Dianne Nowack, Michael Fischbacher, Felicitas Kleber

Institute of Phonetics and Speech Processing, Ludwig-Maximilians-Universität Munich, Germany

Is there a McGurk effect in German tense vowels?

1. Background

- McGurk effect: The influence of visual cues on the perception of speech signals (McGurk & MacDonald, 1976)
  audio signal ≠ visual information → possibility of third sound percept – an articulatory/acoustic "merger" (auditory /b/ + visual /g/ → /d/).
- Listeners rely on visual cues to differentiate rounded from unrounded high, front vowels (Traunmüller & Öhrström 2007; Kleber et al., 2010)
- Perception of openness is less affected by visual cues (Traunmüller & Öhrström, 2007)

2. Hypotheses

H1 There is a McGurk effect for the German tense vowel sets /i, e, ɛ/ (= open) & /i, y, u (=rounded). Listeners perceive a merger when audio and visual information do not match.

H2 The McGurk effect is more pronounced for the set /i, y, u/ i.e. more mergers in the perception of vowels that differ in lip rounding.

H3 The McGurk effect is more pronounced for a /t/ context in the /i, y, u/ set because here the /y-u/-contrast is diminished due to /u/-fronting
(e.g. Kleber et al., 2010)

3. Method

Materials

Video & audio recordings of the sets /i, e, ɛ/ and /i, y, u/ as well as /tit, tyt, tut/ and /pip, pyp, pup/

Stimuli

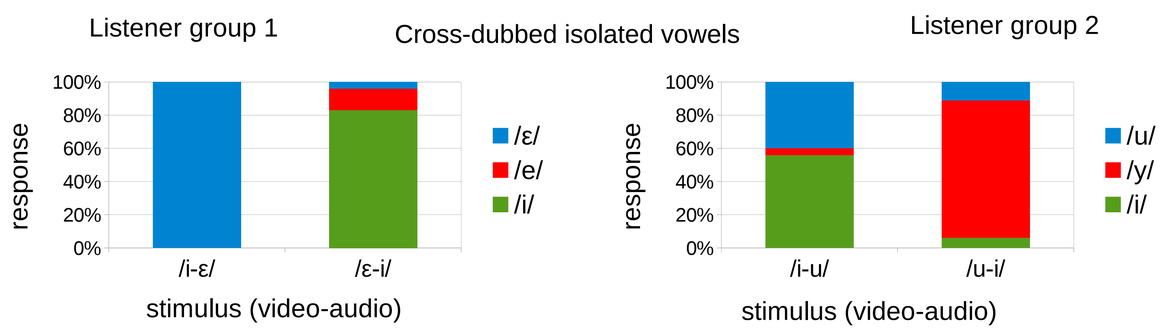
- Mute, only-audio, dubbed and cross-dubbed stimuli within each set
- Pink noise added to all audio files

Participants & Task

- German speakers (21 – 55 yrs)
- 12 for /i, e, ɛ/ set and 12 for /i, y, u/
- Three-alternative forced-choice task run in E-Prime

4. Results

The data were statistically analysed with two GLMM: dependent variables = merger (true = /e, y/ vs. false = /i, ɛ, u/), fixed factors = stimulus and – depending on the analysis – listener group (/i, e, ɛ/ vs. /i, y, u/) or context (/t/ vs. /p/).

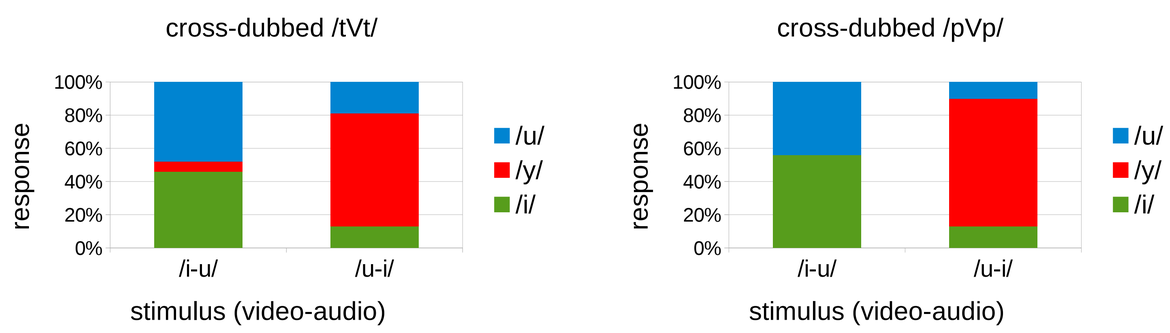


H1 + H2 ✓

→ Listeners perceive a merger, but only when the audio-stimulus is /i/ and when the vowels differ in lip rounding.
(z.B. /u-i/ - /e-i/: $z = 5.385$, $p < 0.001$)

→ Listeners integrate visual cues only in the classification of [± rounded] vowels

→ Some listeners rely more on visuals, others more on auditory cues

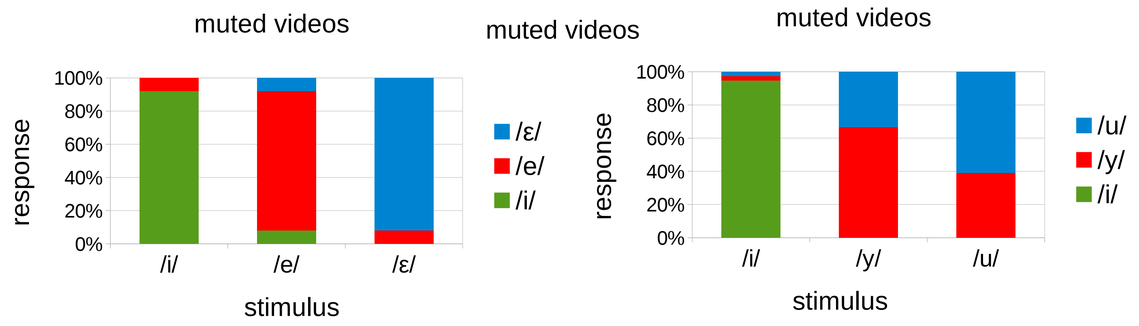


H3 ✗ No significant difference between /t/- and /p/-context stimuli

→ significant effect for Stimulus ($X^2 = 111.9$, $p < 0.001$): only stimuli with auditory /i/ cause perception of mergers

5. Discussion

- /i, e, ɛ/ can already be distinguished without sound input
- /y, u/ could hardly be distinguished without sound input



- /i-ɛ/ vs. /ɛ-i/:
  Listeners rely mostly on audio signal; majority of /i/ responses in /ɛ-i/ presumably due to the fact that /e/ and /i/ are acoustically close and /e/ and /ɛ/ are visually similar

- /i-u/ vs. /u-i/:
  Very few /y/ responses in /i-u/ → large discrepancy between audio and video signal → listeners rely either on audio OR on video signal
  McGurk effect only in /u-i/ because /u/-video and /i/-audio are similar to /y/

6. References

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

Traunmüller, H., & Öhrström, N. (2007). Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics*, 35: 244 - 258.

Kleber, F., Reubold, U., Harrington, J. (2010). /u/-fronting in RP and the implications of perceptual integration of lip gestures for sound change processes. In *12th Conference on Laboratory Phonology*, Albuquerque, USA.