both general-level courses indicated by (A) and advanced or seminar courses (B).

Week 1. Speech is more than a sound

A. Visit Dr. Larry Rosenblum's website, Dr. Julia Irwin's website, Dr. Pat Kuhl's website, read New York Times article.

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

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

B. Visit Dr. Larry Rosenblum's website, Dr. Julia Irwin's website, Dr. Pat Kuhl's website, read New York Times article.

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

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

Green, K. P. (1996). The use of auditory and visual information in phonetic perception. In Speechreading by humans and machines (pp. 55–77). Springer Berlin Heidelberg.

Week 2. Variations in the McGurk Effect: Selected papers

A. Déry, C., Campbell, N. K., Lifshitz, M., & Raz, A. (2014). Suggestion overrides automatic audiovisual integration. Consciousness and Cognition, 24, 33–37.
B. 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. Attention, Perception, & Psychophysics, 50(6), 524–536.

Rosenblum, L. D., Yakel, D. A., & Green, K. P. (2000). Face and mouth inversion effects on visual and audiovisual speech perception. Journal of Experimental Psychology: Human Perception and Performance, 26(2), 806.

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.

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. Perception & Psychophysics, 54(3), 406–416.

Week 3. Audiovisual speech perception in special populations

A. Irwin, J.R., Preston, J.L, Brancazio, L., D'Angelo, M. & Turcios, J. (2014). Development of an audiovisual speech perception app for children with autism Spectrum disorders. Clinical Linguistics and Phonetics. *doi:10.3109/02699206.2014.966395.*
B. Irwin, J. R., Tornatore, L. A., Brancazio, L., and Whalen, D. H. (2011). Can children with autism spectrum disorders “hear” a speaking face? Child Development, 82(5), 1397–1403.

Norrix, L. W., Plante, E., Vance, R., and 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.

Guiraud, J. A., Tomalski, P., Kushnerenko, E., Ribeiro, H., Davies, K., Charman, T., & BASIS Team. (2012). Atypical audiovisual speech integration in infants at risk for autism. PloS One, 7(5), e36428.

Week 4. Limitations to the McGurk Effect and Speech in Noise and an alternative approach.

B Only Irwin, J., Avery, T., Brancazio, L., Ryherd, K., Turcios, J. & Landi, N. (2017). Electrophysiological Indices of Audiovisual Speech Perception: Beyond the McGurk Effect and Speech in Noise.

5 | FOCUS QUESTIONS

Please add 5 focus questions to help readers spring-board into the wider subject matter. See examples below:

(Focus questions should be presented after Seminar Activity demonstration described below).

Why might visible speech be helpful for the listener?

In noisy environments, like cafeterias, classrooms, and playgrounds, children and adults may struggle to recover the message of the listener from sound alone. For many years, we have known that seeing the face of a speaker can help identify their message in the presence of background noise (Sumby & Pollack, 1954).

Do blind individuals who can't see the face show differences in speech perception?

Blind individuals display differences in speech perception and production in comparison to sighted individuals (Ménard, Dupont, Baum & Aubin, 2009), suggesting that visible articulatory information on a speaker's face is a central part of typical perceptual development. This audiovisual or “AV” speech likely fosters native language acquisition (Legerstee, 1990) and visual influence has been demonstrated in infancy (Burnham & Dodd, 1998; Dejardins & Werker, 2004; Meltzoff & Kuhl, 1994; Rosenblum, et al., 1997).

Visual influence in the context of a McGurk task and speechreading ability (identification of a syllable from visual information only) increases with age. *Why might this effect change with development?*

Increased visual influence with development could be due to (a) experience with producing speech sounds (Dejardins et al, 1997 report visual influence for children if they can produce the sound that they see on another speaker's face), (b) ongoing perceptual learning/tuning with respect to visual speech during childhood (Ross et al., 2011; Hockley & Polka, 1994), (c) younger children may be less attentive to the visual source, leading to an attenuated visual effect (Massaro, 1984), or (d) pattern of gaze to the mouth of the speaker with development, that is, less looking at the speaker's mouth in younger children (Irwin, Brancazio & Volpe, 2017).

How does this effect work in individuals from special populations, such as autism spectrum disorders, developmental language disorder or hearing loss?

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 and Turcios, 2014).

What if I don't show a McGurk Effect or Illusion?

There is a great deal of individual variability in the effect—for more details, please see Nath, A. and Beauchamp, M.S. (2011). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage, 59 (1), 781–787.

6 | SEMINAR ACTIVITY

If possible, please suggest an exercise to help bring the subject to life, appropriate either for undergraduate or graduate students, for example, an assessment, a presentation, or other practical assignment.

7 | IN CLASS DEMONSTRATION/DISCUSSION: APPROPRIATE FOR BOTH UNDERGRADUATE AND GRADUATE LEVEL STUDENTS

Present the sample McGurk-type stimuli, visual /ga/ +_auditory /ma/ (at: http://haskinslabs.org/people/julia-irwin) to students in a classroom, free field at a moderate sound level (no headphones). Ask students to watch the speaker's face and report what they hear. (Note: If they report something other than /ma/ or its visual equivalent (e.g., /ba/), such as /na/, /da/, /la/, then they have shown the

McGurk Effect). Then ask students to close their eyes and do the task again. Many students will hear a "na" (or visual equivalent, known as a viseme) with eyes open but will clearly hear a "ma" with eyes closed. Note variability in response. Many listeners will get the illusion, some will not.

Explain to students what the effect is called—the McGurk Effect or Illusion—and that it is robust, occurring even if listeners are aware of the manipulation (Rosenblum and Saldaña, 1996), when female faces and male voices are dubbed (e.g., Green, Kuhl, Meltzoff and Stevens, 1991; Johnson, Strand, and D'Imperio, 1999), or if the audio and visual signals are not temporally aligned (e.g., Munhall, Gribble, Sacco, and Ward, 1996). This demonstrates that visual speech influences what is heard, even if the speech is in clear listening conditions. In other words, we hear what we see!

Query students about what variables they think might influence the effect. Report some of what is known. For example, 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 and Munhall 2003) Quality of the visual signal has also been directly manipulated in the study of AV speech. Even degraded visual signals (MacDonald, Anderson and Bachmann, 2000; Munhall, Kroos, and Vatikiotis-Bateson, 2002) and point-light displays of the face, (where motion of the articulators is shown by placing small light markers on the lips, Rosenblum and Saldaña, 1996; Callan, Jones, Munhall, Kroos, Callan and Vatikiotis-Bateson (2004) yield a visual influence on heard speech. Visual influence is lessened when visual attention is drawn to another stimulus placed on the speaking face (e.g., Alsius, Navarra, Campbell, and Soto-Faraco, 2005; Tiippana, Andersen and Sams, 2004), and suggestibility can reduce the effect under hypnosis (Déry, Campbell, Lifshitz, & Raz, 2014).