Seoul International Conference on Speech Sciences 2017
10-11 November, 2017

Psychoacoustic roughness as a measure of creakiness in two dialects of Zhuang

Jeremy Perkins
(University of Aizu, Aizu-Wakamatsu, Japan)
jperkins@u-aizu.ac.jp

Seunghun J. Lee (International Christian University, Tokyo, Japan & University of Venda, South Africa)
Julián Villegas (University of Aizu, Japan)

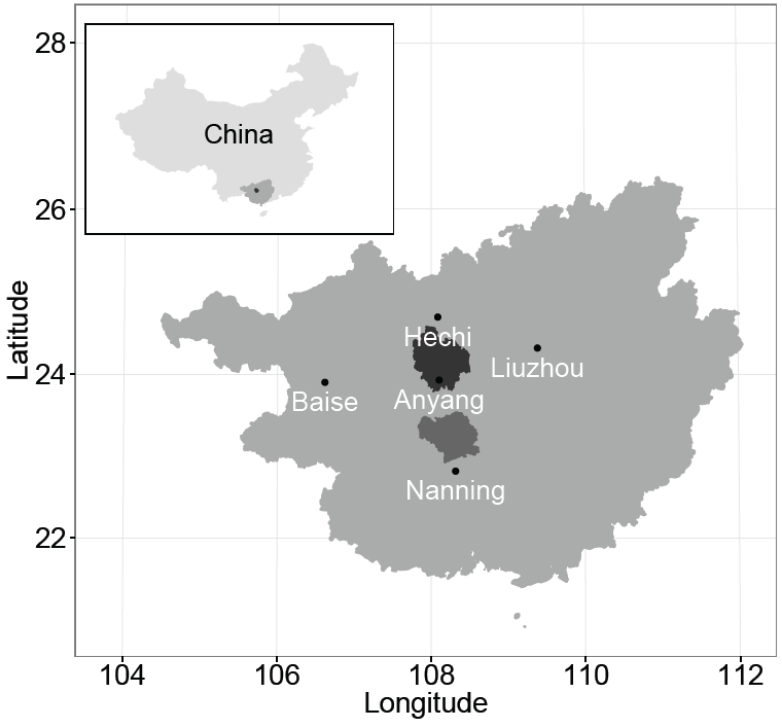
Zhuang

- Zhuang has the largest number of speakers of the 55 official minority languages in China
  - Zhuang is in the Tai-Kadai family (Thai, Laos, Vietnam, Myanmar & China)
  - The variety spoken in Wuming is considered the standard variety (Wei & Qin, 1980).
  - There is a vast degree of dialectal difference within Zhuang.
  - Many Zhuang dialects are not mutually intelligible.

This talk: 2 Zhuang dialects

- Du'an Zhuang is not mutually intelligible with Wuming Zhuang, mainly due to tonal differences.
  - We ultimately aim to describe these differences via phonetic measurement.

- A pilot study with a single Du'an Zhuang speaker suggested contrastive phonation (creaky vs. modal)
    - This study seeks to check this finding in 2 dialects, with more speakers.



DU'AN ZHUANG (DARK)
WUMING ZHUANG (LIGHT)

Wuming Zhuang Tone System (Wei & Qin 1980)

Tone	1	3	5	7 short	7 long
Chao (1930)	24	55	35	55	35
Example	[na]	[na]	[na]	[nap]	[na:p]
Gloss	'thick'	'face'	'arrow'	'to put into'	'to be stuck'
Description	rising	level	rising	level	rising
Tone	**2**	**4**	**6**	**8 short**	**8 long**
Chao	31	42	33	33	33
Example	[na]	[na]	[na]	[nap]	[na:p]
Gloss	'field'	'aunt'	'meat'	'to bind'	'to turn in tax
Description	falling	falling	level	level	level
	unchecked syllables			*checked syllables*	

- Tones and syllable structures
  - Tones 1 to 6 only occur in open syllables or with sonorant codas (*unstopped/unchecked syllables*)
  - Tones 7 & 8 only occur with obstruent codas (*stopped/checked syllables*)

Acoustic correlates of tone

- Phonation and tone (cf. Silverman 1997)
  - laryngeally complex languages
    - Oto-Manguean (e.g. Jalapa Mazatec, Garellek & Keating 2011; Triqui DiCanio 2008; Quiaviní Zapotec, Chávez-Péon 2010), Nilotic (e.g. Dinka, Andersen 1993), Burmese (Watkins 1997, 2000, Lee 2007, 2008, Gruber 2011), Green Mong (Andruski & Ratliff, 2000, Andruski 2006), Chong (DiCanio 2009), Vietnamese (Honda 2006), Black Miao (Mazaudon & Michaud 2008, Kuang 2013, Mazaudon 2014)
- Duration

Tone as a phonological system

- Some factors to consider in order to establish a tonal system
  - **Pitch (F0, in Hz)**
    - Level
    - Contour
  - **Phonation (Spectral shape, H2-H1, H1-A3 etc.)**
    - **Creaky voice**
    - Modal voice
    - Breathy voice
  - Duration (in ms)
    - Short vs. long
  - [Phonology] Syllable types
    - Unchecked CV, CVN (N= sonorants)
    - Checked CVO (O = obstruents)

Data Collection: Consultants

- Five female and three male native speakers of Wuming Zhuang and four female and two male native speakers of Du'an Zhuang, all in their 20's, were recorded in a sound attenuated booth at Guangxi University (Nanning, Guangxi) in 2015.

- At the time of the elicitation, the consultants communicated with their relatives and friends in their respective Zhuang dialect.

- In Nanning, however, the consultants mostly used a Guangxi variety of Putonghua.
  - *Putonghua = Standard Chinese*

Data Collection: Procedure

- Zhuang words were elicited using a frame sentence, presented in Chinese characters (but read in Zhuang).
  - 我正在读__这个词
  - “I am reading this word ____ now”

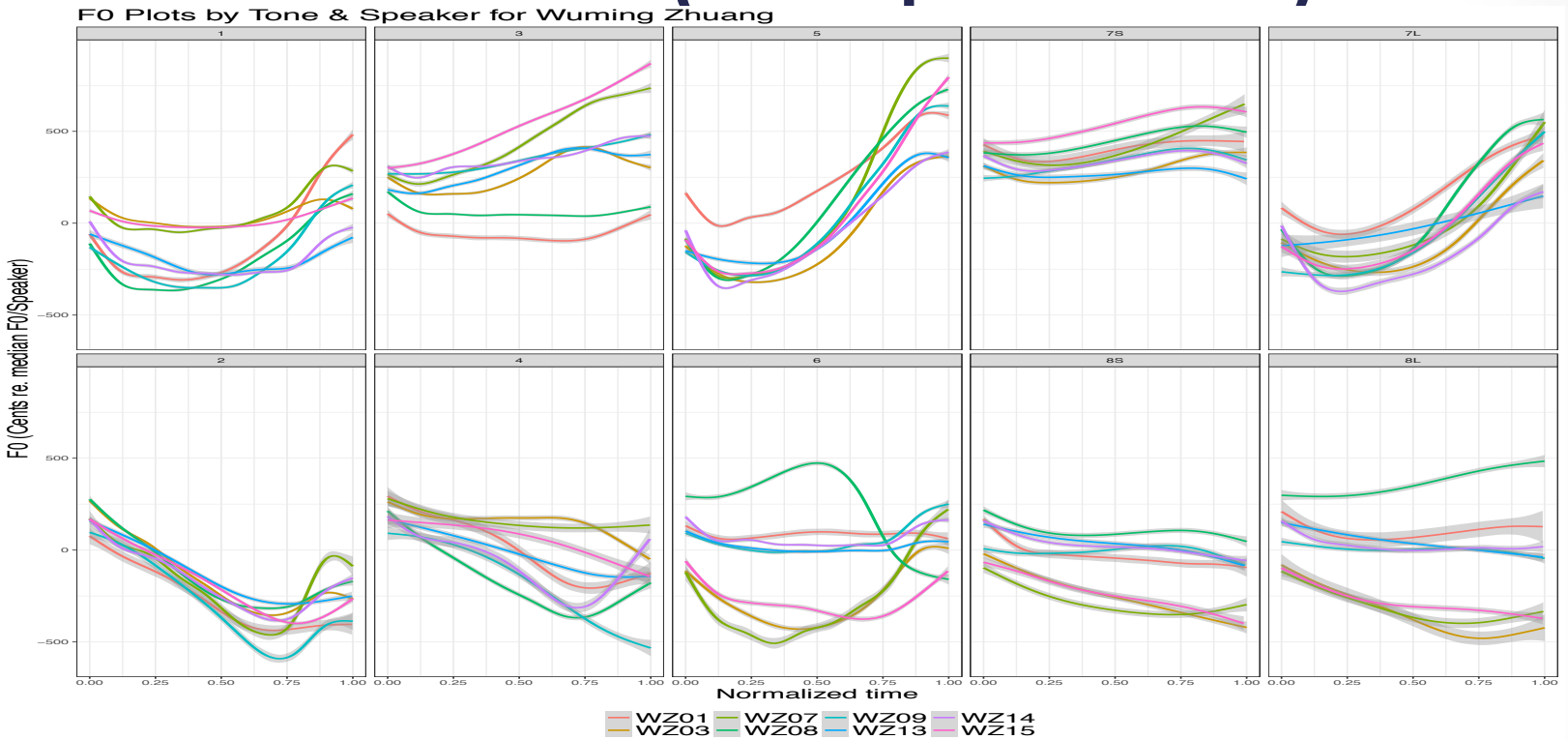
- Words in isolation were then elicited from a list of Chinese characters to exclude possible tone sandhi effects.
- The tones for these words in Wuming Zhuang and Du’an Zhuang are reported, facilitating the analysis (Wei & Qin 1980, Quosheng 1996).
  - In many cases, the tone is unknown in one dialect. We use a tone-classification procedure to assign tone to these words.

Data Collection
Tone Tokens and Exclusion criteria

- Monosyllabic words
  - Only monophthongal vowels were included
  - Onset consonants were limited to alveolar, palatal and velar obstruents
  - Nasal codas were included
    - These yielded higher amounts of creakiness in the COVAREP algorithm
  - Among checked syllables, only alveolar and velar stop codas were included

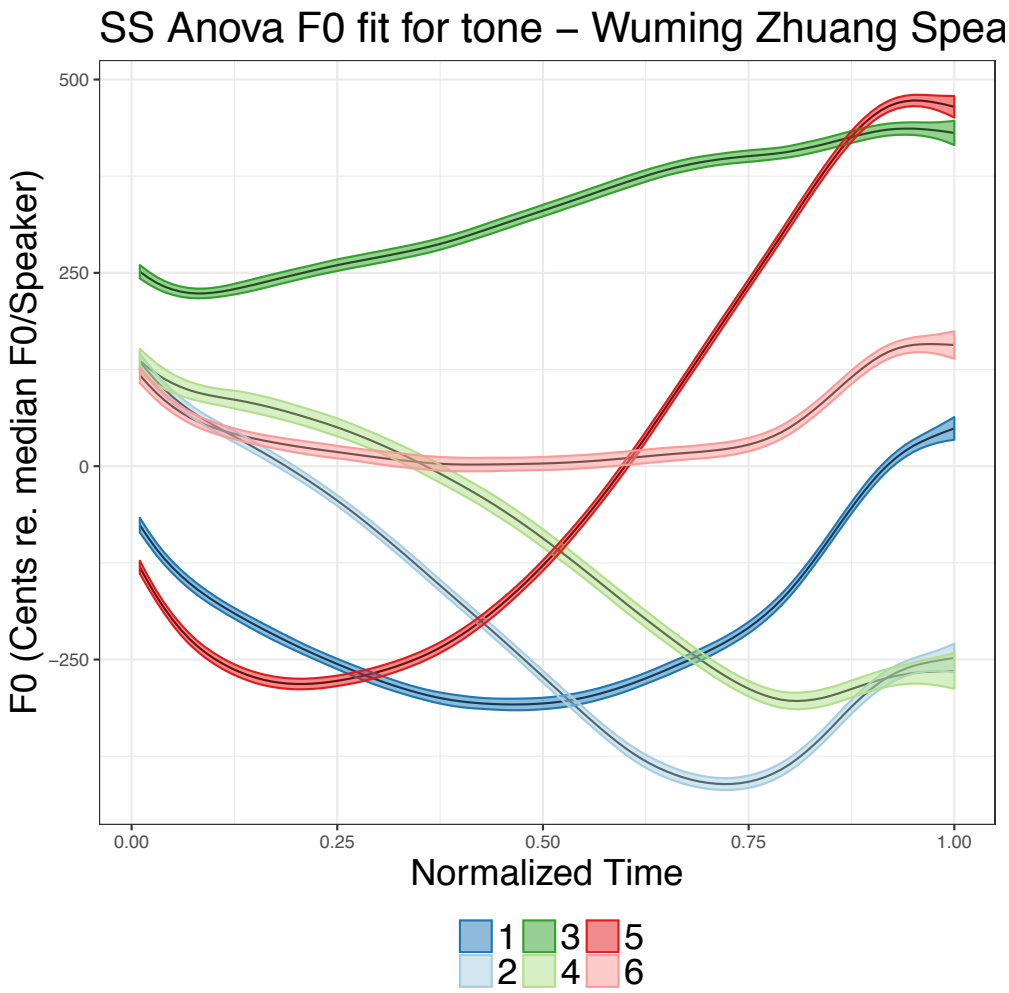
- In sum
  - 197 words with 5 repetitions
  - 985 tokens in total (for each speaker)

Results – F0 (all speakers)

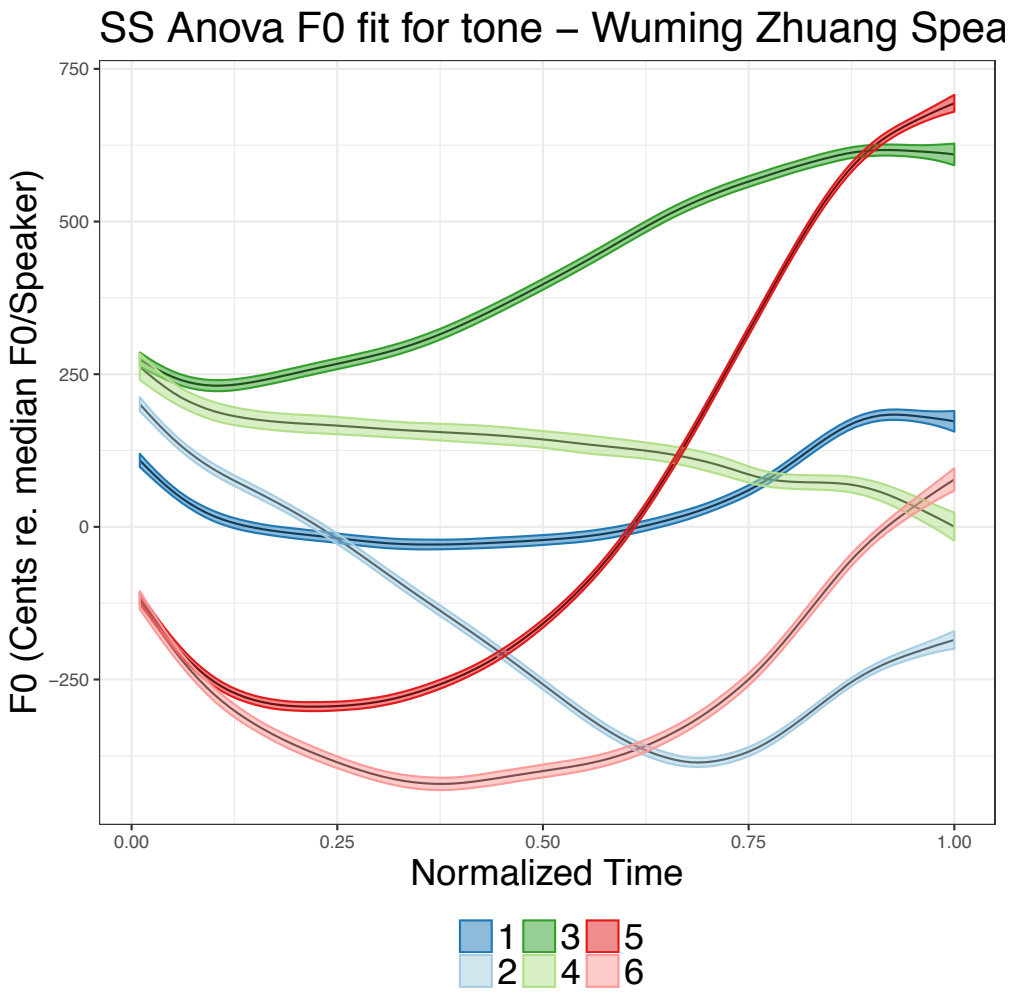


Wei & Qin	1	3	5	7 short	7 long
Tone	24	55	35	55	35
Description	rising	level	rising	level	rising
Tone	2	4	6	8 short	8 long
Chao	31	42	33	33	33
Description	falling	falling	level	level	level
	unchecked syllables			*checked syllables*	

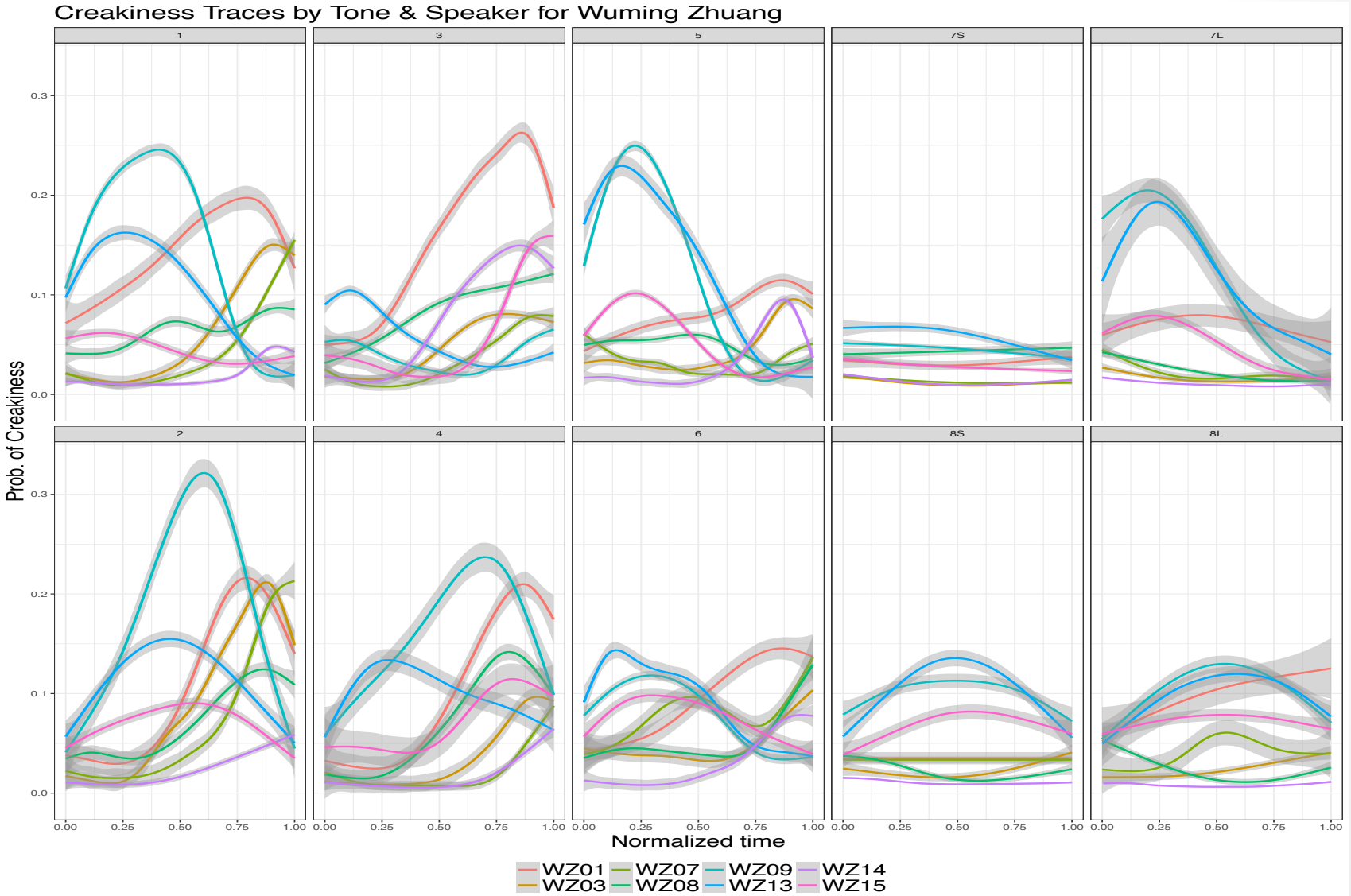
Results – F0 (T6 high group, pink)



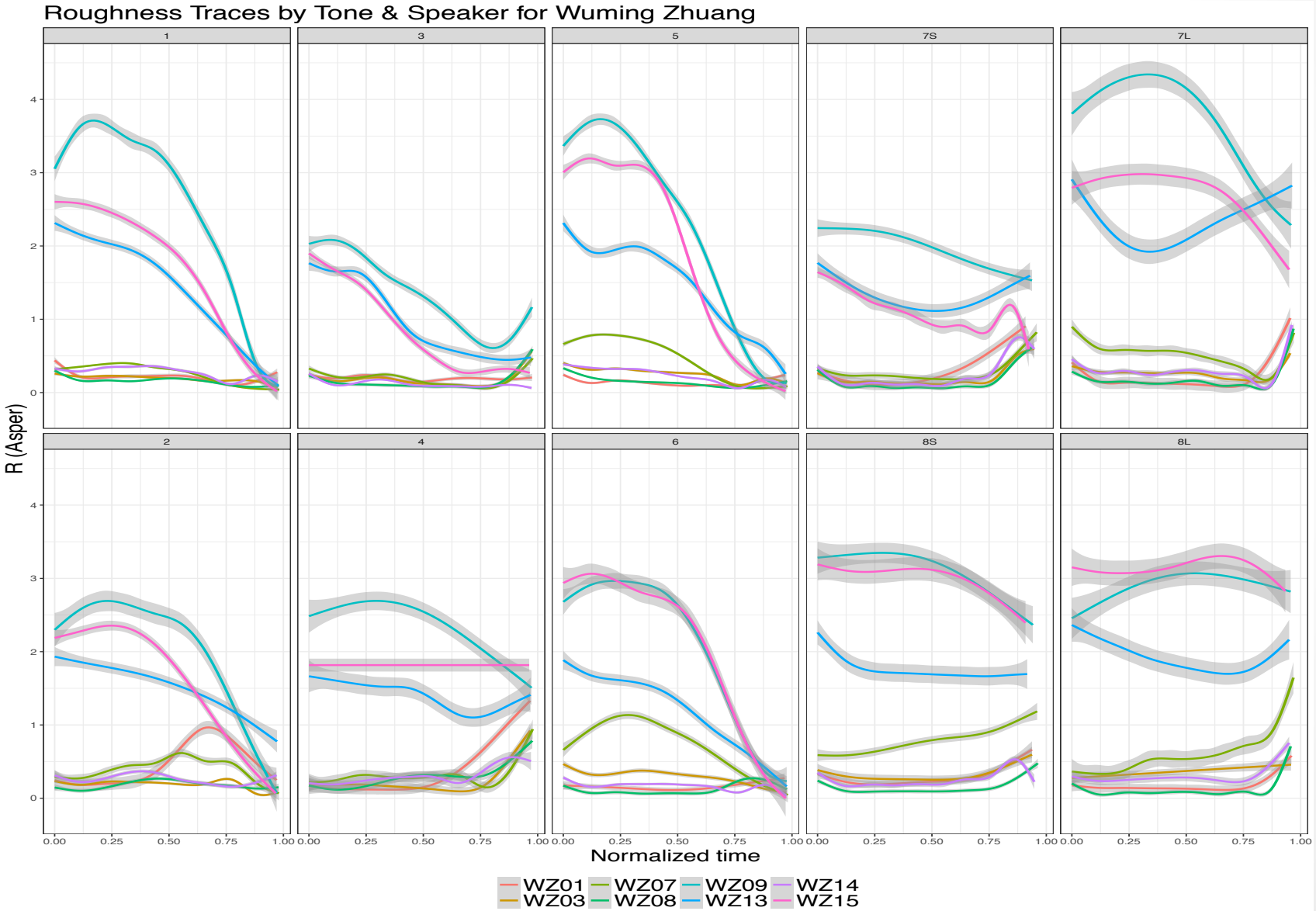
Results – F0 (T6 low group, pink)



Results – Covarep (creakiness)

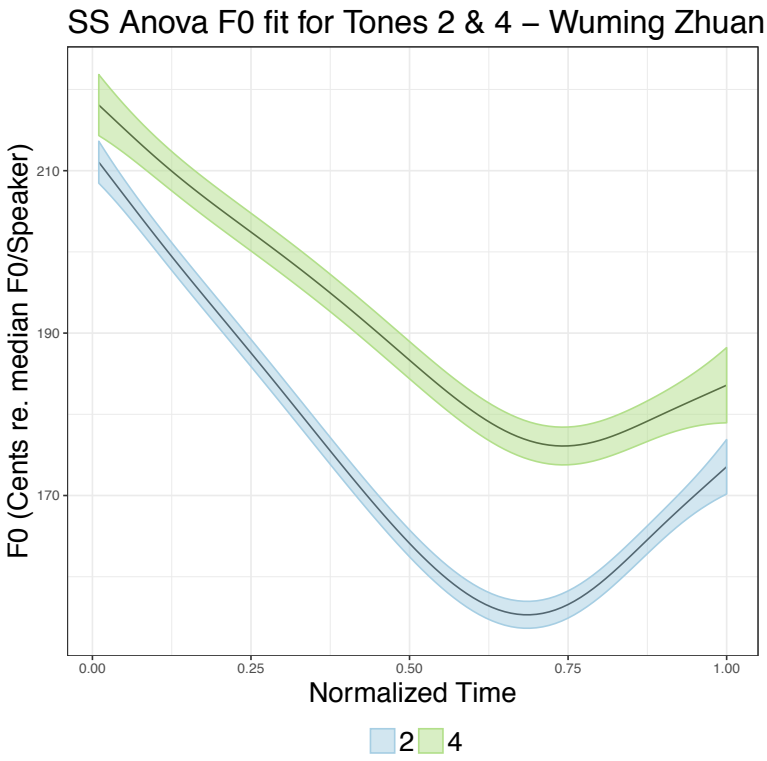


Results - Roughness

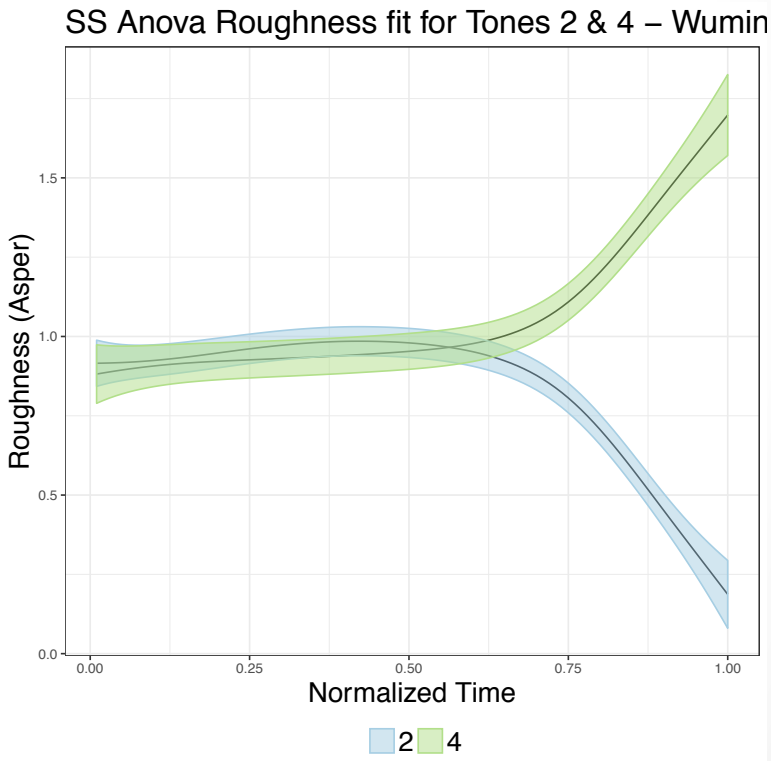


Tone 2 and tone 4 (all speakers)

F0



Roughness



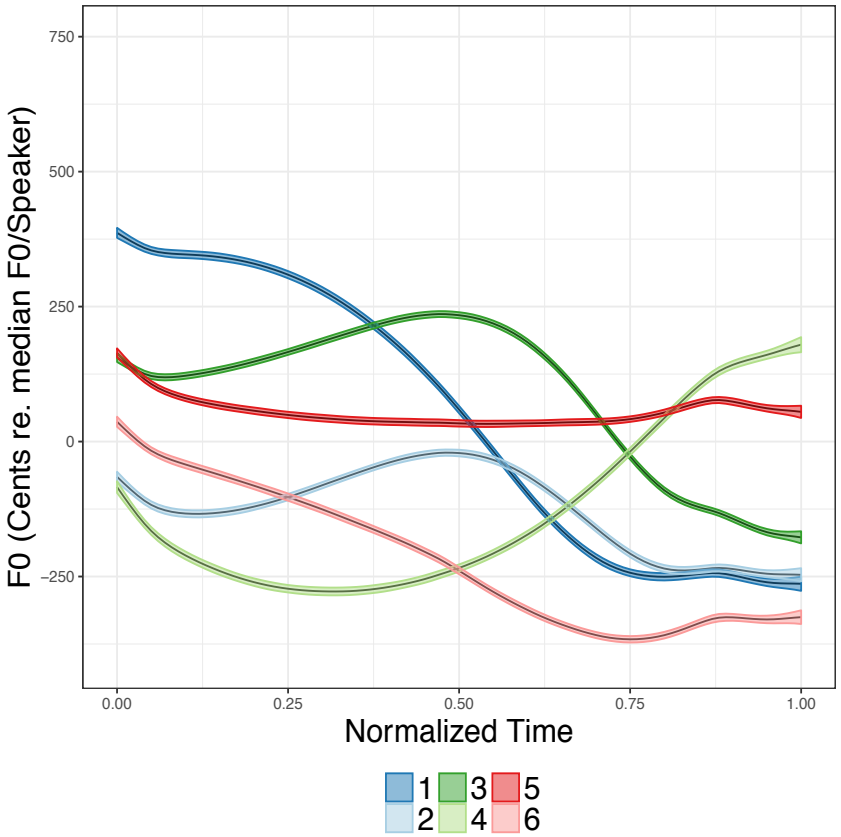
Summary – Wuming Zhuang

- Acoustic methods are used to further understand the interplay between phonation and F0, two known phonetic properties that are related to tone.
  - Creakiness detection algorithm
  - Roughness

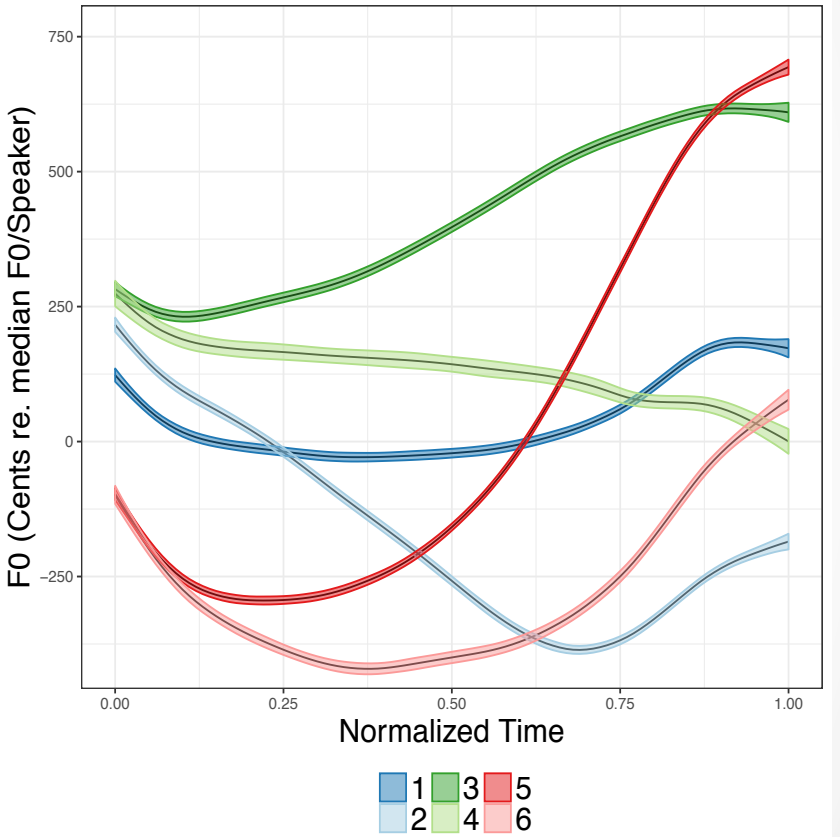
- In Wuming Zhuang, an instrumental investigation of F0 profiles shows that the speakers conform with earlier descriptions in Wei & Qin (1980)
  - There is some evidence that tone 2 and 4 in Wuming Zhuang may be distinguished by phonation, measured by roughness but not by Covarep.

Results – F0 – SS ANOVA

Du'an Zhuang

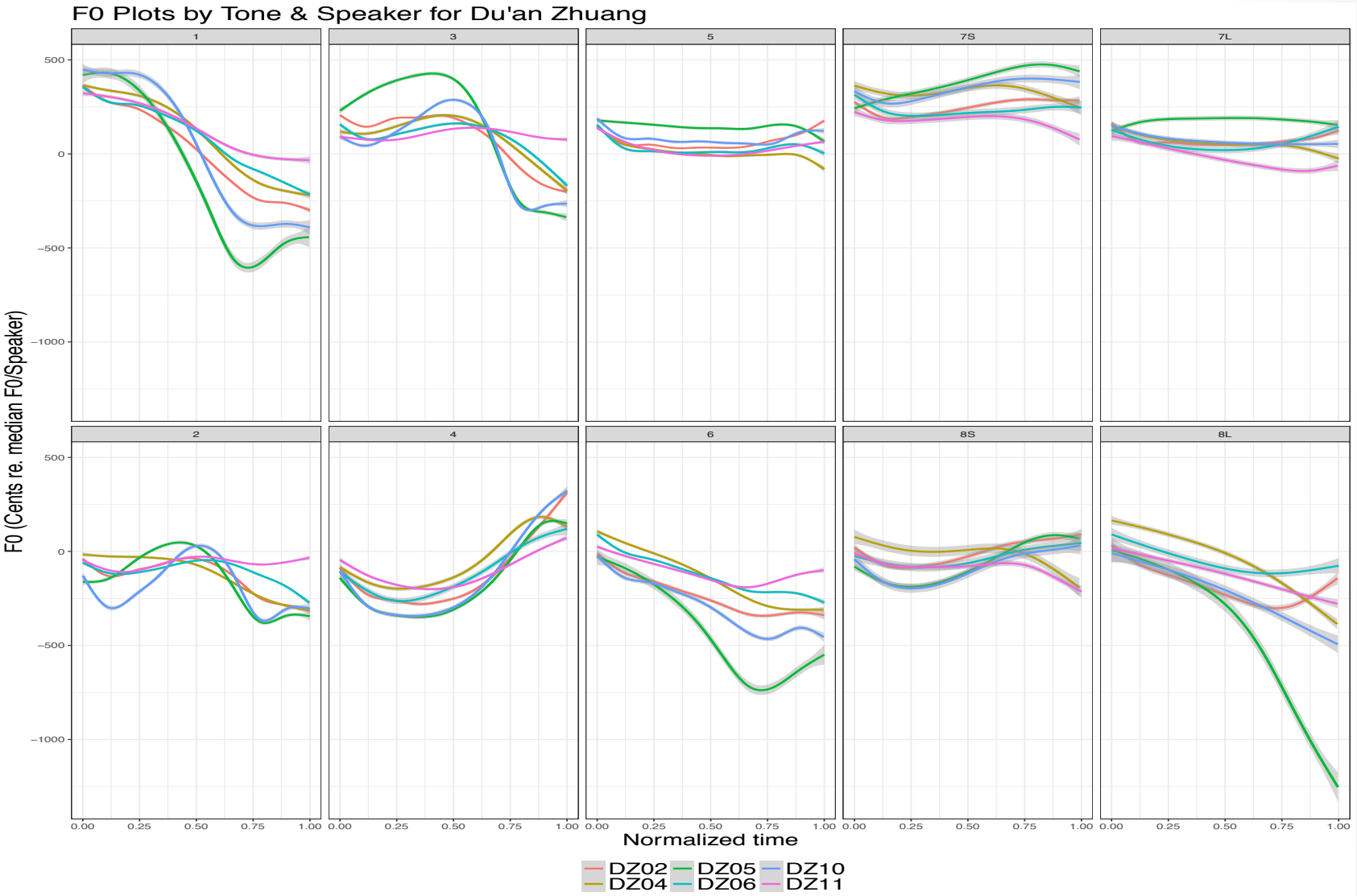


Wuming Zhuang



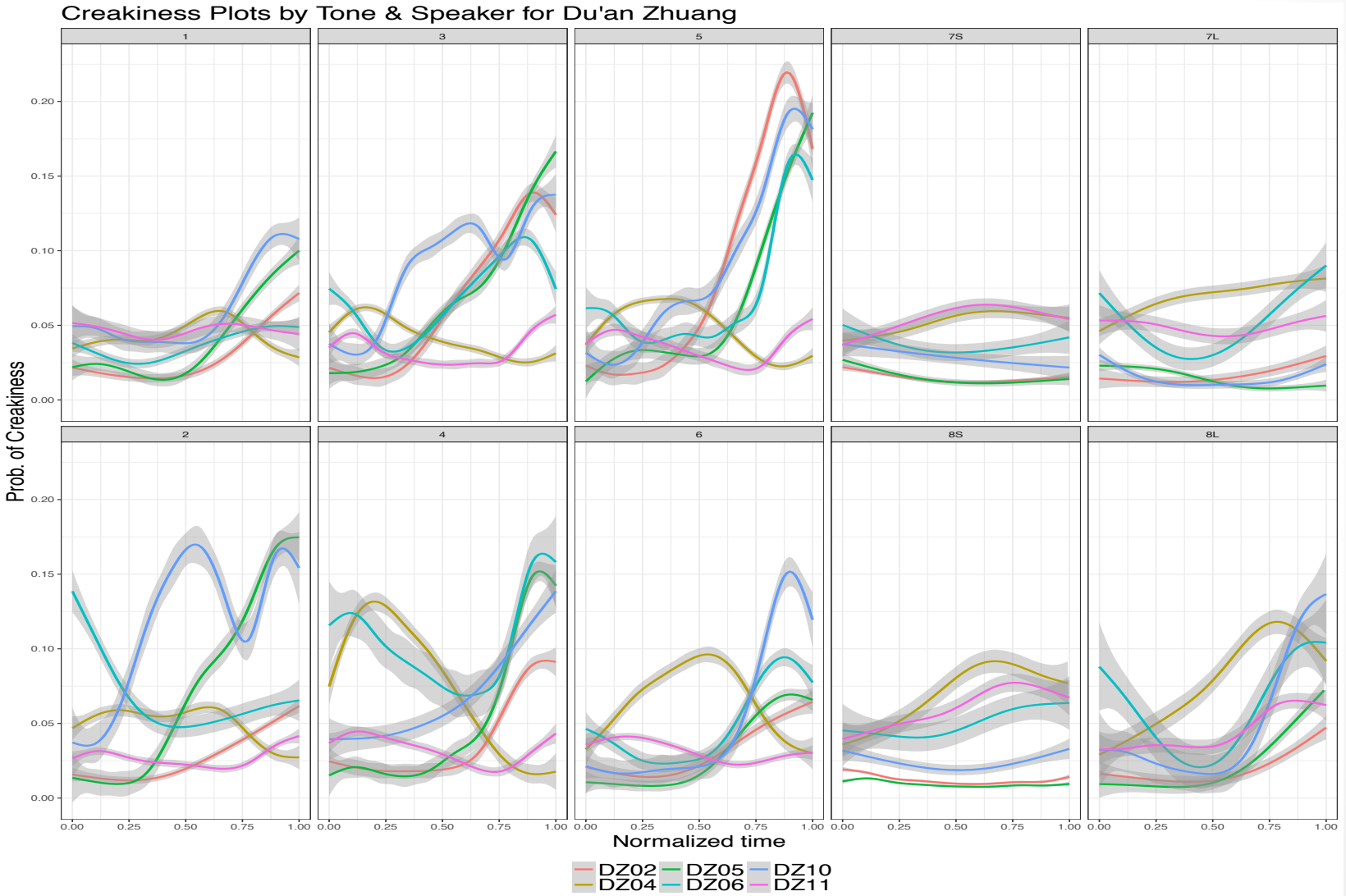
F0 Results Du'an Zhuang

(r^2 = 0.594 unchecked)



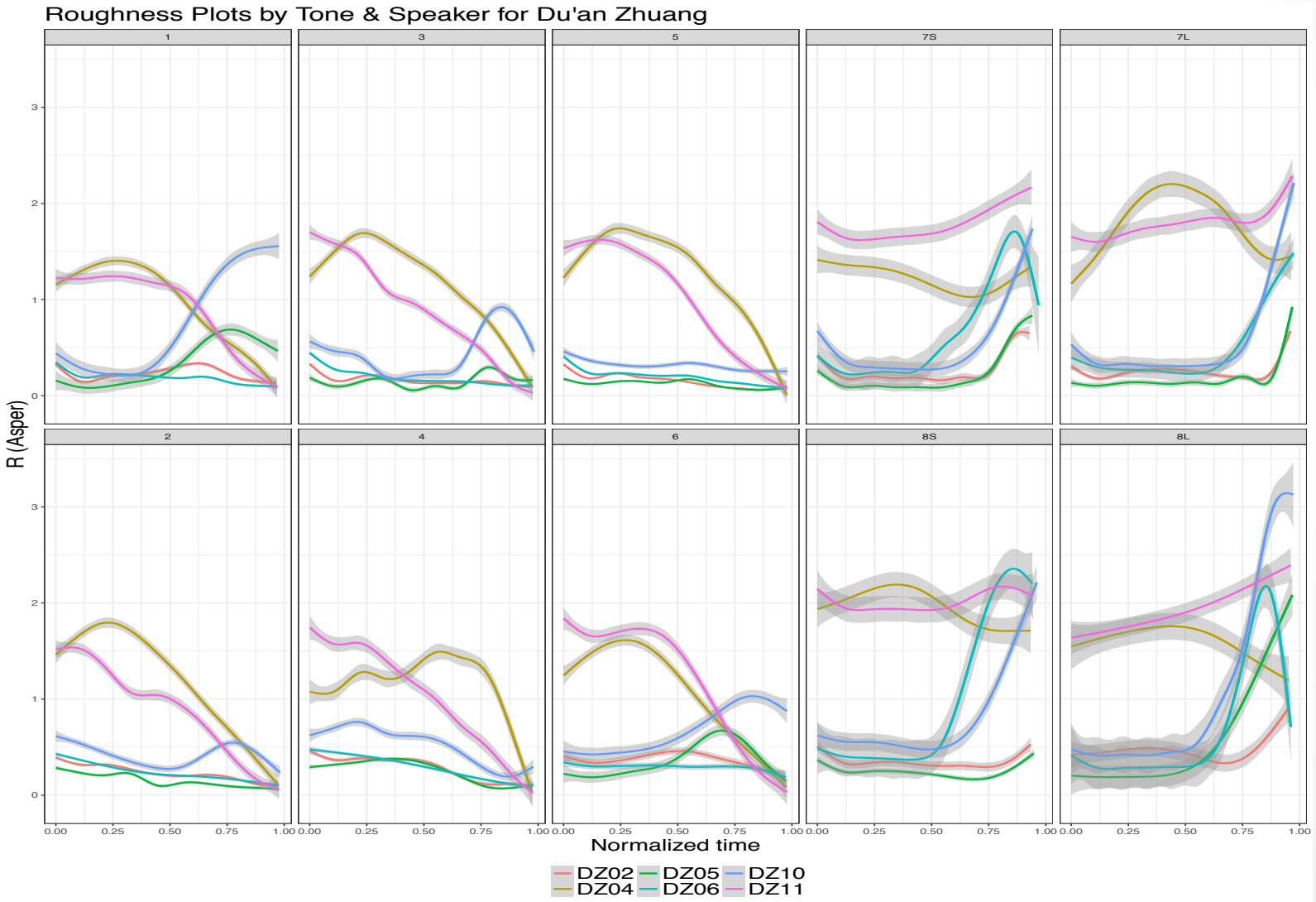
Covarep Du'an Zhuang

(unchecked w/o nasal codas $r^2 = 0.046$, checked $r^2=0.040$)



Roughness in Du'an Zhuang

(unchecked w/o nasal codas $r^2 = 0.040$, checked $r^2 = 0.091$)



Summary – Du'an Zhuang

- The creakiness detection algorithm and roughness, coupled with the F0 contour, shows that F0 and not creakiness is consistent within tones across speakers.
  - This implies that F0 and not creakiness is contrastive in the tonal system of Du'an Zhuang.

Conclusion

- Acoustic methods (such as creakiness detection algorithm and roughness) are used to further understand the interplay between phonation and F0

- A statistical method (SSANOVA) visualizes curves (F0 and creakiness detection) and facilitates comparisons between multiple occurrences of these curves.

- In Zhuang dialects,
  - Roughness suggests that creakiness may play a role in the Wuming Zhuang tone system.
  - F0, and not creakiness is contrastive in the Du'an Zhuang tone system (except for one speaker in our pilot study)

References

- Andersen, Torben (1993) Vowel quality alternation in Dinka verb inflection. Phonology 10, 1–42.
- Chao, Yuen-Ren (1930) A system of tone letters. *Le Maitre Phonetique* 45, 24-27.
- Degottex, Gilles, Kane, John, Drugman, Thomas, Raitio, Tuomo and Scherer, Stefan (2014) COVAREP – A collaborative voice analysis repository for speech technologies. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, 960-964.
- DiCanio, Christian T. (2008) *The Phonetics and Phonology of San Martín Itunyoso Trique*. PhD dissertation, UC Berkeley.
- DiCanio, Christian T. (2009) The phonetics of register in Takhian Thong Chong, Journal of the International Phonetic Association, 39: 162-188.
- Garellek, Mark & Patricia Keating (2011) The acoustic consequences of phonation and tone interactions in Jalapa Mazatec, Journal of the International Phonetic Association 41(2): 185-205.
- Gruber, James (2011) An articulatory, acoustic, and auditory study of Burmese tone. PhD Dissertation, Georgetown University.
- Gu, Chong (2014) Smoothing Spline Anova Models: R Package gss. *J. of Statistical Software* 58(5):1-25.

References

- Ishi, C. T., Sakakibara, K.-I., Ishiguro, H., and Hagita, N. (2008) A method for automatic detection of vocal fry. IEEE Trans. on Audio, Speech and Language Processing 16(1): 47-56.
- Kane, John, Drugman, Thomas, and Gobl, Christer (2013) Improved automatic detection of creak. *Computer Speech & Language*, 27(4):1028-1047.
- Kawahara, Hideki, Masanori Morise, Toru Takahashi, Ryuichi Nisimura, Toshio Irino, Hideki Banno, Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation, Proc. ICASSP 2008, Las Vegas,pp.3933-3936(2008) DOI: 10.1109/ICASSP.2008.4518514
- Keating, Patricia and Esposito, Christina (2007*) Linguistic Voice Quality*. UCLA Working Papers in Phonetics #105, 85-91
- Kim, S.-J. (1998) Positional effect on tonal alternations in Chichewa: phonological rules vs. phonetic timing. Proceedings of the Chicago Linguistics Society 34.
- Lee-Kim, S.-I., Kawahara, S., and Lee, S. J. (2014) The 'whistled fricative in Xitsonga: its articulation and acoustics. Phonetica 71(1) 50-81.
- Moisik, S. Lin, H., and Esling, J. (2013) Larynx Height and Constriction in Mandarin Tones, volume Eastward Flows the great river. City University of HK Press, 187-205.
- Myers, S. (1999) Tone association and F0 timing in Chichewa. Studies in African Linguistics 28: 215–239.
- Rose, Phil (1980) Thai-Phake Tones: Acoustic, Aerodynamic and Perceptual Data on a Tai Dialect wikth Contrastive Creak. Unpublished Manuscript.
- Silverman, Daniel (1997) Phasing and recoverability (Outstanding Dissertations in Linguistics). London: Routledge.
- Stumpf, C. (1898) Konsonanz und Dissonanz, Vol. 1 (J.A. Barth, Leipzig) in German.

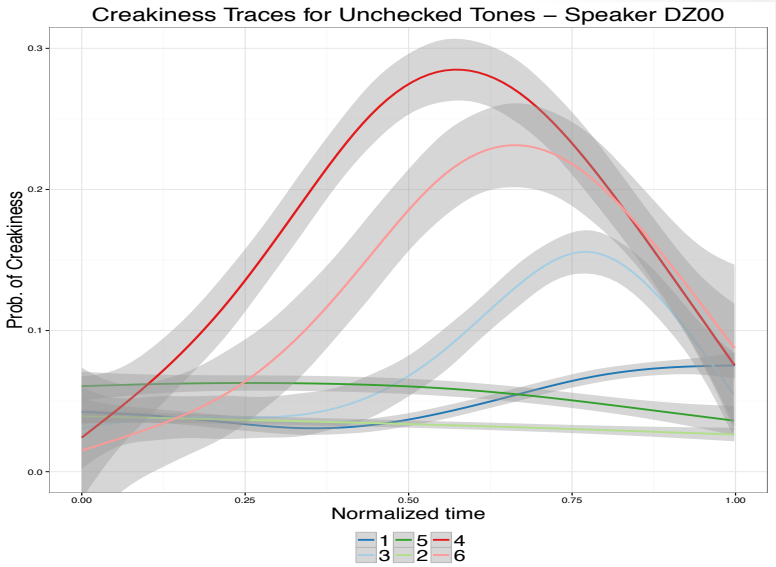
References

- Villegas, J. and M. Cohen (2010) Roughness Minimization Through Automatic Intonation Adjustments, *J. of New Music Research* 39, 75–92.
- Villegas, J. and M. Cooke (2012) Maximising objective speech intelligibility by local f0 modulation, in Proc. Interspeech.
- Villegas, J., J. Perkins, and S. J. Lee (submitted 2017) Automatic prediction of creaky voice with psychoacoustic roughness, J. Acoust. Soc. Am.
- Wang, Y., G. Shen, H. Guo, X. Tang, and T. Hamade (2013) Roughness modelling based on human auditory perception for sound quality evaluation of vehicle interior noise, *J. of Sound and Vibration* 332, 3893–3904.
- Watkins, Justin (2000) Notes on creaky and killed tone in Burmese. SOAS Working Papers in Linguistics 10: 139-149.
- Wei, Qingwen, and Qin, Guosheng (1980) *Zhuang yu jian zhi*. Beijing: Min zu chu ban she : Xin hua shu dian fa xing.
- Xu, Y. (1998) Consistency of tone-syllable alignment across different syllable structures and speaking rates. Phonetica 55: 179–203.
- Xu, Y. (1999) F0 peak delay: when where and why it occurs. In J. Ohala (ed.), International Congress of Phonetic Sciences 1999. San Francisco: n.p., pp. 1881–1884.
- Yip, Moira (2002) *Tone*. Cambridge University Press.
- Zhang et al. (1999) Studies on Zhuang Dialects. Sichuan Minzu Chubanshe. In Chinese.
- Zsiga, E. (2013) The Sounds of Language: An Introduction to Phonetics and Phonology. Wiley.

Acknowledgements

- We would like to thank our Zhuang speakers. We also thank Prof. Wentao Gu for introducing us with the members of the linguistics community at Guangxi University: Prof. Tang, Heng Lin and Lina Jo.
- Earlier parts of this study have also benefitted from comments from the audience of CUNY conference on endangered languages, ICPP 2013, ICPP 2015, TAL 6, HLS 22 and Berner Zirkel.
- Earlier part of this research was partially funded by CCSU research reassigned time and a CSU-AAUP university research grant awarded to Seunghun J. Lee.
- This research was supported by JSPS Kakenhi grant 15K16745 and 16K02641, and by AKS grant AKS-2016-LAB-2250004.

Creakiness Detection Algorithm (Covarep)



- a composite method used to measure creakiness
  - creakiness is estimated every 10 ms using the method in Degottex et al. (2014).
  - the algorithm effectively determines the odds of a frame being creaky based on a combination of acoustic features
    - the difference between the first two harmonics (H2-H1)
    - F0 contour
    - residual peak prominence
    - a group of features used by Ishi et al (2008)
      - including inter-pulse similarity, intra-frame periodicity
    - To minimize false positives, three measures are included
      - normalized signal energy, number of zero-crossings, variance in the very short-term power contour
  - The results are given by an Artificial Neural Network (ANN) trained with creakiness manually labeled.

Roughness

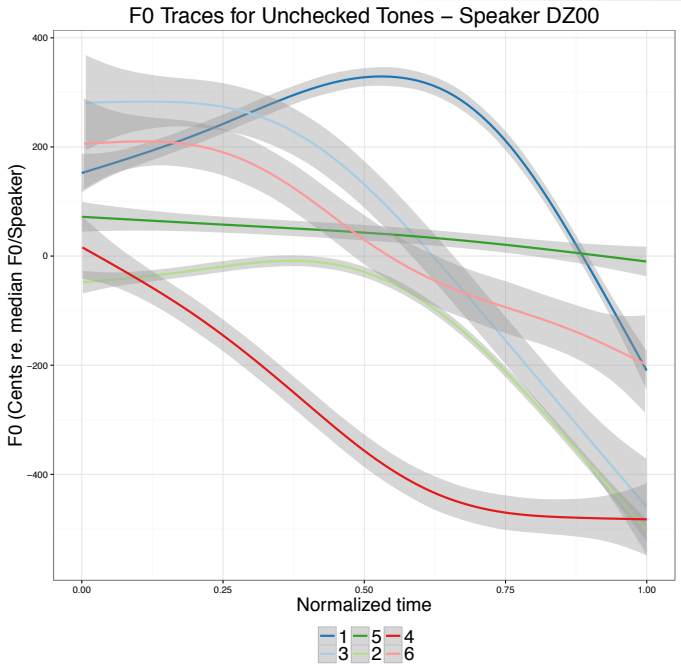
- A psychoacoustic attribute of a sound
  - elicited by rapid changes in the temporal envelop of a sound (between 15–300Hz, approximately)
  - has its maximum at about 70Hz.

- When two frequency components are sufficiently close, their interaction is perceived as rougher than otherwise separated frequency components.

- The roughness unit may shed light on what listeners are hearing when it comes to phonation types such as creakiness.

SSANOVA

- Fit smoothing spline ANOVA models in Gaussian regression (Gu 2014)
  - **F0, roughness,** and **creakiness** are explained by the factors *Tone* (6 unchecked Zhuang tones) and *normT* (the normalized duration of each token), and their interaction
  - We used a Restricted Maximum Likelihood (REML) method with the default smoothing parameters given by the R library.

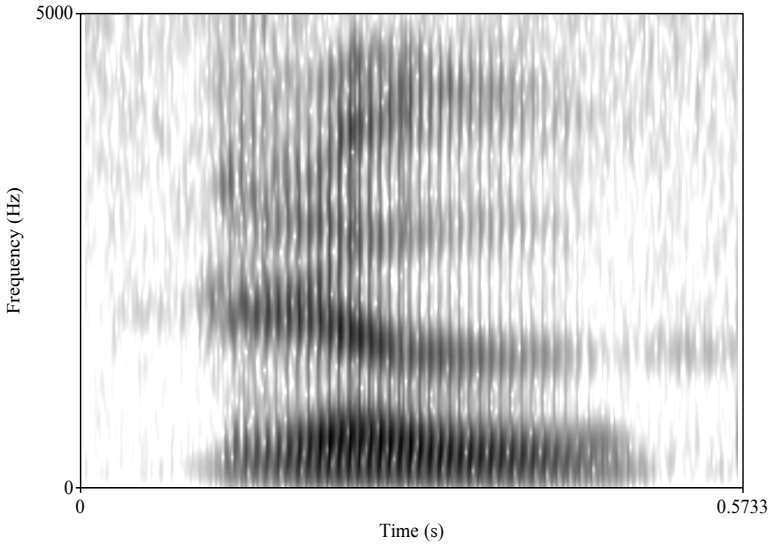


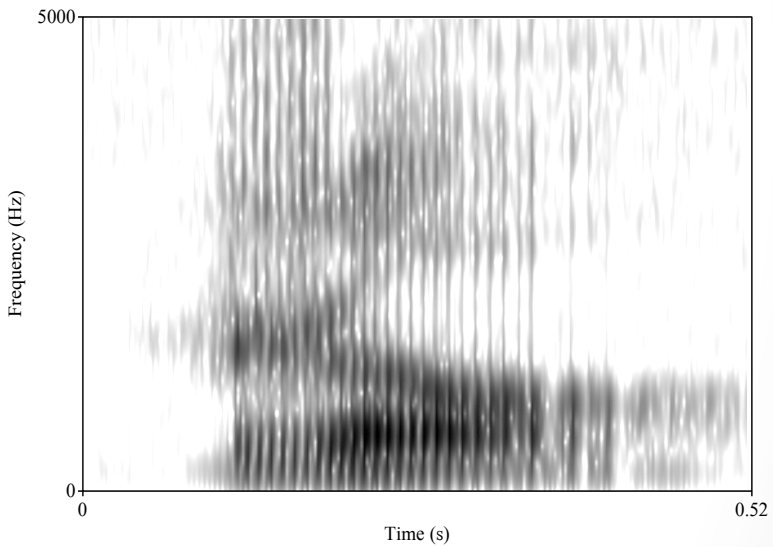
- A similar methodology was used in analyses of the lingual and labial articulation of whistled fricatives (Lee-Kim et al. 2014)
- F0 contours and larynx height for Mandarin tones (Moisik et al. 2013)

Examples – Tone 2 & Tone 4

- Creakiness can be observed as irregularly spaced glottal pulses

Modal tone 2 “ears” vs. Creaky tone 4 “meeting”





Part II – Data Collection

- Consultants
  - Two male speakers and four female native speaker of Du'an Zhuang in their early 20's were recorded in a sound-attenuated booth at Guangxi University, Nanning, China in Dec. 2015.
  - The speakers were bilingual Du'an Zhuang/Mandarin speakers and spoke regularly with their family in Du'an Zhuang. They grew up in Du'an.