Reconciling perception with production in Southern speech

Rachel M. Olsen, Joseph A. Stanley, Michael L. Olsen,
Lisa Lipani, & Margaret E. L. Renwick

The American Dialect Society
January 6, 2019

Southern diphthong weakening

- Weakening of canonical /aɪ ɔɪ aʊ/ occurs in Southern speech (Thomas 2003)
- /ai/ weakening
  - Most prevalent
  - Triggering feature of the Southern Vowel Shift (SVS) (Labov, Ash & Boberg 2006)
- /ɔɪ/ weakening
  - Most prevalent amongst African Americans, and older European Americans in the South
  - For everyone before laterals (Thomas 2008)
- /aʊ/ weakening
  - Widespread in European American Southern English (Thomas 2008)

Transcription & Production

- Many studies have compared transcriptions by multiple speakers
  - Henderson (1938), Ladefoged (1960)
  - Buckeye Corpus vowels had 74% agreement. (Pitt et al. 2005)
  - Read RP had 83% agreement (Eisen 1993)
  - Transcriptions are affected by transcriber and speech variables (Cucchiarini 1993)
- Fewer have compared transcriptions to production data.
  - Kerswell & Wright (1990) conclude that reliability is not certain.
- Sometimes Linguistic Atlas transcriptions are questioned
  - Johnson (2010:28–29) flat out doubts some LANE transcriptions
  - Regarding the LOT-THOUGHT merger, Mouton (1968:464) says the LANE fieldworkers were “hopelessly and humanly incompetent”.

This Study

Research question

How does perception compare to production in southern speech in Linguistic Atlas data?

Hypothesis

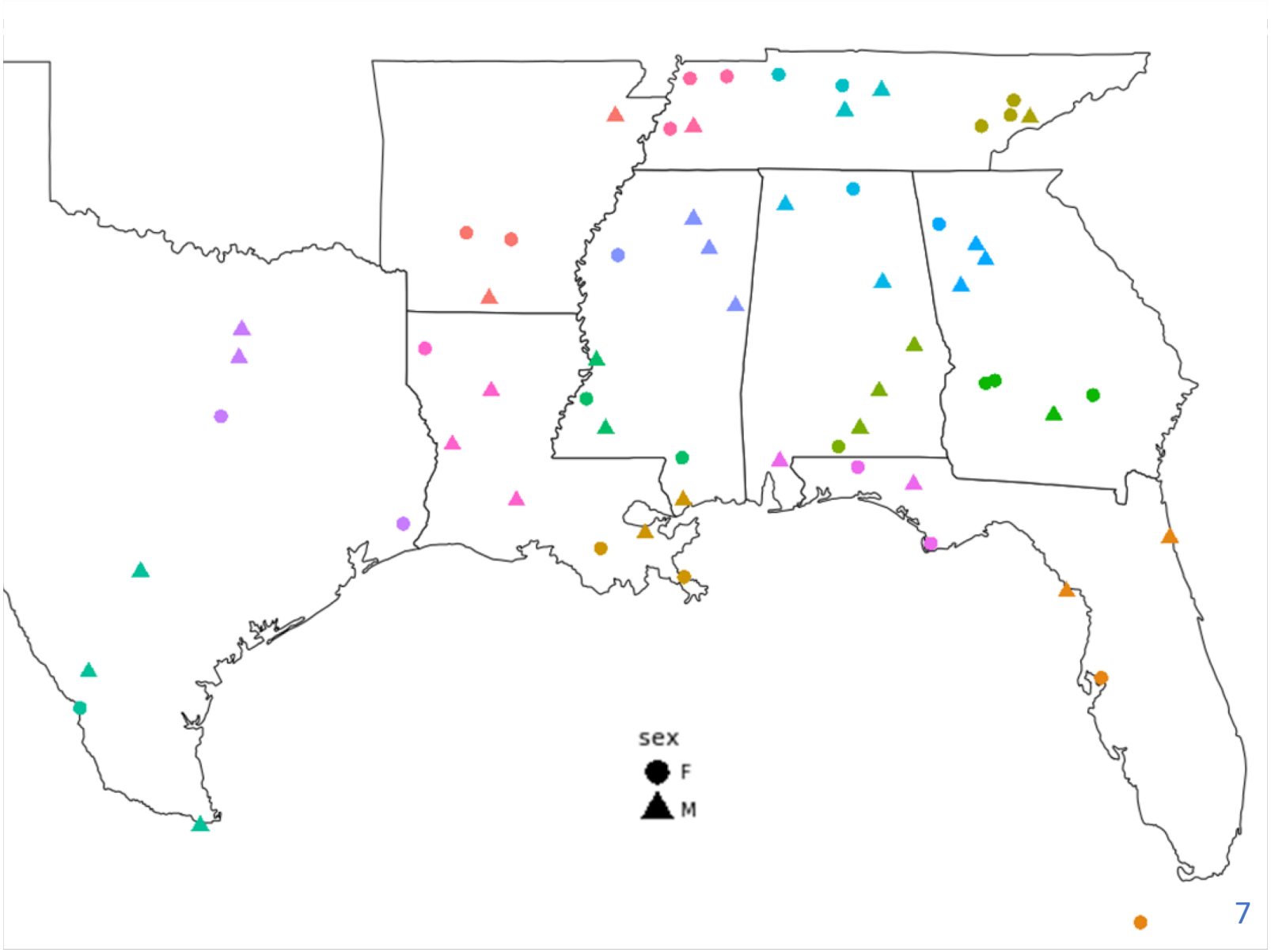
We expect impressionistic glide weakening to correlate with less dynamic vowels.

Methods

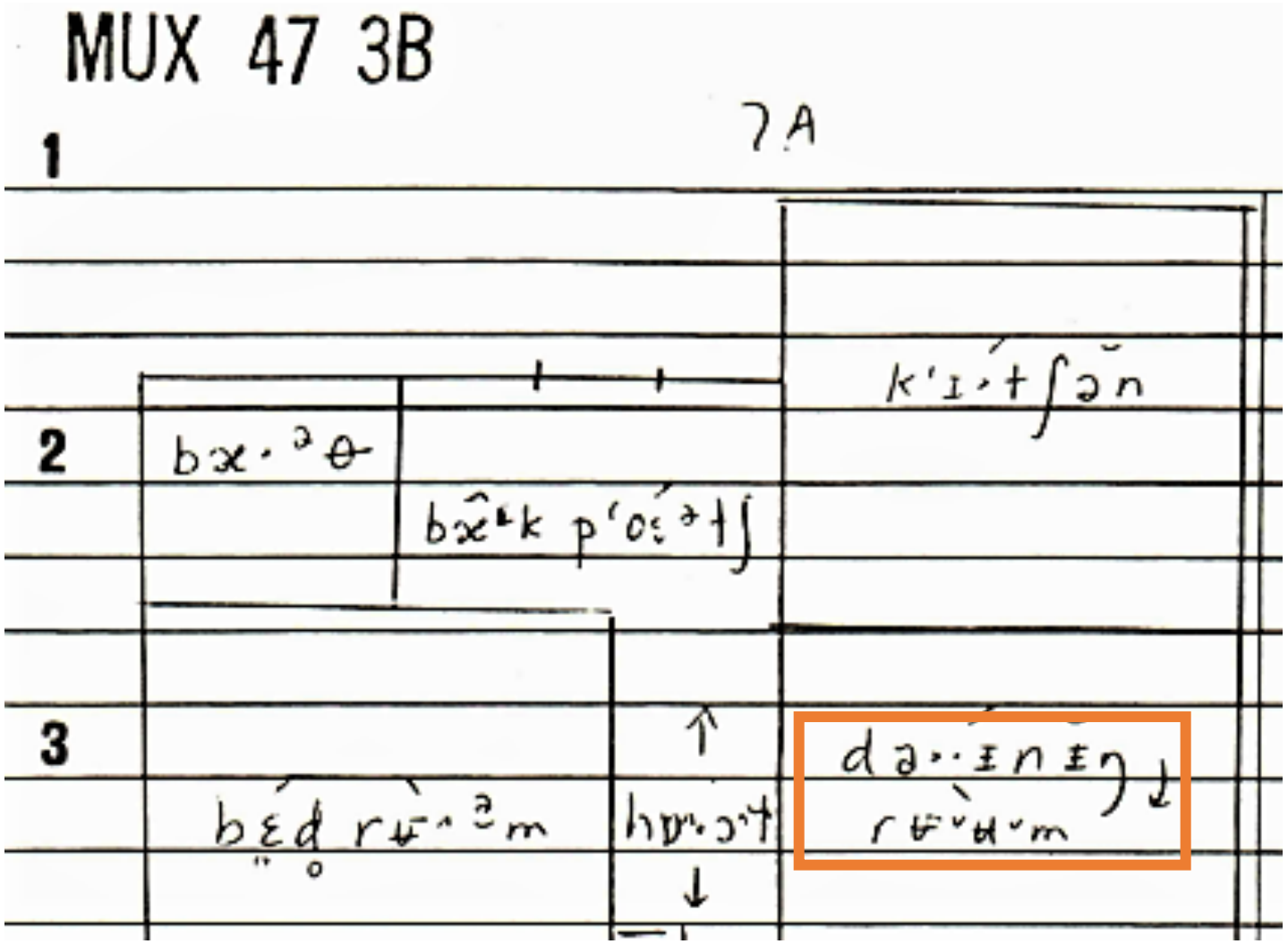
The Digital Archive of Southern Speech (DASS)

- Sociolinguistic audio corpus; subset of Linguistic Atlas of the Gulf States (Pedersen *et al*. 1986, Kretzschmar *et al.* 2013, Olsen *et al.* 2017)
- 64 speakers (30F)
  - Born 1886–1965 (μ = 61 years old)
  - Recorded 1970–1983
- 367 hours of audio (2.5–10 hours per interview; μ = 5.75 hours)
- 4 speakers for each of 16 geographical sectors
  - 1 African American (AA) speaker
  - 3 European American (EA) speaker "DASS Types"
    - Type I: Folk
    - Type II: Common
    - Type III: Cultured

Geographic distribution of DASS speakers

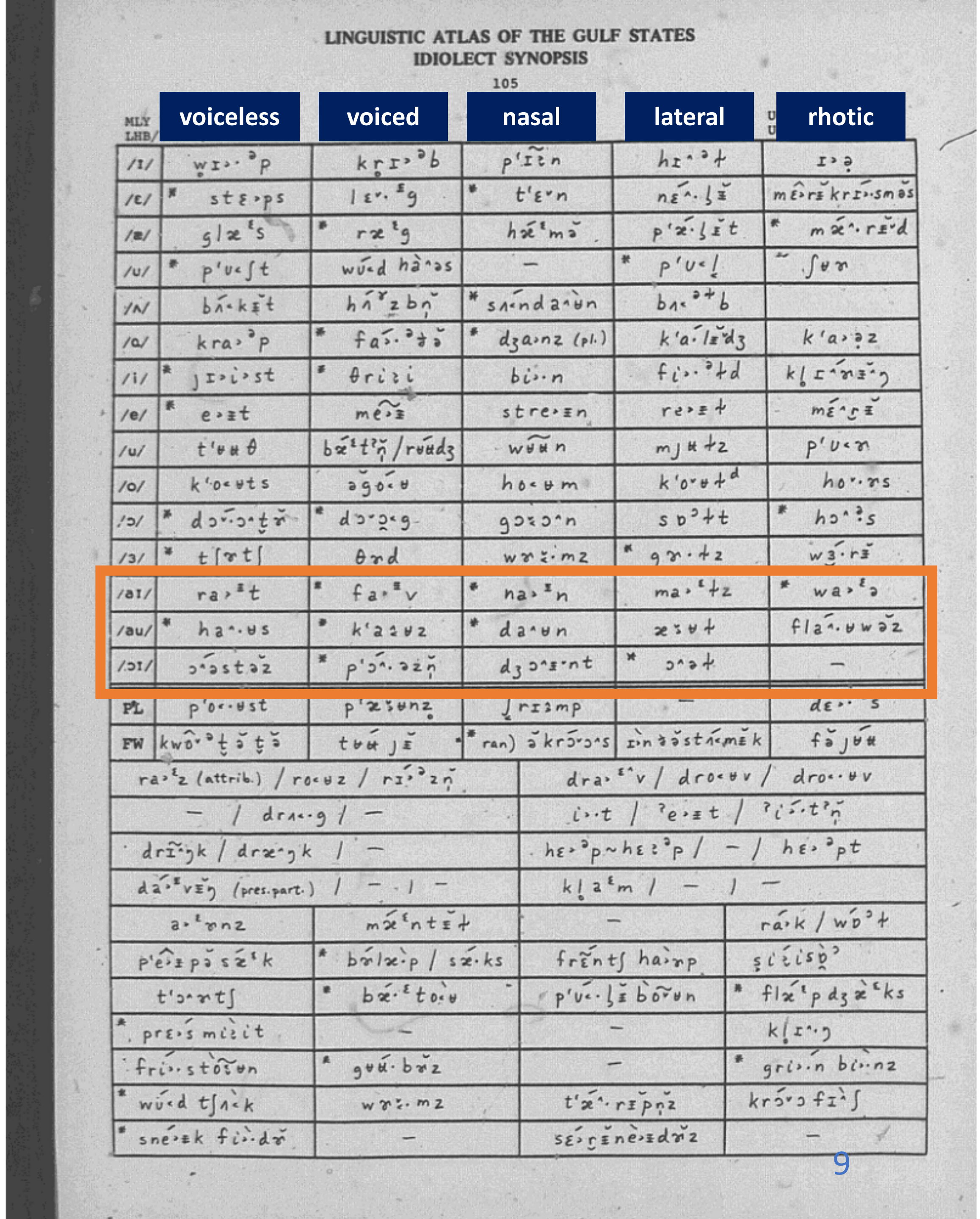


DASS Protocols



Idiolect Synopsis

- Summary of target phonemes in various environments for each speaker
- Impressionistic, and thus a record of perception



Methods

Compare:

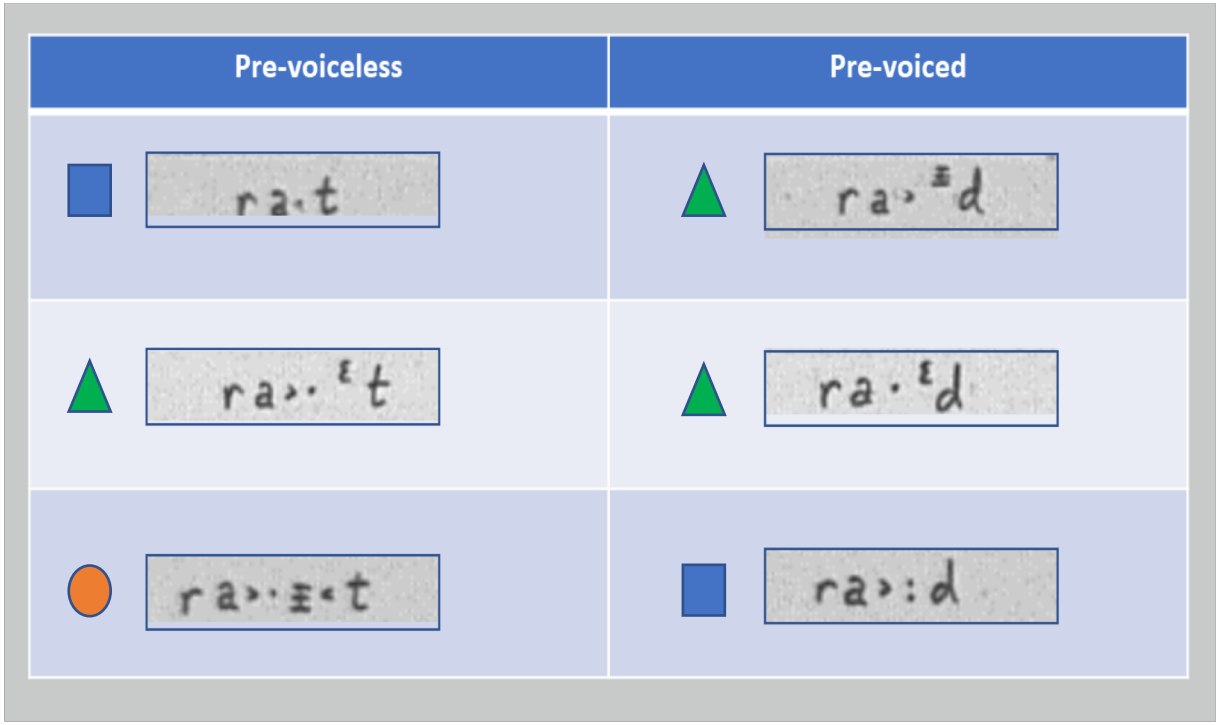
Impressionistic transcriptions from Idiolect Synopses.

63 speakers: 1,323 tokens

Acoustic data from the same speakers.

107,854 tokens

Impressionistic/Perception Analysis



- Calculated percentage of monophthongized diphthongs for each speaker.

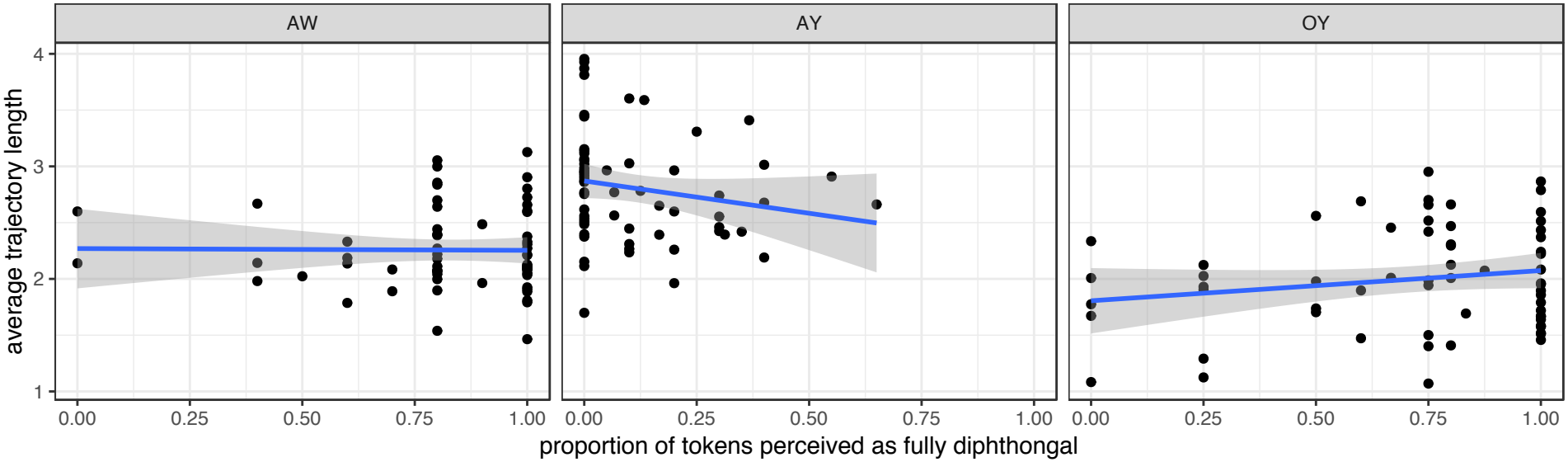
Acoustic/Production Analysis

- Data Processing
  - Fully transcribed DASS interviews force-aligned using DARLA (Reddy & Stanford 2015)
  - F1 and F2 extracted at five time points for stressed tokens of /aɪ ɔɪ aʊ/ (Rosenfelder et al. 2014)

- Trajectory length used to quantify diphthongization (Fox & Jacewicz 2009; Farrington *et al.* 2018)
  - Method
    - Composite measurement of F1 and F2 length between points in the vowel trajectory
    - 20%, 35%, 50%, 65%, 80%
    - Captures the amount of vowel movement across time
  - More dynamic vowels (i.e. diphthongs) have longer TL

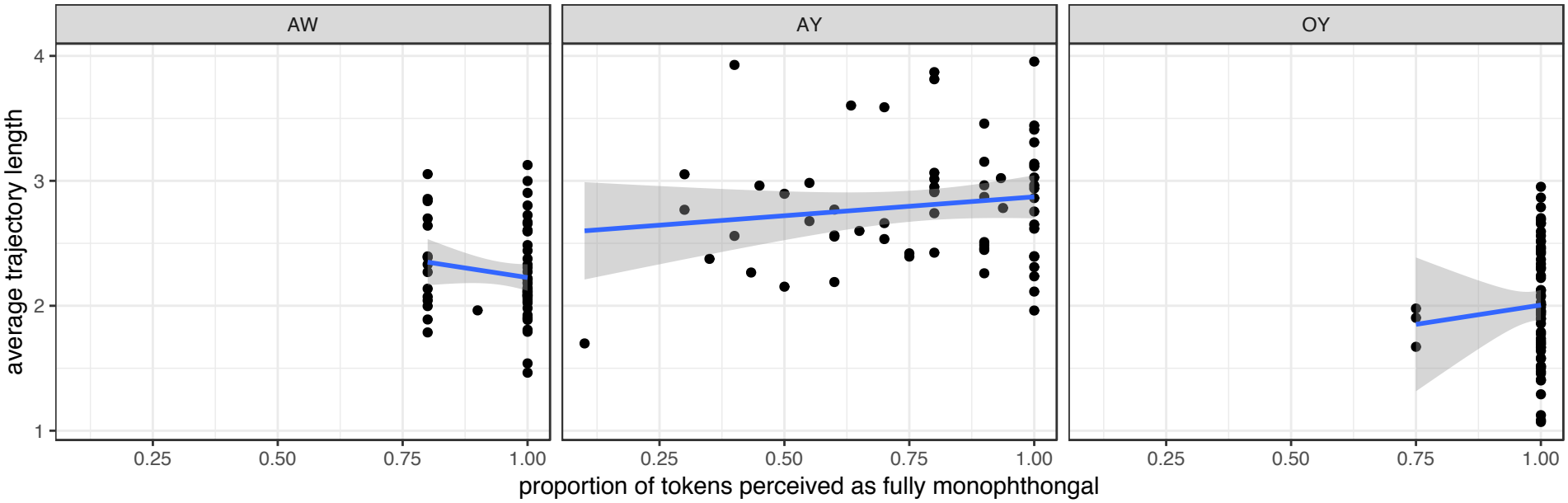
Results

Trajectory length by percentage of tokens transcribed as fully diphthongal



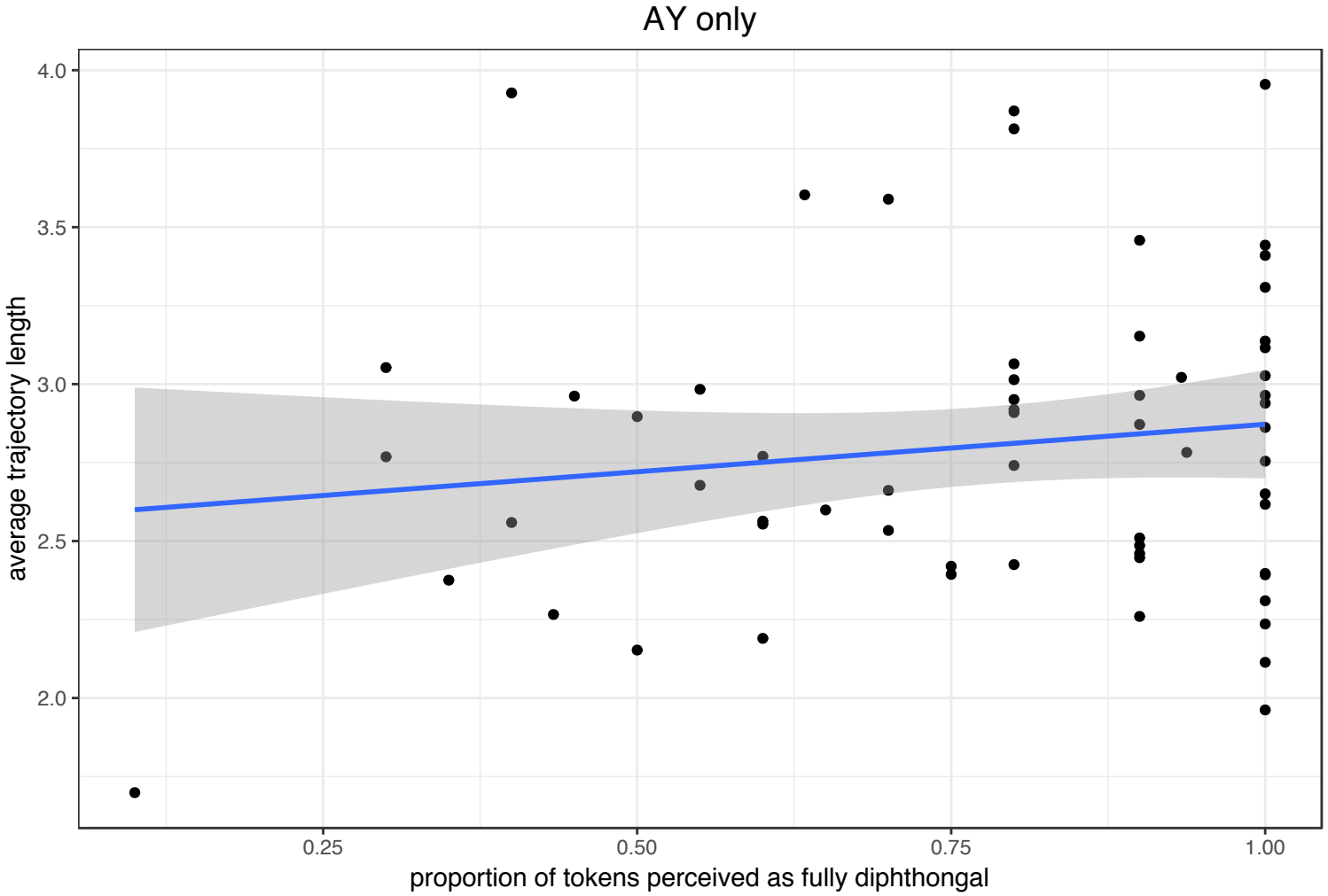
Monophthong = one vowel in transcription or superscripted offglide.
Diphthong = full vowel as offglide

Trajectory length by percentage of tokens transcribed as fully monophthongal



Monophthong = one vowel in transcription.
Diphthong = offglide as full vowel or superscript

Trajectory length by percentage of tokens transcribed as fully diphthongal



Monophthong = one vowel in transcription or superscripted offglide.
Diphthong = full vowel as offglide

Mixed Modeling
Perceptual/Transcribed Data

- Dependent variable: Percentage of tokens transcribed as fully monophthongal
- Random effect: Speaker
- Fixed effects: Phonological environment, Default=voiceless
  Sex, Default=Female
  Ethnicity, Default=African American
  Year of birth

- Phonological environment significant
  - Rhotics transcribed as monophthongal more
- No social factors were significant

	β	*p*
(Intercept)	-0.087	0.962
Lateral	0.042	0.461
Nasal	0.023	0.688
Rhotic	0.247	< 0.001***
Voiced	-0.024	0.669

Mixed Modeling
Production/Acoustic Data

- Same as previous, only trajectory length is dependent variable

- Phonological environment significant
  - Laterals (not rhotics) most weakened

- European Americans have significantly longer TL (i.e. more diphthongal production) than African Americans

	B	*p*
(Intercept)	-0.254	0.944
Lateral	-0.137	0.043*
Nasal	0.235	< 0.001***
Rhotic	0.193	0.011*
Voiced	0.149	0.027*
Voiceless	0.144	0.034*
European American	0.253	0.007**

Discussion and Conclusion

Discussion

- Acoustic and perceptual data also tell different stories.
  - Perceptual data suggest:
    - Social factors are not significant
    - Pre-rhotic glides are the most weakened
  - Production data suggest:
    - Ethnicity is significant
    - Pre-lateral glides are the most weakened
- As percentage of monophthongal perception goes up, Trajectory Length goes down, as expected, but the correlation is VERY weak.
- Trajectory Length is not necessarily reflective of impressionistic transcriptions of glides in DASS.

Conclusion

- The acoustic correlates to perception are not always straightforward.

- Production and perception must be considered in concert with one another.

- Future considerations:
  - Other potential acoustic measures of glide weakening, such as trajectory shape
  - A more fine-grained analysis of the Idiolect Synopsis transcriptions
    - Take into account the proposed second vowel of transcribed diphthongs, and weak diphthongs, as well as triphthongization

References

Bassett, Marvin Winslett, Susan Leas McDaniel & Lee Pederson. 1986. *Linguistic atlas of the Gulf States. [microform :] a concordance of basic materials.* Ann Arbor, Mich. : University Microfilms International, 1986.

Cucchiarini, Catia. "Phonetic Transcription: A Methodological and Emperical Study." 1993. Ph.D. Dissertation, Katholieke Universiteit Nijmegen.

Eisen, Barbara. "Reliability of Speech Segmentation and Labelling at Different Levels of Transcription." *EUROSPEECH'93*, 1993, 673–76.

Farrington, Charlie, Tyler Kendall, and Valerie Fridland. "Vowel Dynamics in the Southern Vowel Shift." *American Speech* 93, no. 2 (May 1, 2018): 186–222. https://doi.org/10.1215/00031283-6926157.

Fox, Robert Allen & Ewa Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America* 126(5). 2603–2618. doi:10.1121/1.3212921.

Johnson, Daniel Ezra. *Stability and Change Along a Dialect Boundary: The Low Vowels of Southeastern New England*. Publication of the American Dialect Society 95. Durham, NC: Duke University Press, 2010.

Kerswill, Paul, and Susan Wright. "The Validity of Phonetic Transcription: Limitations of a Sociolinguistic Research Tool." *Language Variation and Change* 2, no. 3 (October 1990): 255–75. https://doi.org/10.1017/S0954394500000363.

Kretzschmar, William A., Paulina Bounds, Jacqueline Hettel, Lee Pederson, Ilkka Juuso, Lisa Lena Opas-Hänninen & Tapio Seppänen. 2013. The Digital Archive of Southern Speech (DASS). *Southern Journal of Linguistics* 37(2). 17–38.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The Atlas of North American English: Phonetics, phonology, and sound change: A multimedia reference tool*. Berlin; New York: Mouton de Gruyter.

Moulton, William G. "Structural Dialectology." *Language* 44, no. 3 (1968): 17.

Olsen, Rachel M., Michael Olsen, Joseph A. Stanley, Margaret E. L. Renwick, & William A. Kreztschmar, Jr. 2017. Methods for transcription and forced alignment of a legacy speech corpus. Proceedings of Meetings on Acoustics 30, 060001; doi: http://dx.doi.org/10.1121/2.0000559.

Pederson, Lee. 1981. *Protocol of a field record for the linguistic atlas of the Gulf States*. Ann Arbor, MI: University Microfilms International.

Pederson, L., McDaniel, S. L., and Adams, C. M. (eds.). 1986. *Linguistic Atlas of the Gulf States*, University of Georgia Press, Athens, Georgia, Vols. 1–7.

Pitt, Mark A., Keith Johnson, Elizabeth Hume, Scott Kiesling, and William Raymond. "The Buckeye Corpus of Conversational Speech: Labeling Conventions and a Test of Transcriber Reliability." *Speech Communication* 45, no. 1 (January 1, 2005): 89–95. https://doi.org/10.1016/j.specom.2004.09.001.

Reddy, Sravana & James N. Stanford. 2015. Toward completely automated vowel extraction: Introducing DARLA. *Linguistics Vanguard* 1(1). 15–28. doi:https://doi.org/10.1515/lingvan-2015-0002.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gormon, Hilary Prichard & Jiahong Yuan. 2014. FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2. https://github.com/JoFrhwld/FAVE.

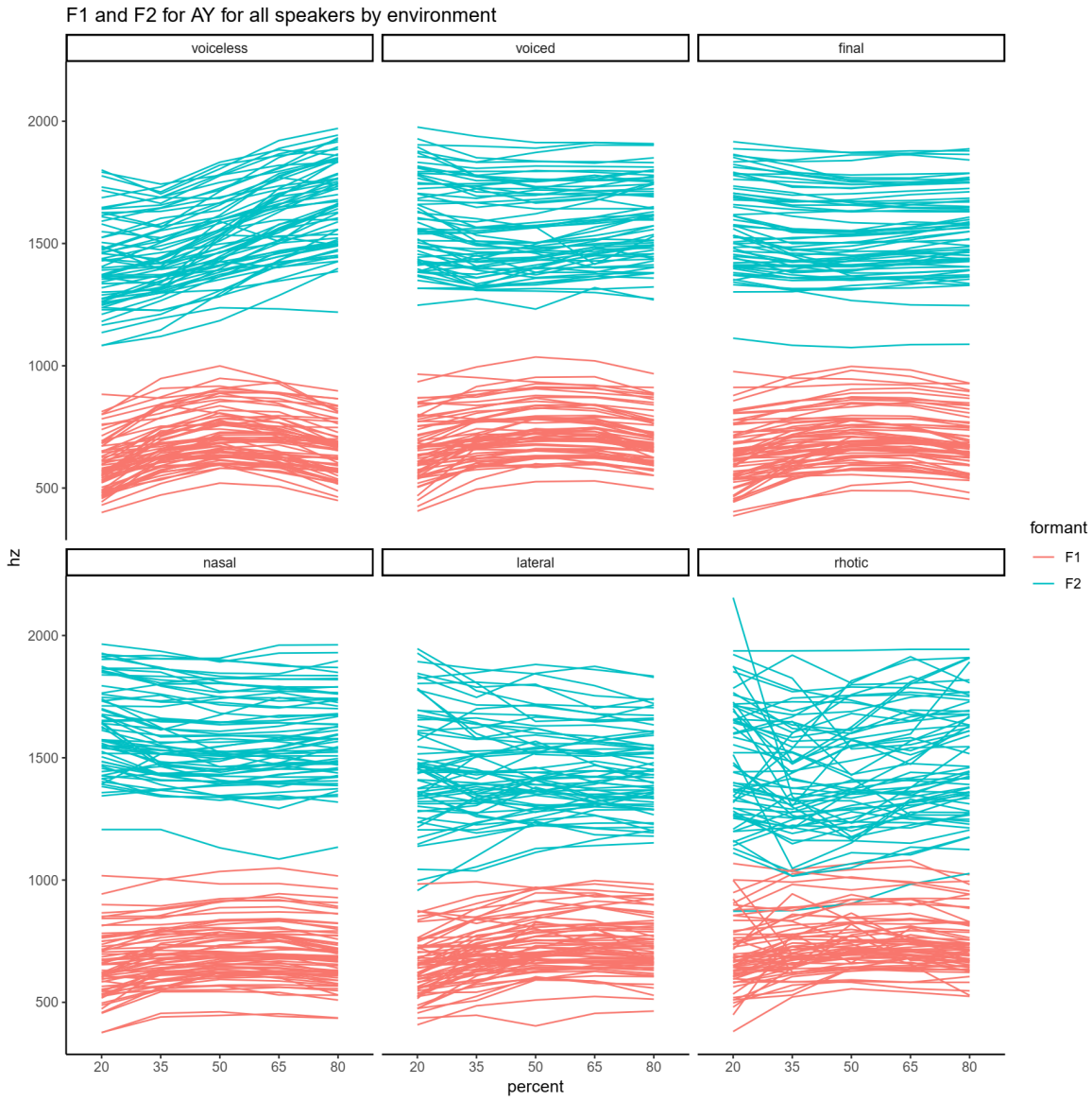
Thomas, Erik R. 2003. Secrets Revealed by Southern Vowel Shifting. *American Speech* 78(2). 150–170. doi:https://doi.org/10.1215/00031283-78-2-150.

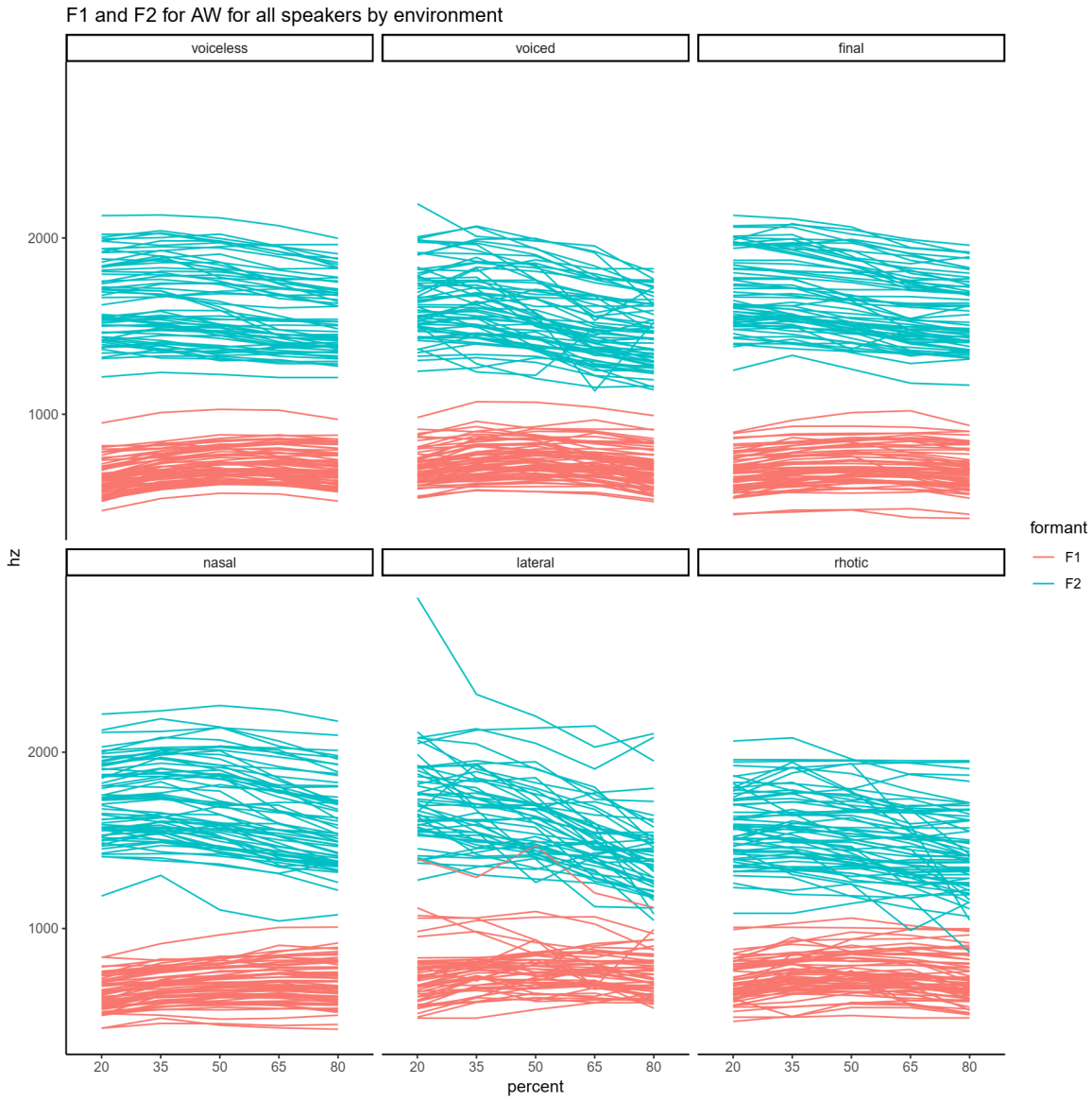
Thomas, Erik R. 2008. Rural Southern White Accents. In Edgar W. (ed. Schneider introds., and summaries) & Bernd (introd.) Kortmann (eds.), *Varieties of English, 2: The Americas and the Caribbean*, 87–114. Berlin, Germany: Mouton de Gruyter.

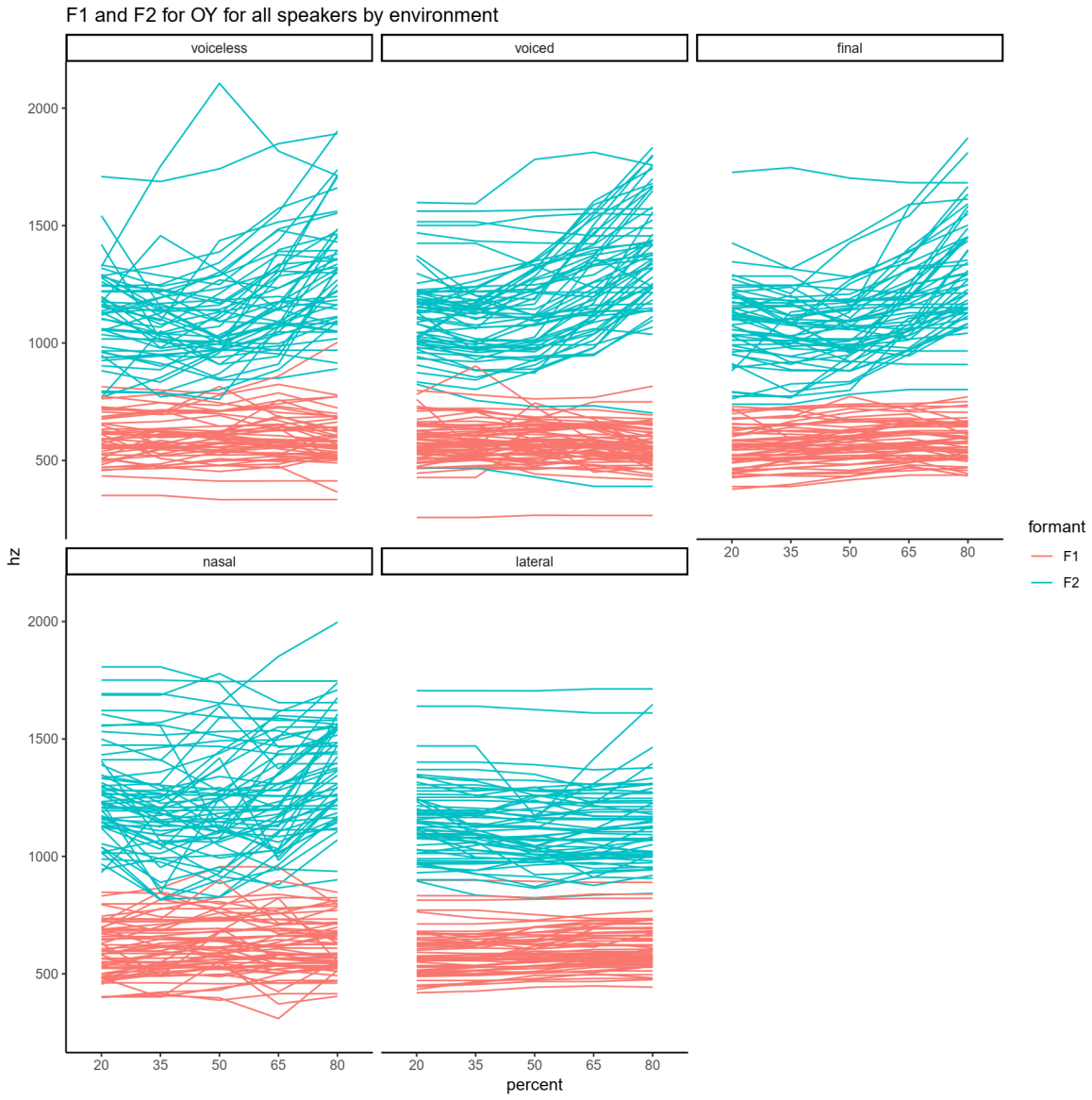
Thank You!

- This research was funded by a grant from the National Science Foundation (BCS1625680), co-PIs Dr. Margaret E. L. Renwick and Dr. William Kretzschmar.

- Download these slides at joeystanley.com/ads2019


F1 and F2 for AY for all speakers by environment
voiceless
voiced
final
nasal
lateral
rhotic
hz
2000
1500
1000
500
20
35
50
65
80
percent
formant
F1
F2


F1 and F2 for AW for all speakers by environment
voiceless
voiced
final
nasal
lateral
rhotic
hz
2000
1000
20
35
50
65
80
percent
formant
F1
F2


F1 and F2 for OY for all speakers by environment
voiceless
voiced
final
nasal
lateral
hz
2000
1500
1000
500
20
35
50
65
80
percent
formant
F1
F2