Consonant and vowel rounding: same acoustics, different visuals (3aSC36)
Baichen Du1,2, Alexandra Pfiffner3, Keith Johnson1
1Department of Linguistics, University of California, Berkeley; 2Department of Linguistics, University of Hong Kong | baichen@connect.hku.hk

Overview

The phonological feature [±round] is associated with vowel and consonant contrasts, e.g., retroflex/non-retroflex in Mandarin sibilants1,2.

Rounding increases the length of the front resonating cavity, decreasing the filtered frequency3.

This study: whether visual cues to [±round] is consistent within category and between consonants and vowels, comparing acoustics and visuals of any rounding.

Rounding for Mandarin vowels is characterized by small horizontal and vertical apertures, whereas retroflex consonants have higher vertical aperture.

Background

What exactly is rounding?


Increasing length of the front cavity and lower resonating frequency.

Articulatory variation

A phonological feature can be realized with different articulatory gestures.

They are all labeled as the same feature based on acoustic parameters or perceptual categorizations.

Rounding can be variable, e.g., compressed /u/ vs. protruded /u/ in German8.

Measurement and construction of [+round]

Existing studies: acoustic parameters.

Lower resonating frequency (e.g., F2, F3)

However, articulatory properties that are visible outside of oral cavity can be part of total cues to the contrast employed by speaker and available to listener.

The current study

What is the difference in visual cues between consonant rounding and vowel rounding in Mandarin, if any?

Do visual properties of consonant rounding fall into any category of the rounding patterns?

Are visual cues integrated with acoustic cues for consonants the same way as vowels?

The Production Experiment

• Audiovisual production of consonants and vowels
  /i, ɨ, y, w, ey, /ts, ʈʂ, ʈʂʰ, tʂʰ/ contrasts.

• To avoid gestural blending with adjacent sounds:
  ○ Palatoalveolar + /i/-/ɨ/ and retroflex/non-retroflex + /ɨ/

• 60 Mandarin native speakers born and raised in China aged 18 – 22.

• Speak after a beep (articulatory onset) and stay still
  ○ Though the ComputerVision tools are not very affected by movements9.

• Video segmentation and camera calibration using Python scripts

• 68 critical facial landmarks tracked by OpenFace10.

Analysis and Results

• The calibrated distance between both lip corners (A1), and between upper and lower lips (A2).

• [±Round] uniformly decreases resonating frequency:
  ○ F2 is significantly lower for /ɨ/ than /i/ (t=31.99, df=59, p<0.001***) by 341.2Hz.

• Retroflexes’ CoG is 3685Hz lower compared to non-retroflexes (t=29.811, df=89, p<0.001**).

• Rounded high-front vowels have smaller A1, A2 and smaller lip opening areas
  ○ /ɨ/ , A1: 54.66mm; A2: 22.45mm; lip opening area: 857.38mm².

• /ɨ/ , A1: 50.82mm; A2: 21.46mm; lip opening area: 759.72mm².

• Rounded consonants have smaller A1, but larger A2 and lip opening areas
  ○ Retroflex, A1: 52.05mm, A2: 24.65mm; lip opening area: 896.8mm².

• Non-retroflex, A1: 54.46mm, A2: 22.2mm; lip opening area: 846.7mm².

• Significant differences in all parameters.

• A1 and A2 strongly predicted acoustic cues for both classes in LMER.

• Vowels: A1 (β =1.98, t = 6.07, p<0.001***) and A2 (β = 11.71, t = 19.64, p<.001***)

• Consonants: A1 (β =39.83, t=35.6, p<.001) and A2 (β=6.8, t=8.9, p<0.001***)

Discussion

• Phonetic descriptions of features have relied on acoustic parameters.

• Finer differences in articulation and visual cues were found within vowels, but it also exists cross-category between vowels and consonants.

• Ambiguity in audiovisual signals can mislead perceptual reanalysis and result in “innocent misperception”, mapping phonetic cues to a category that is not intended.

• Visual cues to features might be stored in representations with acoustic cues. The construction of [+round] might need to consider both acoustic and visual parameters.

• For acoustics, [+round] decreases resonating frequency for both vowels and consonants.

• For visuals, [+round] covers distinctive visual properties, suggesting multiple articulatory/visual correlates.

Conclusion

[±round] has both acoustic and visual correlates in the contrasts we tested.

Acoustic consequence for rounding is consistent, whereas visual cues and articulatory gestures for [+round] have more variation within vowels, and between vowels and consonants.

This suggests that the same phonological feature may have multiple phonetic correlates, and distinct correlates for different categories.

Selected References


Contact

• Scan the QR code to download the poster after ASAIB5 and know more about the 1st and 2nd authors.

We thank all subjects who participated in this study. The presenting author also thanks both co-authors, Prof. Keith Johnson and Alexandra Pfiffner for their amazing mentorship and supervision for the project. He also thanks Prof. Jonathan Havenhill at the University of Hong Kong, whose expertise in articulatory phonetics inspired many aspects of the study.