Source and filter contributions to voice quality differences

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

[1]Haskins Laboratories, U.S.A., Kanazawa Medical University, Japan, [2]Kanazawa Institute of Technology, Japan, [3]Institute of Automation Chinese Academy of Sciences, China, [4]Japan Society for the Promotion of Science, [5]Keio University, Japan, [6]The Ohio State University, U.S.A,[7]Japan Advanced Institute of Science and Technology, Japan

Ericksondonna2000@gmail.com, tsayoko@neptune.kanazawa-it.ac.jp, liyongwei2000@126.com, jiayin.gao@ed.ac.uk, kawahara@icl.keio.ac.jp, kerriebobert@gmail.com, kyoko.takahashi@jaist.ac.jp, akagi@jaist.ac.jp

This talk reports an exploratory study which examined how the source and the filter contribute to voice quality differences. The speaker was a female phonetician (first author) trained in the Estill Voice Training Method® of singing [1], who produced nine sustained vowel sounds with different voice qualities by varying F0, mode of phonation, and articulatory (pharynx and tongue) positions. Acoustic and MRI recordings were made at ATR, Inc. Kyoto, Japan, using the BAIC MRI recording facilities for two /i/-vowels produced at around 500 Hz with two modes of vocal fold vibration: THIN and THICK. THIN vocal folds were produced using the upper edge and cover of the vocal folds, while THICK folds, with using greater mass and body of the folds and both upper and lower medial edge (see [2] for description of vocal fold composition). Roughly speaking, THIN folds correspond to falsetto-type phonation, and THICK folds, to modal-type (chest) phonation [3]. Acoustic analysis of the two /i/ vowels were done using the ARX-LF model [4, 5]; in addition, EGG (electroglottograph) recordings (Glottal Enterprises, EG2-PCX2) of THIN and THICK /i/vowels at around 500 Hz were made separately in the soundproof booth of the Arai-lab, Sophia University, Japan. Open quotients (OQ) were measured using the derivative of the EGG signals as in [3, 6]. How these different voice qualities impact perception have also been explored but is separately reported in [7].

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.

MRI images of the two modes of phonation (THIN and THICK folds) are shown in Figure 1. Visual inspections suggest that even though the speaker kept the larynx position the same for the two modes of phonation, the vocal tract area is different. 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, the posterior oral cavity and oropharynx of the THICK voice (the middle panel) appears to be bigger than that of the THIN voice (the left panel). Quantitative analyses of the vocal tract areas of the two sounds are on-going and will be reported and compared with the estimated acoustic outputs generated by the ARX-LF model [8].

In applying the ARX-LF model, we assume the glottis=1. The obtained formant values, spectral tilt and open quotient (OQ), are shown in Table 1. We see that F1 is higher and F2 is lower for the THICK voice than for the THIN one, thus bringing F1 and F2 closer together. This result is consistent with the findings from examination of the relation of vocal tract shape to three voice qualities, which reported that a wide oral cavity and increased tract length may cause F1 and F2 to be closer together [9].

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 (falsetto)	500	510	1980	2479	4377	-15.6	0.47	0.78
6	THICK (modal)	520	533	1740	2654	3105	-12.7	0.4	0.55

Table1 also shows the THIN voice has a sharper spectral tilt and a larger OQ than the THICK voice, indicating that the former produced a more breathy voice than the latter. Previous studies, e.g., [10], have similarly reported that falsetto voices have less glottal contact than chest voices, resulting in more breathy voice quality. Two OQ values are shown in the table, one estimated by the ARX-LF model, the other from EGG recordings made at a separate time. Both show the same tendency of the THIN voice having a higher OQ than the THICK one.

The results of this exploratory study suggest that when a speaker changes phonation modes (e.g. THIN vs THICK), supralaryngeal articulation is also changed. The study by [10] reported formant differences between falsetto and chest voices, but could only speculate about "putative articulatory changes associated with the change in laryngeal mechanism" (p. 500). Our study with MRI shows that supralaryngeal gestures do indeed change when phonation mode is changed, but it also indicates more specifically what these changes involve.

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.

References

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

[2] 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.

[3] Henrich, N., d'Alessandro, C., Doval, B. , Castellengo, M. (2004) On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation, *J. Acoust. Soc. Am.,* 1321-1332.

[4] 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.

[5] 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, 1-8. 10.1007/s11265-019-01510-4

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

[7] Erickson, D., Kawahara, S., Rilliard, A., Hayashi, R., Sadanobu, T., Li, Y., Daikuhara, H., de Moraes, J., Obert, K. submitted. Cross cultural differences in arousal and valence perceptions of voice quality. Submitted to *Speech Prosody 2020*.

[8] 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.

[9] 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

[10] 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