McGurk Doesn't Work:
Using EEG to Investigate the Time Course of the McGurk Effect

Alexandra Griffin, Sarah Mann, Izzie Ramirez, & Laura M. Getz
Department of Psychological Sciences, University of San Diego

www.lamplab.org

INTRODUCTION

Speech perception is an inherently **multimodal process**, where auditory signals provide concurrent information to visual information from the speaker's mouth movements.

The **McGurk effect** is an illusion often used to study audiovisual speech integration; it occurs when presented with **incongruent auditory and visual speech cues**. In the original McGurk study, listening to the spoken syllable /pa/ while watching visual mouth movements for /ka/ resulted in a 'fusion' perception of /ta/ in 81% of participants (McGurk & MacDonald, 1976). Since the original study, others have found much lower percentages of fusion responses that are influenced by participant and task differences (Basu Mallick et al., 2015; Getz & Toscano, 2021).

Research on audiovisual integration has largely looked at effects in isolated syllables. **Our goal was to enhance the ecological validity of the McGurk effect** by creating word stimuli mimicking everyday conversations (e.g., pairing audio for /pig/ with lip movements for /tig/ to determine listeners' interpretation).

In this study, we varied **task** (forced-choice vs. open-ended) between-participants and **stimuli** (words vs. non-words) within-participants. We argue that an **ecological model of speech perception** showing differences in fusion based on the word or nonword stimuli used and the response demanded by the task necessarily mean the **McGurk illusion cannot be a perceptual effect.**

METHODOLOGY & DESIGN

Participants: 46 Introduction to Psychology students.

Task: Each participant was randomly assigned to respond to a three-alternative forced choice (3AFC) or open-ended task. In the **3AFC task**, participants (*N* = 31) only had the response options P, T, and K. In the **open-ended** task, participants (*N* = 15) typed whatever they thought the speaker said.

Design: Experiment was completed using **Open Sesame** software. Participants completed 10 practice trials, followed by **three blocks** of **90 AV trials** (3AFC task) or **two blocks** of **90 AV trials** (open-ended). There was also a **visual-only block** with 54 trials and an **audio-only block** with 54 trials.

STIMULI

All auditory, visual, and combined files are available on Google Drive using the QR code

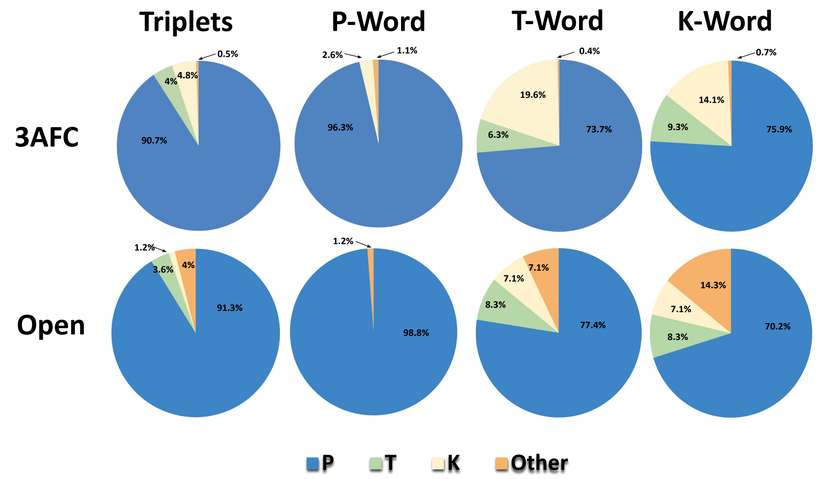
Visual stimuli: A male speaker was videotaped with a neutral background and neutral facial expressions. He looked directly at the camera and spoke individual words.


Auditory stimuli: Stimuli were recorded in a sound-attenuated booth by a male speaker of American English. Stimuli stimuli consisted of **congruent and incongruent AV examples** of one-syllable words beginning with /p/, /t/, and /k/. Stimuli were created by formulating triplets in which all PTK words were real words ("triplets" condition, e.g., pail/tail/kale) or where only one PTK word was real ("P-word", "T-word", and "K-word" conditions, e.g., pig/tig/kig). We had nine unique word sets for triplets and nine unique word sets for nonword triplets (three each for P, T, and K).

Triplets	P-word	T-word	K-word
Pail / Tail / Kale	**Pig** / Tig /Kig	Pooth / **Tooth** / Kooth	Pid / Tid / **Kid**
Pear / Tear / Care	**Park** / Tark / Kark	Pip / **Tip** / Kip	Pamp / Tamp / **Camp**
Pick / Tick / Kick	**Path** / Tath / Kath	Peam / **Team** / Keam	Purb / Turb / **Curb**
Pill / Till / Kill			
Pin / Tin / Kin			
Pool / Tool / Cool			
Poll / Toll / Coal			
Post / Toast / Coast			
Pour / Tore / Core			

Audiovisual stimuli: We began with **congruent AV** versions of /P/, /T/, and /K/ words/non-words. The **incongruent AV** stimuli were created using **iMovie**; we combined auditory /P/ with visual /K/ and auditory /K/ with visual /P/. For the audio to align with the speaker's lip movements, congruent auditory and visual parts of the AV files were separated. The new incongruent audio track was overlaid to ensure that the onset of the consonant bursts matched before the congruent audio track was removed.

RESULTS



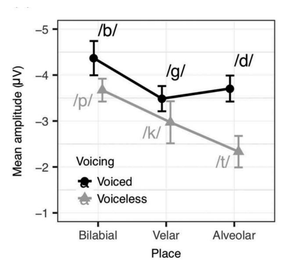
- The auditory ("P") option was selected most frequently overall.
- There was not much difference between 3AFC and open in response frequencies overall.
- The fusion ("T") option was selected more in the T-word and K-word conditions than the P-word and Triplet conditions.

CONCLUSIONS/FUTURE DIRECTIONS

We conclude that the McGurk effect occurs at a decision level rather than a perceptual level because task and stimulus differences influenced McGurk effect likelihood.

To better understand the McGurk effect and audiovisual speech perception more generally, future EEG use will allow time course isolation in which participants are making decisions regarding the task and stimuli. Doing so will allow us to more capture early perceptual processes separately from later categorization and decision-level processes (e.g., Toscano et al., 2010) and also allow for the measurement of top-down effects on perception (e.g., Getz & Toscano, 2019).

Following Pereira et al. (2018), measuring the N1 amplitude in response to a variety of congruent and incongruent /p/, /t/, and /k/ will allow us to see whether *perception* is actually changing when viewing incongruent "mcgurk" stimuli. Given our behavioral results, we would predict that the N1 would match what participants hear rather than a fusion phoneme (e.g., N1 would match a /p/ rather than /t/ response when viewing incongruent aP-vK stimulus).



REFERENCES

Basu Mallick, D., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review, 22*(5), 1299-1307.
Getz, L. M., & Toscano, J. C. (2019). Electrophysiological evidence for top-down lexical influences on early speech perception. Psychological Science, 30(6), 830-841.
Getz, L. M., & Toscano, J. C. (2021). Rethinking the McGurk effect as a perceptual illusion. *Attention, Perception, & Psychophysics, 83,* 2583–2598.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748.
Pereira, O., Gao, Y. A., & Toscano, J. C. (2018). Perceptual encoding of natural speech sounds revealed by the N1 event-related potential response. Auditory Perception & Cognition, 1(1-2), 112-130.
Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. Psychological Science, 21(10), 1532-1540.

Special thanks to USD Office of Undergraduate Research for an OUR Travel Award and the Department of Psychological Sciences for a Student Travel Award for Research (STAR) to fund conference travel to present this research.