WWW.OEAW.AC.AT

EXPLORATION AND CLASSIFICATION OF VOCAL FRY, PERIOD DOUBLING, AND MODAL VOICE USING ACOUSTIC AND EGG MEASURES

Yaqian Huang | Acoustics Research Institute, ÖAW

yaqian.huang@oeaw.ac.at | yaqianhuang.netlify.app

I. OVERVIEW

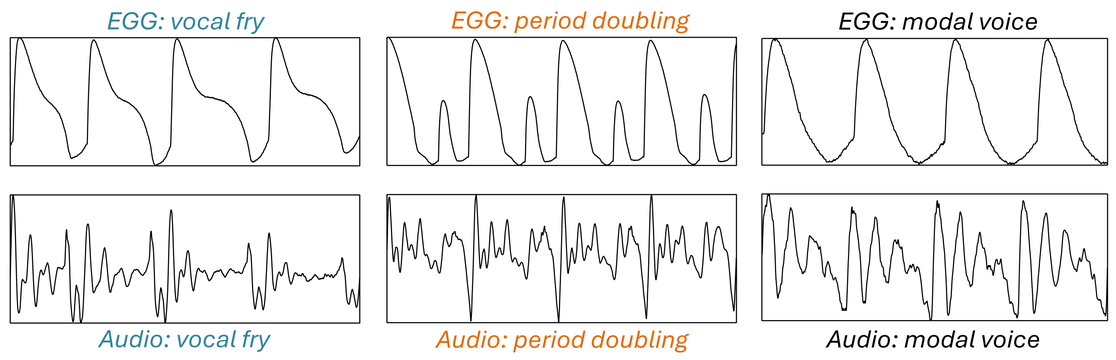
Creaky voice differs from modal voice in acoustic & phonatory properties, but how subtypes of creaky voice differ is less clear.

Existing classification of creak subtypes are mainly based on acoustic waveforms (Hedelin & Huber 1990, Redi & Shattuck-Hufnagel 2001)

***Q:* How do subtypes of creaky voice differ in their acoustic and phonatory properties?**

Common acoustic attributes of creaky voice	Low f0	Low spectral tilt	Irregular f0	Others
prototypical	x	x	x	
vocal fry	x	x		damped pulses
period doubling	alternation	x		subharmonics
aperiodic		x	x	

After Keating et al. 2015, Huang 2022



***Goal*: evaluate the importance of source and filter characteristics in distinguishing between vocal fry, period doubling (PD), and modal voice**

- Most important **acoustic** measures: **f0, H1-H2, H1, SoE, H2, H4, and HNR (0-500Hz)**
- Most important **phonatory** measures: **duration of the glottal opening phase** and **contact quotient** of the glottal pulse

II. METHODS

- Simultaneous recordings of audio and electroglottography (EGG) of continuous read speech in Mandarin (Huang 2024)
- 20 university students (10F); 480 sentences/recording
- Fixed carrier sentence with varying trisyllabic words + picture fillers
- wo3 tɕau1 nji3 STIMULUS tsən3 mɤ5 ʂwo1
- "I teach you how to say the STIMULUS."
- Tokens of vocal fry, PD, and modal voice were located using EGG

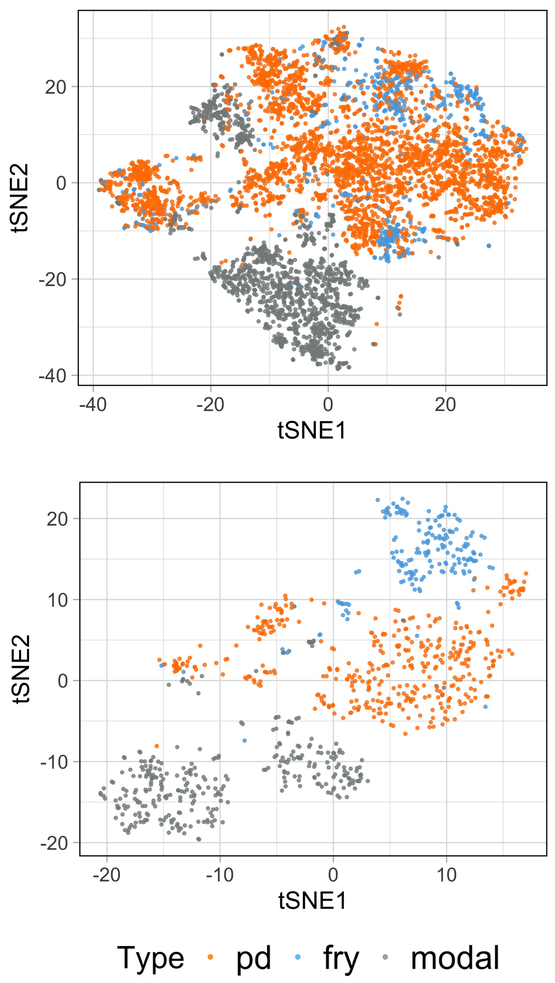
32 acoustic & 11 EGG normalized measures (VoiceSauce, EGGWorks):

- Harmonics & spectral tilts: e.g., H1*, H2*, H4*, H2K*, A1*, A2*, A3*, H1*–H2*, H2*–H4*, H1*–A1*, H1*–A2*, H1*–A3*, H4*–H2K*
- HNRs, CPP, SHR (subharmonic-to-harmonic ratio)
- Formants & bandwidths; Energy measures
- Contact quotient (CQ), Cycle peak/minimum velocity/time, Speed Quotient (SQ); Contacting and opening duration

t-SNE clustering & ML models:

- Multinomial logistic regression with *l1* regularization (Lasso)
- Random forest model
- Cross-validation: 2/3 training & 1/3 test sets

III. T-SNE CLUSTERING



Acoustics dataset

	Vocal fry	Period doubling	Modal
Women	482	2354	1175
Men	156	943	428
Total	638	3297	1603

5538 rows x 33 cols (gender coded as 0/1)

Acoustics + EGG dataset

	Vocal fry	Period doubling	Modal
Women	154	324	187
Men	25	91	137
Total	179	415	324

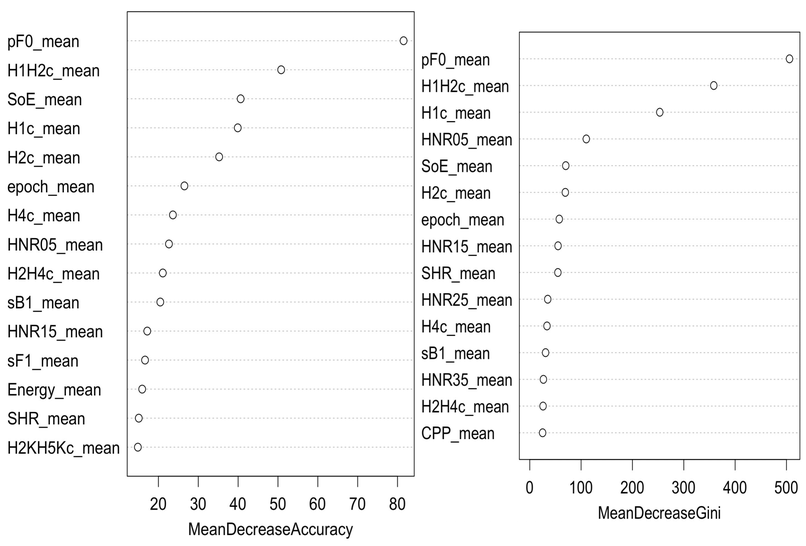
918 rows x 44 cols (gender coded as 0/1)

IV. CLASSIFICATION RESULTS

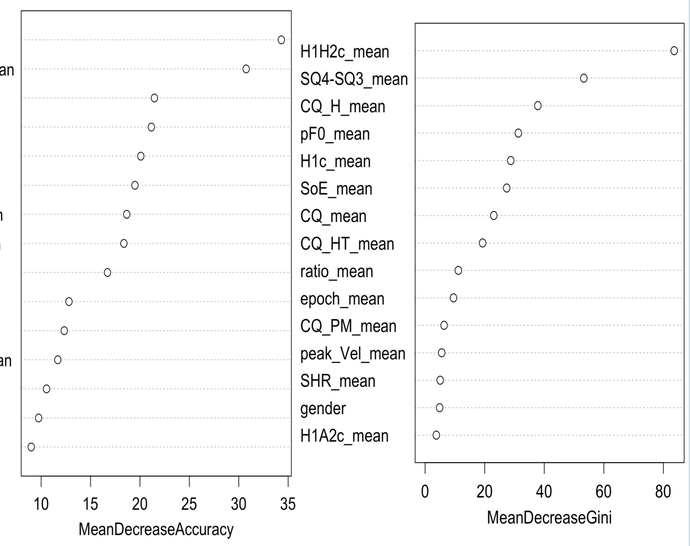
Accuracy, precision, recall scores in two datasets

Acoustic	Lasso regression	Random forest
Accuracy	0.9112	0.9312
Macro avg. precision	0.8749	0.9098
Macro avg. recall	0.8137	0.8529
Acoustic + EGG	**Lasso regression**	**Random forest**
Accuracy	0.9837	0.9967
Macro avg. precision	0.9840	0.9975
Macro avg. recall	0.9784	0.9970

Top 15 important acoustic measures



Top 15 important acoustic + EGG measures



V. DISCUSSION & CONCLUSIONS

- Phonatory measures such as EGG contribute to the separation of subtypes of creaky voice
- Voicing types have stronger ties to the source dynamics associated with our vocal folds than the filter
- Mapping between perception and acoustics is direct, but what about phonatory characteristics? How are they accessible to listeners?

REFERENCES

Hedelin, P. and Huber, D. (1990). Pitch period determination of aperiodic speech signals. International Conference on Acoustics, Speech, and Signal Processing, 361–364.

Huang, Y. (2022). Articulatory properties of period-doubled voice in Mandarin. Proc. Speech Prosody 2022, 545-549.

Huang, Y. (2024, April 29). F0 and voice quality of coarticulated Mandarin tones. https://doi.org/10.31219/osf.io/nyva6

Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. Proceedings of the 18th ICPhS, 2–7.

Redi, L. and Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. Journal of Phonetics, 29(4), 407–429.

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. Proceedings of the ICPhS XVII, 1846-1849.

Tehrani, H. (2009). EGGWorks: a program for automated analysis of EGG signals. Computer program.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11), 2579-2605.

ARI – ACOUSTICS RESEARCH INSTITUTE