Lip postures of high vowels in Taiwan Mandarin

Chenhao CHIU, Po-Hsuan HUANG

NATIONAL TAIWAN UNIVERSITY

National Taiwan University

Lip postures

- Lip rounding is more of *phonetic* classification than *phonemic* identification (Lisker & Rossi, 1992).
  - Lip rounding can be individually variable: some speakers do not have rounded feature for English /ʃ/.
- Feature [round] is associated with lip rounding and protrusion (Catford, 1988; Lisker & Rossi, 1992; Jackson & McGowan, 2012, etc.).
- Two types of LABIALS: (Catford, 1988)

/u/, /o/ → ❖ *endolabial*: the rounding of the inner circle of the lips, substantial protrusion, pull corners together (laterally).

/y/, /ø/ → ❖ *exolabial*: the outer rounding of the lips, limited protrusion, elliptical aperture.

National Taiwan University

Modularized speech

	Lip protrusion (Linker, 1982) (Catford, 1988) (Saito, 2016)	Lateral pull (Linker, 1982) (Catford, 1988)	Lip aperture (Linker, 1982) (Montgomery & Jackson, 1983) (Mayer et al., 2021)	Lip approximation (Linker, 1982) (Jackson & McGowan, 1992) (Catford, 1988) (Saito, 2016)
/i/	(unspecified)	spread	substantial	less affected
/u/	obvious	centered	limited; circular	most substantial
/y/	some but limited; varies across lngs.	constrained	≧ aperture for /i/; varies across lngs; elliptical	substantial; varies across lngs.

- /i/ and /u/ differ not merely in tongue backness but also significantly across all lip measurements.
- /y/ is expected to contrast with /i/ in lip postures only. But their contrasts were not the greatest. → Different strategies? Different weightings?

National Taiwan University

Introduction

- Taiwan Mandarin has three high vowels: /i/, /u/, and /y/ (Ladefoged & Maddieson, 1996; Lin, 2007).

	[back]	[round]
/i/	–	–
/u/	+	+
/y/	–	+

/i/

/u/

/y/

- [round] provides more visual cues than [back] for the tongue (Lisker & Rossi, 1992).
- The **+/–** values are dichotomic, implying that sounds with the same feature values share the same articulatory gestures.

→ (articulatorily) true? visually distinct?

National Taiwan University

Research questions

1. What are the lip postural differences in Taiwan Mandarin high vowels?
2. Are these differences consistently realized in production?
3. To what degree automatic tracing + machine learning lend supports to physiological bases and their role in production?

- (Potential) Implications:
  - multimodal processing (e.g., McGurk effect)
  - cue-weighting shift (within visual cues/postural dimension)
  - sound changes.

National Taiwan University

Data examined

1. Lab speech, collected from a **production** experiment
2. Natural speech, used as **stimuli** in a perception experiment.

Lab speech

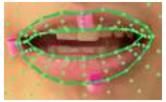
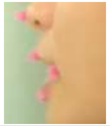
National Taiwan University

Methods

- Participants: 18 native speakers of TM
  23.44 y.o.)
- Stimuli and procedure:
  - ❖ Critical stimuli: /i/, /u/, /y/ in Taiwan Ma
  - ❖ Self-paced reading task; 10 tokens each.
  - ❖ Tones and environments:
    - ➢ monosyllabic: /i/ ‘*clothes*’, /u/ ‘*house*’, /y/ ‘*mu*
    - ➢ disyllabic: /ta-i/ ‘*big chair*’, /ta-u/ ‘*fifth year in*
      ‘big ant’
  - ❖ MediaPipe (powered by Google): face detection and tracking.
  - ❖ Melting beads: on upper lip, lower lip, and right corner of the mouth.

National Taiwan University

Measurements and analyses

Data acquisition	Measurement	Analysis
MediaPipe 	1. (max.) Lateral distance 2. (max.) Vertical distance 3. Axial ratio (= lateral/vertical) 4. Lip aperture area	Linear Mixed Models (LMM) • Fixed effects: VOWEL, WORD • Random slopes: VOWEL, WORD • Random intercept: PARTICIPANT • Baseline: monosyllabic /y/ • All data z-transformed.
Bead 		

National Taiwan University

Results – Distances

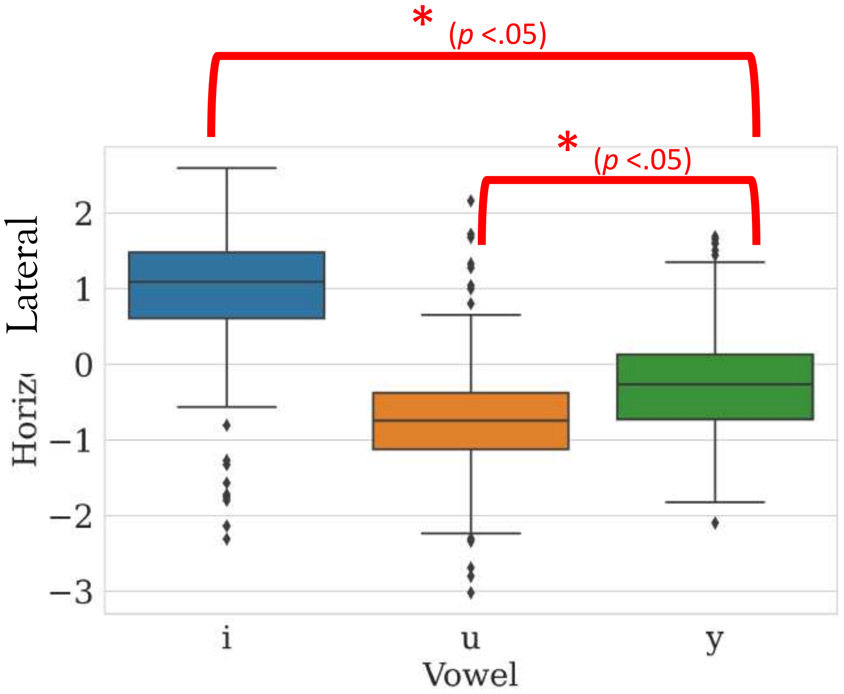


Figure 1. Results of *lateral distance*

National Taiwan University

Results – Ratio and area

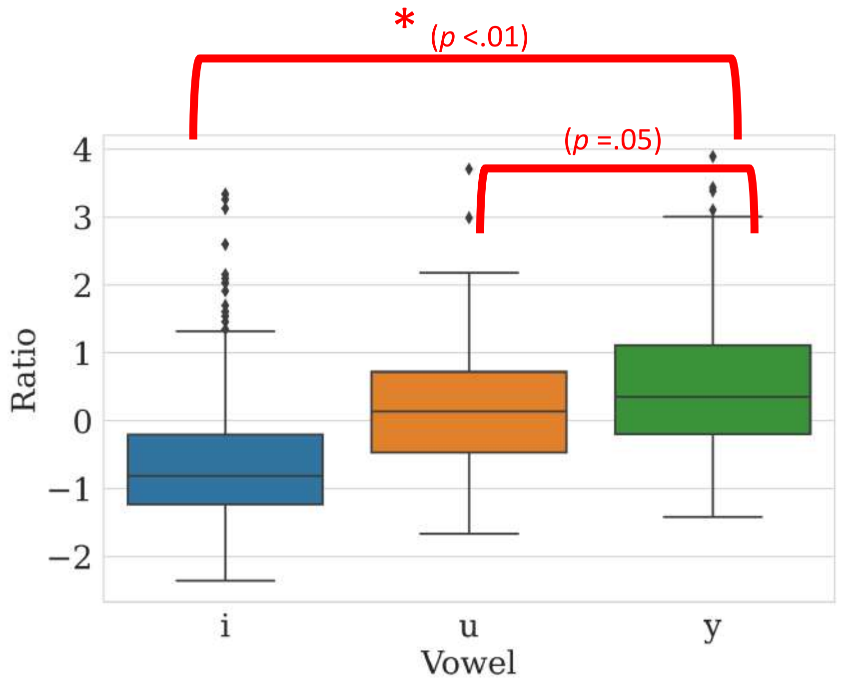


Figure 3. Results of *axial ratio*

National Taiwan University

Results – Lip protrusion

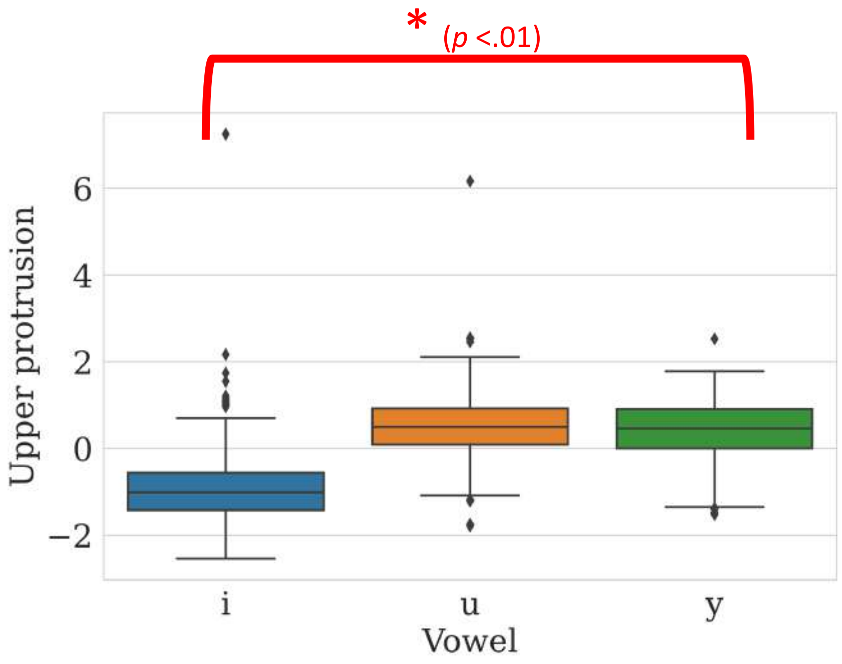


Figure 5. Results of *upper lip protrusion*

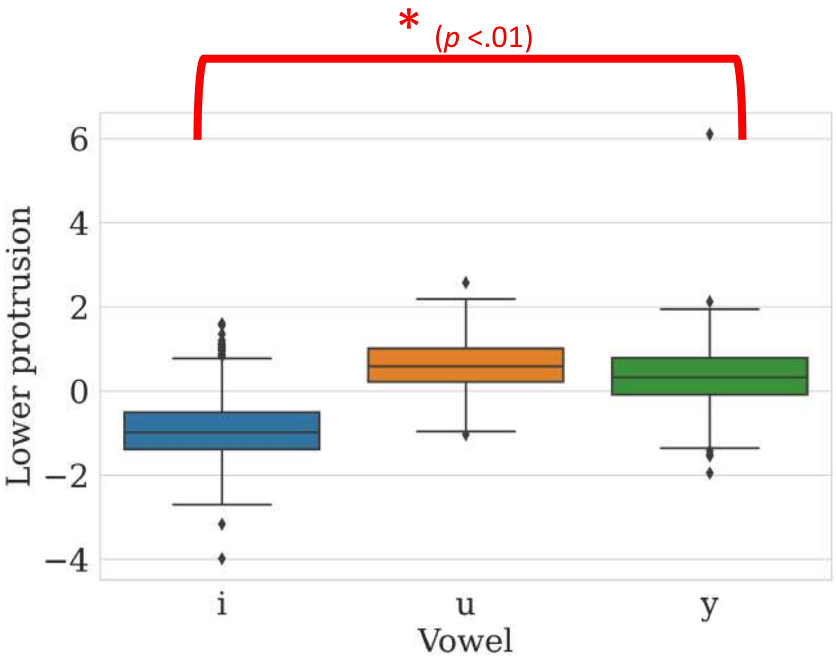
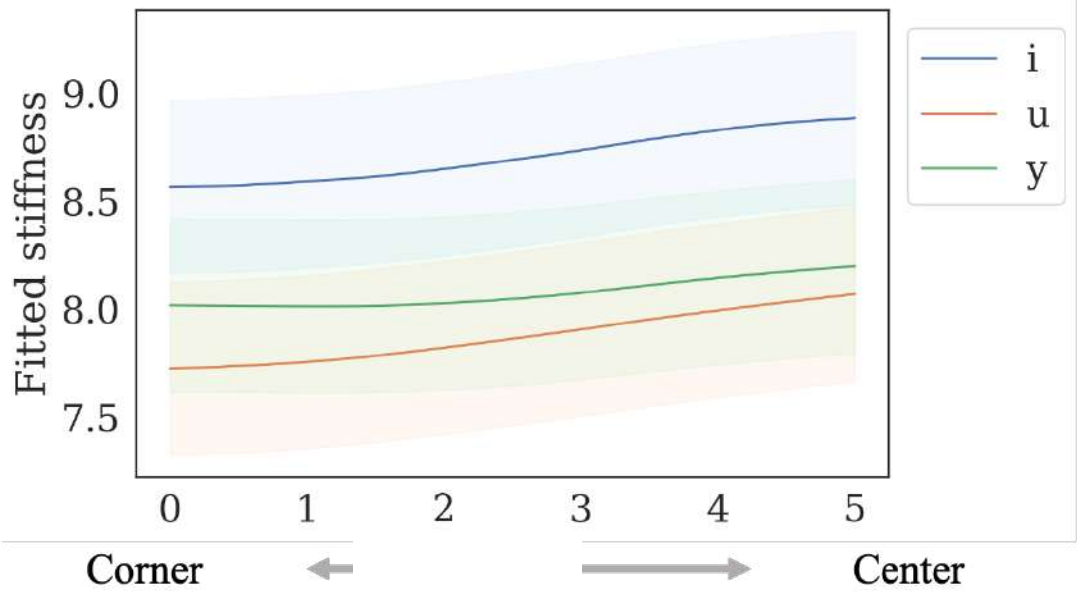
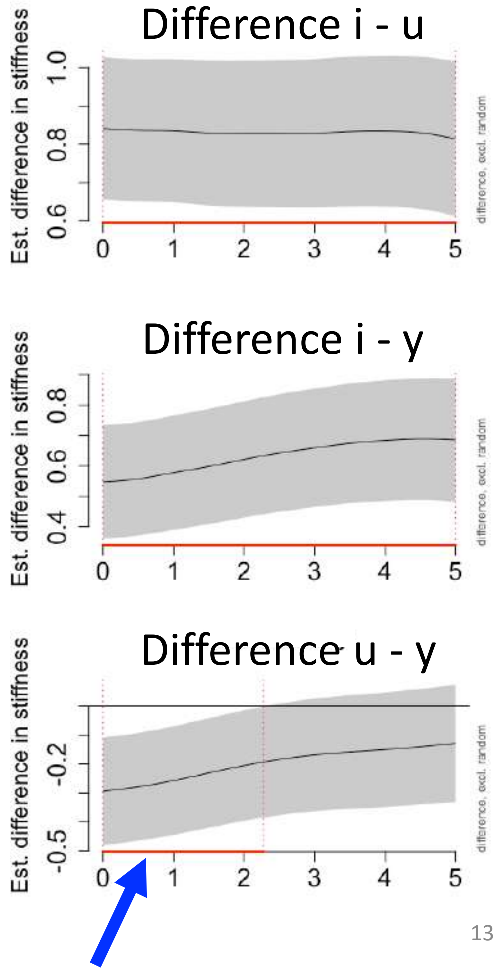


Figure 6. Results of *lower lip protrusion*

National Taiwan University

Results – Kinematic stiffness





Summary (lab speech)

- /i/ is different from /u/ and /y/ in all six measurements.
- Both /u/ and /y/ are associated with shorter lip distances, smaller lip aperture, and more protrusion.
- /u/ contrasts with /y/ *only* in lateral distance; /y/ associated with more ellipsis aperture, suggested by the axial ratio results.
- Earlier research reported different degrees of protrusion for /u/ and /y/, but no difference was found in our results.

  → Comparable protrusion from [+round].

Natural speech

Methods

- Images of /i/, /u/, /y/ from natural speech (i.e., public lectures, interviews, etc.) are used as identification stimuli.
- 168 tokens for each vowel (504 in total), collected from 12 speakers (6 female).
- Visual identification without acoustic signals.
- 3 alternative forced choice paradigm.
- 20 TM native participants (9 female).

National Taiwan University

Analyses

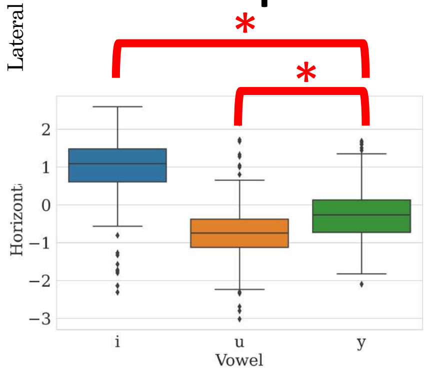
- The best 25% accurately identified stimuli are analyzed.
- Images processed by MediaPipe (each images with 468 landmarks in 3D).

1. Measurements
    - Lateral distance, vertical distance, axial ratio, and aperture area.
    - Linear Mixed Models

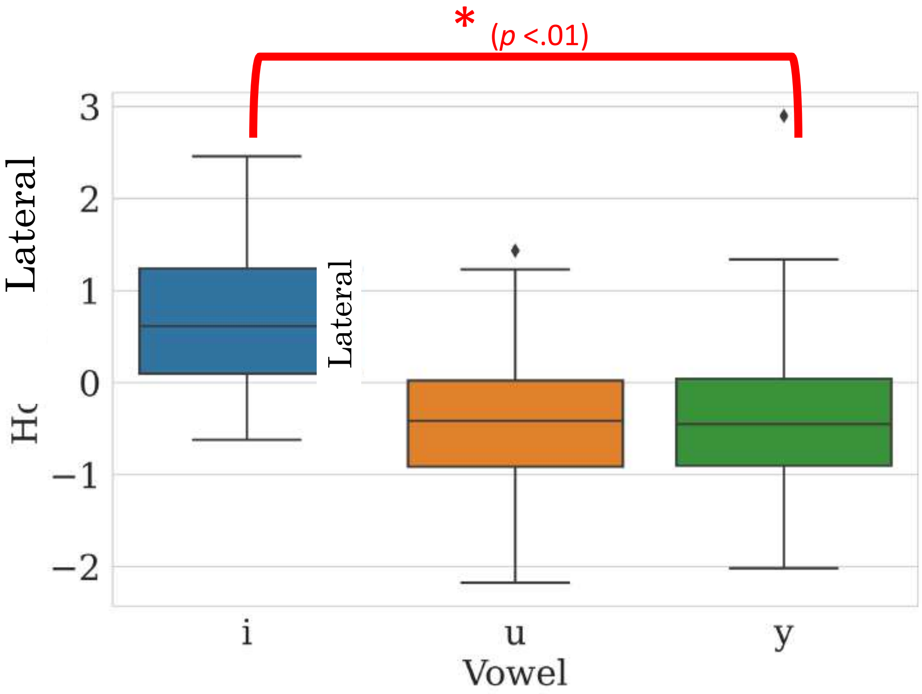
2. Machine Learning
    → Individual models for each vowel.
    → Each image 1404 (=468 landmarks *3 dimensions) datapoints.
    → Datapoints were to predict whether the image was canonical.
    - Feature importance for the 1404 dimensions was calculated and plotted, using *random forest* classifier model.

Results – Lateral distance

Lab Speech

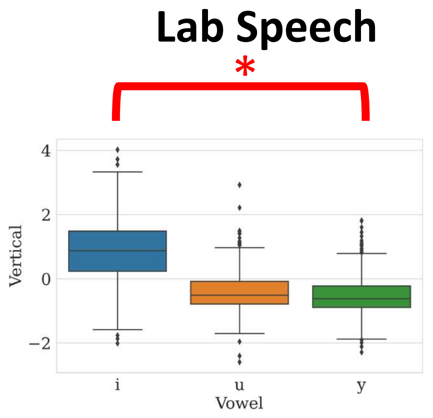


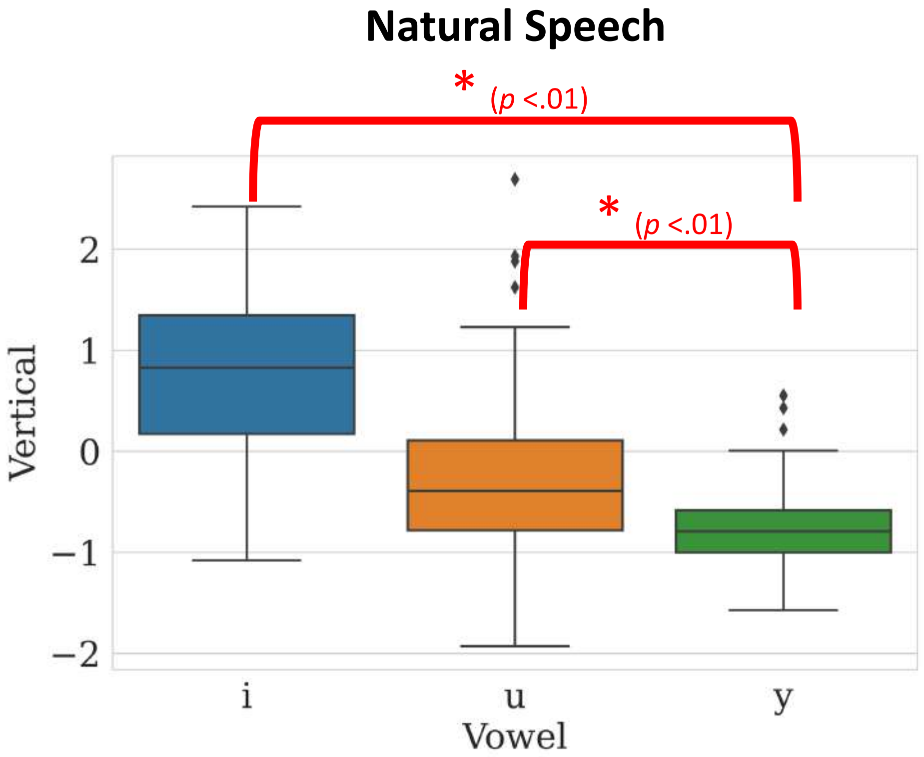
Natural Speech



National Taiwan University

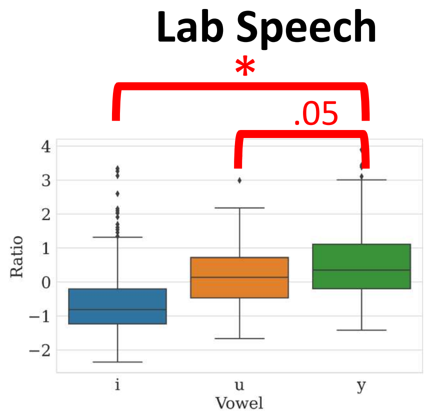
Results – Vertical distance

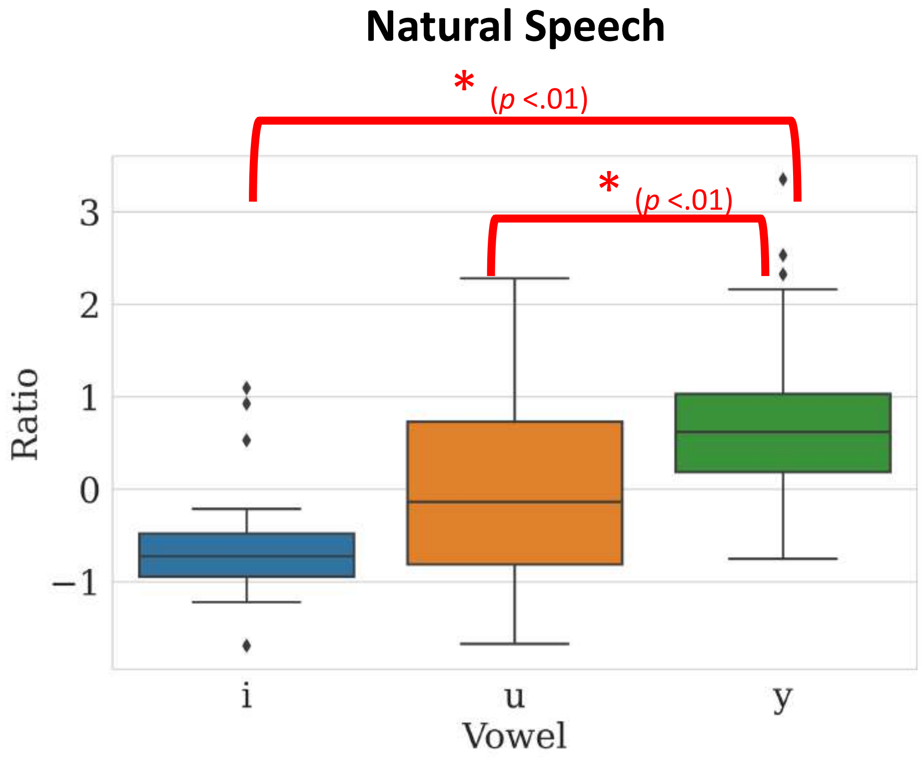




National Taiwan University

Results – Axial ratio

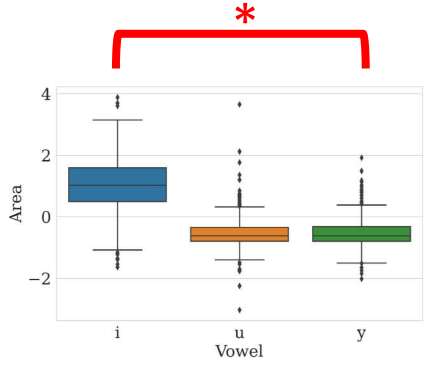




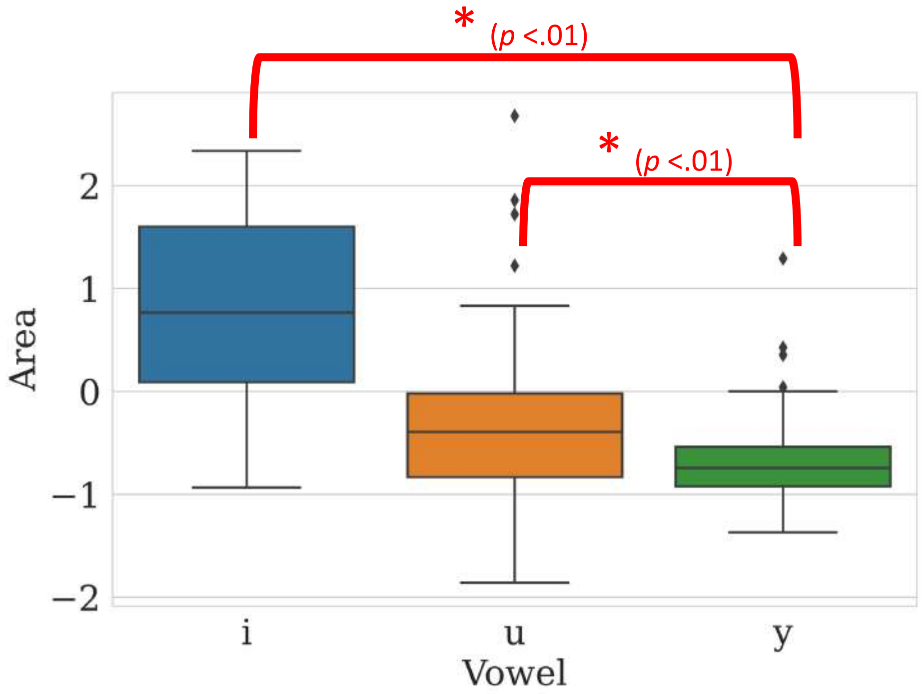
National Taiwan University

Results – Aperture area

Lab Speech

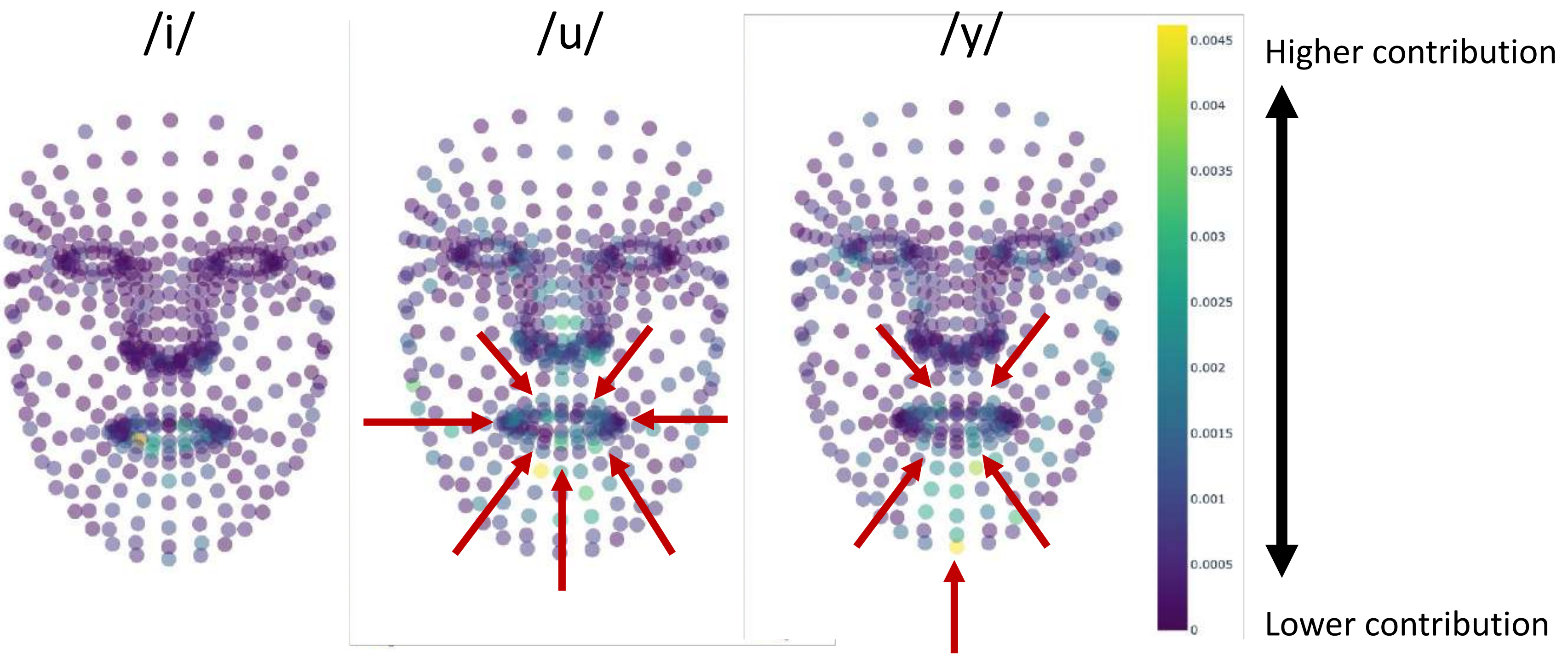


Natural Speech



National Taiwan University

Result – Random forest



National Taiwan University

Summary (natural speech)

- Natural and lab speeches are similar in:
  - /i/ different from /u/ and /y/.
  - Axial ratio: /i/ < /u/ < /y/ → circular vs. elliptical
- Variations observed between lab and natural speech:
  - Natural speech contrasts /u/ and /y/ in *vertical distance* and *aperture areas.*
  - Lab speech contrasts /u/ and /y/ in lateral distance.

- Based on machine learning results,
  - /i/ : spread lips
  - /u/ : circular rounding
  - /y/ : lip approximation with limited corner activity

  → Asymmetry between machine learning and stiffness results.

National Taiwan University

General Discussion

National Taiwan University

General discussion

RQ1: What are the lip postural differences in Taiwan Mandarin high vowels?

- /i/ is distinct from the other two.
- /u/ and /y/ are best distinguished in axial ratio of the aperture, and kinematic stiffness, but not in protrusion (cf. Catford 1988).

→ [+round] may be associated more with lip protrusion than other postures.

→ /u/ and /y/ are visually distinct, and are associated with different muscular compositions, as supported by Mayer et al. (2021).

National Taiwan University

General discussion

RQ2: Are these differences consistently realized in production?

- For /i/, yes.
- For /u/ and /y/, lab speech and natural speech demonstrated different patterns in different measurements.

National Taiwan University

General discussion

RQ3: To what degree automatic tracing + machine learning lend supports to physiological bases and their role in production?

- Asymmetry between machine-detected cues and human-approached parameters.
- Machine learning results may indicate that the production involves coordinating physiological structures (e.g., lip approximation but no corner centering for /y/), and the postures we measured might offer a limited perspective.

National Taiwan University

General discussion

Beyond research questions:

- Asymmetry between detected cues and measured parameters may therefore allow variations in production (cf. Gick & Stavness, 2013; Gick et al. 2020).
- The weighting between kinematic postures may not necessarily be equal or fixed; may vary over time.

  → The distinction between /y/ and /u/ may weigh more in the dimension of tongue positions.

Acknowledgement

National Taiwan University

This work is supported by the research grant awarded to the first author (MOST 111-2410-H-002 -049 -MY2). Our thanks also go to other research assistants (Jian-zhi Huang and Chi-wei Wang) at the Speech Behavior and Science Lab, NTU.

SBS Lab

Thank you!

References

Catford, J. C. (1988). A Practical Introduction to Phonetics. Oxford: Clarendon Press.

Gick, B., & Stavness, I. (2013). Modularizing speech. *Frontiers in psychology, 4*, 977.

Gick, B., Mayer, C., Chiu, C., Widing, E., Roewer-Després, F., Fels, S., & Stavness, I. (2020). Quantal biomechanical effects in speech postures of the lips. *Journal of neurophysiology, 124*(3), 833-843.

Jackson, M. T.-T., McGowan, R. S. (2012). A study of high front vowels with articulatory data and acoustic simulations. *JASA,* 131, 3017–3035.

Ladefoged, P., Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford, England: Blackwell.

Lin, Y. H. (2007). *The Sounds of Chinese*. Cambridge University Press.

Linker, W. (1982). Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study. *UCLA Working Papers in Phonetics* 56.

Lisker, L., & Rossi, M. (1992). Auditory and visual cueing of the [±rounded] feature of vowels. *Language and speech, 35*(4), 391-417.

Mayer, C., Chiu, C., Gick, B. (2021). Biomechanical simulation of lip compression and spreading. *Canadian Acoustics, 49*(3), 38 – 39.

Montgomery, A. A., & Jackson, P. L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. *JASA*, *73*(6), 2134-2144.

Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *JASA, 112*(4), 1627-1641.

Saito, H. (2016). Lip movements for an unfamiliar vowel: Mandarin front rounded vowel produced by Japanese speakers. *JSLHR, 59*(6), S1558-S1565.

Van Lieshout, P. H., Bose, A., Square, P. A., & Steele, C. M. (2007). Speech motor control in fluent and dysfluent speech production of an individual with apraxia of speech and Broca's aphasia. *Clinical Linguistics & Phonetics, 21*(3), 159-188.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *J. of Phon. 70*, 86-116.

National Taiwan University

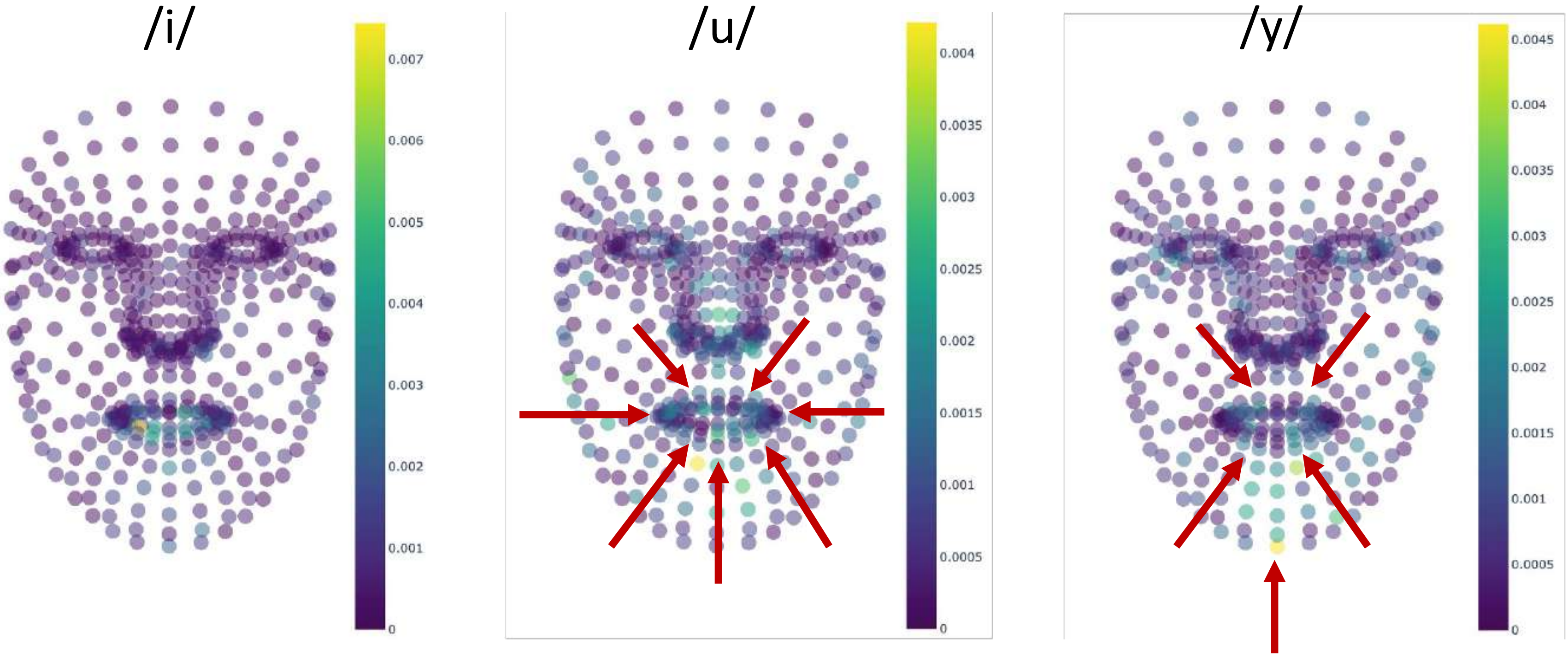
Modularized speech

	Lip protrusion (Linker, 1982) (Catford, 1988) (Saito, 2016)	Lateral pull (Linker, 1982) (Catford, 1988)	Lip aperture (Linker, 1982) (Montgomery & Jackson, 1983) (Mayer et al., 2021)	Lip approximation (Linker, 1982) (Jackson & McGowan, 1992) (Catford, 1988) (Saito, 2016)
/i/	(unspecified)	spread	substantial	less affected
/u/	obvious	centered	limited; circular	most substantial
/y/	some but limited; varies across lngs.	constrained	≧ aperture for /i/; varies across lngs; elliptical	substantial; varies across lngs.

- Speech as neuromuscular modules, allowing biomechanical saturation effects (Gick & Stavness, 2013; Gick et al., 2020).
- Any dependency among these kinematic postures?

National Taiwan University

Result – Random forest



General discussion

National Taiwan University

- Muscular compositions in a speech module may automatically yield the lip postures for the intended vowel.
  - → Variations and overlapped lip postures between /y/ and /u/ may suggest that these measurements may not be the defining postures between the two vowels.
  - →
  - → How to quantify the dependency across these kinematic postures?

Appendix

Measurement	model	AIC	BIC	p	sig
LL protrusion	w/o gender	1929.6	2014.4	0.0156	*
	w/ gender	1926	2040.6		
UL protrusion	w/o gender	2116.5	2201.3	0.5574	
	w/ gender	2123.6	2238.3		
Vertical distance	w/o gender	2133.3	2218	0.1671	
	w/ gender	2136.2	2250.8		
Lateral distance	w/o gender	1636.4	1721.2	0.2881	
	w/ gender	1641.1	1755.7		
Axial ratio	w/o gender	2343.1	2427.8	0.0668	.
	w/ gender	2343.3	2457.9		
Aperture area	w/o gender	1919.8	2004.5	0.2298	
	w/ gender	1923.7	2038.3		

Future work

- Identification, goodness rating, etc.
- Multimodal perception (congruent vs. incongruent inputs).
- Lip perturbation (ICPhS poster ID #656).
- EMG activity during the production of TM high vowels.

National Taiwan University

Lip postures

- The difference between /i/ and /y/ lies in vertical distance (i.e., lip approximation) (Jackson & McGowan, 2012) .
- Inner border (i.e., lip aperture) of lips for different vowels also vary from one another (Montgomery and Jackson, 1983) .

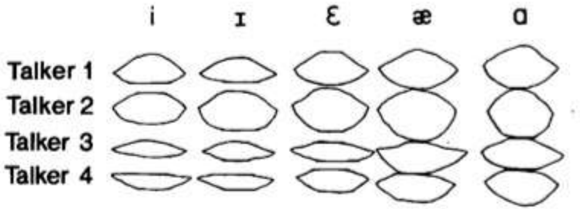


FIG. 2. Tracings of the inner lip border of vowel tokens used as lipreading stimuli. Images from four talkers are shown and represent the digitized tracings of the videotape field chosen to represent the vowel maximum for each token of the vowels /i, ɪ, ɛ, æ, ɑ/.

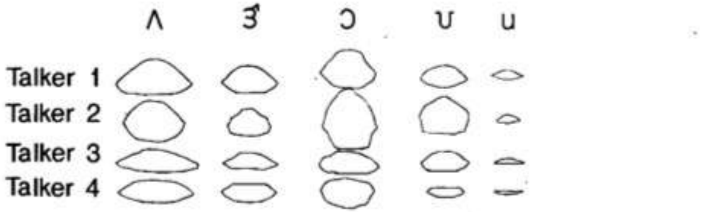
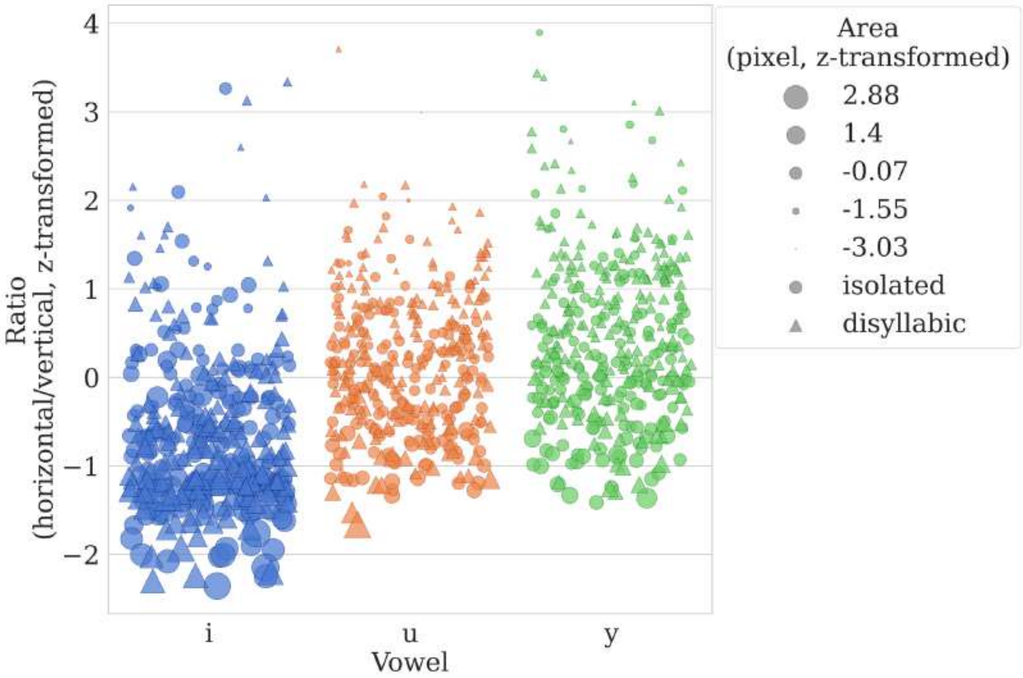
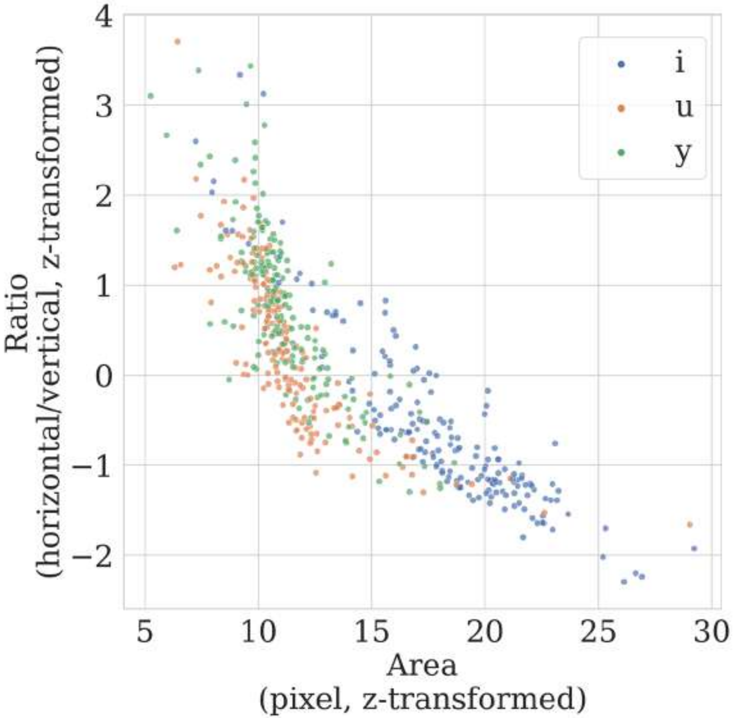


FIG. 3. Tracings of the inner lip border of vowel tokens used as lipreading stimuli. Images from four talkers are shown and represent the digitized tracings of the videotape field chosen to represent the vowel maximum for each token of the vowels /ʌ, ɝ, ɔ, ʊ, u/.


Area
(pixel, z-transformed)
2.88
1.4
-0.07
-1.55
-3.03
isolated
disyllabic
Ratio
(horizontal/vertical, z-transformed)
Vowel
i
u
y


i
u
y
Ratio
(horizontal/vertical, z-transformed)
Area
(pixel, z-transformed)