Visual cues in sound change: A cross-modal perceptual account for the typological rarity of labial palatalization.

Jonathan Havenhill1, Baichen Du1,2
1Department of Linguistics, University of Hong Kong; 2Department of Linguistics, University of California, Berkeley | jhavenhill@hku.hk

Background
Listener-based sound change

- (Mis-)perception of acoustically similar sounds contributes to diachronic sound changes (Ohala, 1981, 1993; Blevins, 2004).
- Visual cues in speech perception

- Enhance speech intelligibility (Sumby and Pollack, 1954)
- Alter linguistic percept when presented with conflicting auditory cues (McGuirk & McDonald, 1976)
- Visually perceived lip movements activate brain regions responsible for auditory perception (Calvert et al., 1997).

Visual cues in sound change: Understudied

- Unrelated variants of CAUGHT are more likely to be misperceived as COT for Northern Cities-shifted listeners (Havenhill, 2018; Havenhill & Do, 2018).

The typological asymmetry of palatalization

- Palatalization of labials:
  - Extremely rare across languages (Bhat 1978, Kochetov 2011)
  - Limited to secondary but not full palatalization. Change from /p/ to /tŋ/ is attested in only 2-3 lgs (Bateeman, 2010).
- Accounts for labial palatalization in Setswana:
  - Direct sound change due to acoustic similarity (Ohala, 1978).
  - Indirect change via telescoping: Hardening of fricative noise and subsequent deletion of labials (Bateeman, 2010).

The current study

- To what extent do visual cues contribute to or inhibit misperception-based sound change?
- Given diminished auditory place cues, why is [p] seemingly not misperceived as [t]/[t͜a]?
- Hypothesis: Visible lip closure offers additional cue to place of articulation (PoA) and inhibits misperception

Study Design

- 3 conditions: Audio-only, Audiovisual Congruent, and Audiovisual Incongruent
- Audio stimuli: Interpolation of natural tokens produced by female Korean speaker in Matlab.

- All acoustic cues were interpolated in 5 steps.
- 3 sets of continua: /p/ to /tŋ/ /tŋ/ to /k/ /k/ to /h/ with vowels /i, a, o, u, e, / and glide /j, f, ɹ/.
- Visual stimuli:
  - Congruent: original pairing of audio and video
  - Incongruent: swapping lip-opening and lip-closure fillers: /m, n, h, j, w/ /f, a, o, u, e, /

Procedure

- 2-alternative forced choice task through Prolific
- Fixation cross to Audio/Audio+video → Response
- Reaction time and keyboard response collected
- Participants: 75 Korean speakers (42 f/31 M/2 other)

Analysis and Results

Table 1: The experiment design

<table>
<thead>
<tr>
<th>Condition</th>
<th>Audio Only</th>
<th>Visual cue</th>
<th>Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>5-step continuia for each on-set vowel combination</td>
<td>Labial: lip-closure Non-labial: lip-opening</td>
<td>stops vs. affricate</td>
</tr>
<tr>
<td>Incongruent</td>
<td>Labial: lip-opening Non-labial: lip-closure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Logistic Mixed-Effects Regression

| Item | Estimate | Std. Err | z value | Pr(>|z|) |
|------|----------|----------|---------|---------|
| (Intercept) | -4.22 | 0.14 | -29.65 | <0.001 *** |
| Step | 1.18 | 0.02 | 54.49 | <0.001 *** |
| Glide (PreGlide) | 1.02 | 0.05 | 20.76 | <0.001 *** |
| Condition (Congruent) | -0.89 | 0.16 | -5.64 | <0.001 *** |
| Condition (Incongruent) | -0.43 | 0.16 | -2.66 | <0.01 ** |
| Onset (ph) | -0.25 | 0.06 | -4.30 | <0.001 *** |
| Onset (th) | 0.27 | 0.06 | 4.63 | <0.001 *** |

Discussion

- Both acoustic and visual cues significantly influence perception of stop-affricate place contrast in Korean.
- Perception of cues in both dimensions is line with typological patterns.
  - High vowels, glide /j/, and lip opening gesture favor synchronic misperception of stops as affricates.
  - Low vowels, absence of glide, and visible lip closure gesture disfavor such misperception.
- Audiovisual congruence improves identification of phonological category.
- Incongruence for onsets originally with lip opening gesture, increases reaction time and cognitive load.
- Future directions:
  - How do acoustic place cues (F2, burst/fricative noise) trade between each other and with visual cues?
  - To what extent do speakers optimize their productions for visual perceptibility? (see 2pSC22, this session)
  - How do visual cues of the same feature/category differ? (see 3aSC63, tomorrow)

Conclusion

- Full labial palatalization is cross-linguistically rare, while coronals and dorsals are frequently palatalized.
- Acoustically palatalized stimuli are less likely to be perceived as [t] when shown with visible lip closure.
- This study suggests that both visual cues (i.e., lip closure) and articulatory factors contribute to the asymmetry.
- Additional analyses also support previous findings about the role of audiovisual congruence in speech processing.

Selected References

- Bateman, N. (2010). The change from labial to palatal as glide hardening. Linguistic Typology 14(3-4), 167-211.

Contact