Source and filter contributions to voice quality differences

Donna Erickson[1], Sayoko Takano[2], Yongwei Li[3], Jiayin Gao[4], Shigeto Kawahara[5], Kerrie Obert[6], Kyoko Takahashi[7], Masato Akagi[7]

1 Haskins Laboratories, U.S.A., 2 Kanazawa Institute of Technology, Japan, 3 Institute of Automation, Chinese Academy of Sciences, 4 Sophia University (JSPS), Japan, 5 Keio University, Japan, 6 The Ohio State University, USA, 7 Japan Advanced Institute of Science and Technology

[1]ericksondonna2000@gmail.com, [2]tsayoko@neptune.kanazawa-it.ac.jp, [3]yongwei.li@nlpr.ia.ac.cn, [4]jiayin.gao@gmail.com, [5]kawahara@icl.keio.ac.jp, [6]kerriebobert@gmail.com, [7]kyoko.takahashi@jaist.ac.jp, [7]akagi@jaist.ac.jp

Our study

- We use MRI imaging and electrography (EGG) to investigate source and filter contributions to voice quality differences.
- We also use the ARX-LF model [1, 2] as a way to estimate source and filter contributions to the acoustic signal and for comparison with the MRI and EGG results.
- Two sustained /i/ vowels were produced using different phonation modes, a la cover-body theory of phonation [3];
  - 1. "thin" folds (cover, upper edge of the folds only)

"Thin" folds may be similar to "Mechanism 2"

  - 2. "thick" folds (cover & body)
  - "Thick" folds may be similar to "Mechanism 1"

Questions

- 1. How does "thickness" of vocal folds affect Open Quotient (OQ)?
- 2. Does "thickness" of vocal folds affect supralaryngeal settings?
- 3. What are some acoustic changes associated with thick fold vs thin fold mode of phonation?

Data

- **MRI recordings:** MRI & simultaneous acoustic recordings were made at ATR, Inc. Kyoto, Japan, using the BAIC MRI recording equipment.
- **EGG recordings:** Vowels recorded separately in a soundproof room at Arai Lab at Sophia University with electroglottograph (Glottal Enterprises EG2-PCX2) and an electret condenser microphone (Sony ECM-MS957) connected to a laptop computer via an audio interface (Edirol UA-25EX). Acoustic and EGG signals were recorded simultaneously using Audacity at a 44.1 kHz sampling rate,
- **Speaker:** Phonetician trained in Estill Voice Production Method [4]
- **Sustained vowels: /i/**, produced with 2 modes of phonation: (1) thin & (2) thick, keeping F0 approximately the same.

Methodology

- **EGG Analysis**: Praatdet (Kirby 2017, [5]). OQ estimated by detection of closing and opening peaks, using vowel sounds recorded at Sophia.
- **Acoustic Analysis**: ARX-LF Model of sounds recorded with MRI

Results

MRI images of 2 /i/ vowels

➢Figure 1. MRI images. From left to right THIN vocal folds, THICK vocal folds, overlay of THIN (pink) and THICK (green). For the overlay, pink (THIN folds) were moved upward and forward, and green (THICK folds) were moved backward and downward.

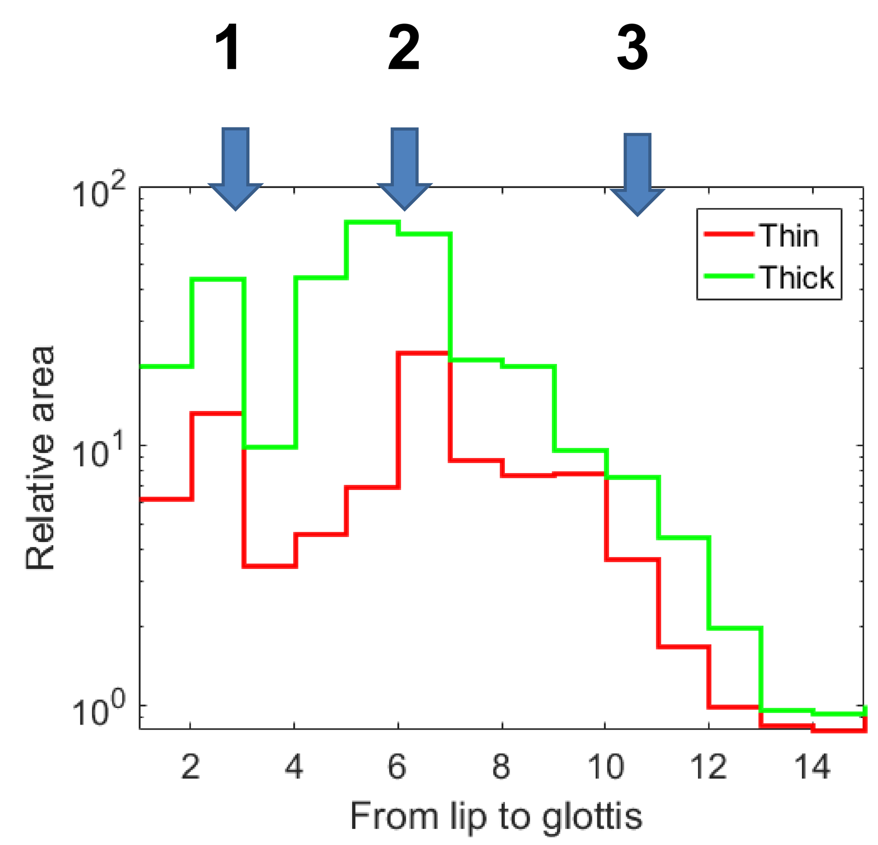
➢Comments

➢Speaker kept larynx position the same for the two modes of phonation,

➢But the vocal tract midsagittal contours are different for THICK vs. THIN fold phonation.

- ➢For THICK folds, the **velum is more raised** and **the tongue is more bunched**, suggesting that more articulatory adjustments were required for the THICK fold phonation.
- ➢Also, for THICK folds, the posterior oral cavity and oropharynx **appears to be bigger** than that of the THIN folds voice. Also, the front cavity is larger for the THICK than for the THIN.

➢Estimated area functions generated by the ARX-LF model [6]



➢Comments

➢Notice the estimated area for the THICK folds is much larger than for the THIN folds, as shown also in the MRI images,

➢The area for the THICK folds is greater for three different parts of the vocal tract, (1) from lips to 3, (2) from 4 to 7 and (3) from 7 to 14

➢ Estimated Formants and Spectra

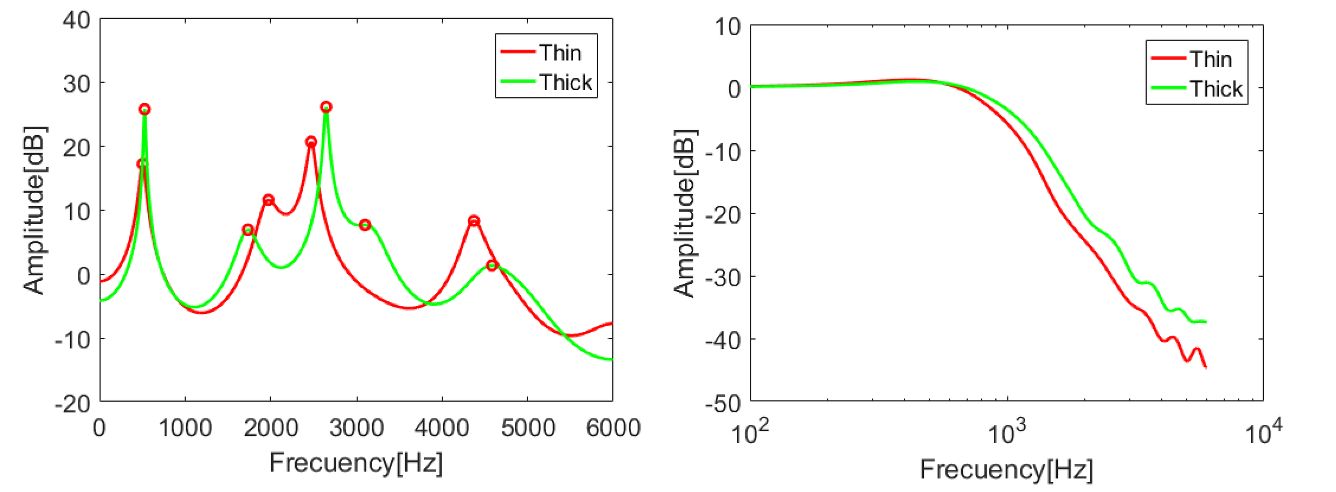


Table 1. ARX-LF model estimates of F0, formants, spectral tilt, and open quotients (OQ). Also, shown are OQ derived from egg from separate recordings.

ID	Phonation mode	F0	F1	F2	F3	F4	Tilt	OQ_{ARX-LF}	OQ_{egg}
5	THIN	500	510	1980	2479	4377	-15.6	0.47	0.78
6	THICK	520	533	1740	2654	3105	-12.7	0.4	0.55

➢ Comments

1. For THICK voice, F1 is higher and F2 is lower than for THIN, bringing F1 and F2 closer together.
1. This result is consistent with the report that a wide oral cavity & increased pharyngeal area => F1 & F2 to be closer together [7].
2. The OQ estimates with the ARX-LF model and those from EGG show that OQ is greater for THICK folds than for THIN folds.
3. Also, the spectral tilt is steeper for THIN folds than THICK ones

➢ Summary

- 1. THIN folds have larger OQ than THICK folds (also reported by [8]).
- 2. When a speaker changes phonation modes (e.g. THIN vs THICK), supralaryngeal articulation also changes, & formant frequencies & spectral tilt change accordingly.
- 3. For THICK fold phonation, the oral cavity increases consistently throughout the vocal tract (front, mid and back) (adding to findings reported in [9]).

References

[1] .Li, Y., Li, J., M. Akagi, M. (2018) Contributions of the glottal source and vocal tract cues to emotional vowel perception in the valence-arousal space", *J. Acoust. Soc. Am.*, **144**, doi: 10.1121/1.5051323.

[2] Li, Y., Sakakibara, K-I., Akagi, M. (2019) Simultaneous Estimation of Glottal Source Waveforms and Vocal Tract Shapes from Speech Signals Based on ARX-LF Model, J. Signal Processing Systems, 92, 831-838, 2020. 10.1007/s11265-019-01510-4

[3] Hirano, M. (1977) Structure and vibratory behavior of the vocal folds, *Dynamic Aspect of Speech Production*, University of Tokyo Press, Tokyo, Japan, pp. 13–27.

[4]Estill, J., Steinhauer, K., McDonald, M. (2017) *The Estill Voice Model: Theory and Translation*. Estill Voice International, LLC: Pittsburgh PA.

[5] Kirby, J. (2017) Praatdet: Praat-based tools for EGG analysis (v0.1.1). https://doi.org/10.5281/zenodo.1117189.

[6] Kawahara, H., Sakakibara, K.-I., Banno, H., Morise, M., Toda, T., Irino, T. (2015) Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation. *Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 520–529. IEEE Press.

[7] Story, B.H., Titze, I.R., Hoffman, E. A. (2001) The relationship of vocal tract shape to three voice qualities, *J. Acoust. Soc. Am*. **109**,1651-1667

[8] Erickson, D., Yun, J., Gao, J., and Obert, K. (2020). Interaction between phonation mode and pharyngeal narrowing: A pilot EGG study. ISSP 2020.

[9] Henrich-Bernadoni, N., Smith, J., Wolfe, J. (2014) "Vocal Tract resonances in singing: variation with laryngeal mechanism for male operatic singers in chest and falsetto registers" *J. Acoust. Soc. Am.* **135**, 491-501.

Acknowledgments. This study was partially supported by a Grant-in-Aid for JSPS Fellows 17F17006 to Takayuki Arai and Jiayin Gao, and by research money granted by the Keio Institute of Cultural and Linguistic Studies to Shigeto Kawahara.