Spectral energy properties of non-modal phonations

Yuan Chai[1], Padmini Bhagavatula[1], Serene Wong[1], and Patricia Keating[2]
[1]University of Washington, [2]University of California, Los Angeles

Falsetto and breathy voices have distinct articulatory mechanisms and auditory percepts. Falsetto voices are produced with thinly stretched vocal folds (Titze, 2000; van den Berg, 1968). Breathy voices are produced with little vocal fold constriction (Kreiman et al., 2012). Despite the articulatory and perceptual differences, previous literature has described the spectral energy properties of breathy and falsetto voices to be similar to each other: they both have high energy in the fundamental frequency, and low energy in the high-frequency harmonics (Colton, 1972; Zhang, 2016). The current study aims to provide a quantitative comparison in harmonic energy among falsetto voice and breathy, modal, and creaky voices to examine whether and how falsetto voice is different from these phonations.

We used an existing corpus of the "Little Red Riding Hood" story narrated by Mandarin speakers from the "Production and Perception of Linguistic Voice Quality" project at UCLA. Speakers often chose to use different voice qualities for different characters in the story. For this study, each vowel was manually annotated with its perceived phonation type. The current analysis includes 988 vowel tokens (falsetto: 222; breathy: 119; creaky: 93; modal: 554) produced by ten female and nine male speakers; we continue to add results from additional speakers in the corpus. We measured H1, H2, H3, H4, H1K (the harmonic closest to 1000 Hz), H2K, H3K, H4K, and H5K using VoiceSauce (Shue et al. 2011). Note that some of these parameters (H3, H1K, H3K, H4K) were newly added to VoiceSauce by us to ensure comprehensive coverage of the spectral energy. In order to approximate the harmonic energy of the source of phonation, we performed several steps of outlier exclusion and formant correction. First, all the tokens adjacent to nasal consonants were excluded to avoid the influence of nasal resonance. Second, we plotted the f0 track of each token and excluded tokens with f0 tracking errors, because a wrongly tracked f0 will result in errors in the frequency of all harmonics. Third, we plotted the time courses of the first three formants of each vowel, and excluded intervals with formant tracking errors, because wrongly tracked formants will cause errors in the formant correction procedure. After excluding all tokens or intervals with f0 and formant tracking errors, for each harmonic, we subtracted the energy amplification caused by its adjacent formants using the formulaby Iseli et al. (2007). Lastly, we calculated the spectral tilt of every two adjacent harmonics (H1*-H2*, H2*-H3*, H3*-H4*, H1K*-H2K*, H2K*-H3K*, H3K*-H4K*, H4K*-H5K*).

The values of H1*, H2*, H3*, and H4* are presented in Figure 1a. The values of H1K*, H2K*, H3K*, H4K*, and H5K* are presented in Figure 1b. We compare the spectral values in a mixed linear effect model with phonation as the fixed effect and participant as independent intercept. In the corpus, falsetto voice has a higher f0 than other phonations in general. Thus, we also added f0 as a controlling variable to examine whether the differences observed in the spectral tilt of different phonations areinherent to the phonations or can beexplained by theirf0 differences. The results of the statistical comparisons are in Table 1.

The important differences between falsetto and other phonations are in H1*-H2*, H2*-H3*, H3*-H4*, and H4K*-H5K*. Falsetto voice has higher H1*-H2* than modal and creaky voices, but does not differ from breathy voice. Falsetto voice has lower H2*-H3*, higher H3*-H4*, and lower H4K*-H5K* than all other phonations.

Measuring spectral energy for non-modal phonations has been challenging because the noise introduced by irregular vocal fold vibration and air turbulence can cause errors in f0 and formant tracking. The current study provides a methodological protocol for extracting phonation information from a corpus, adding new spectral parameters to the VoiceSauce program, and performing outlier detection and formant correction for spectral energy measurements.

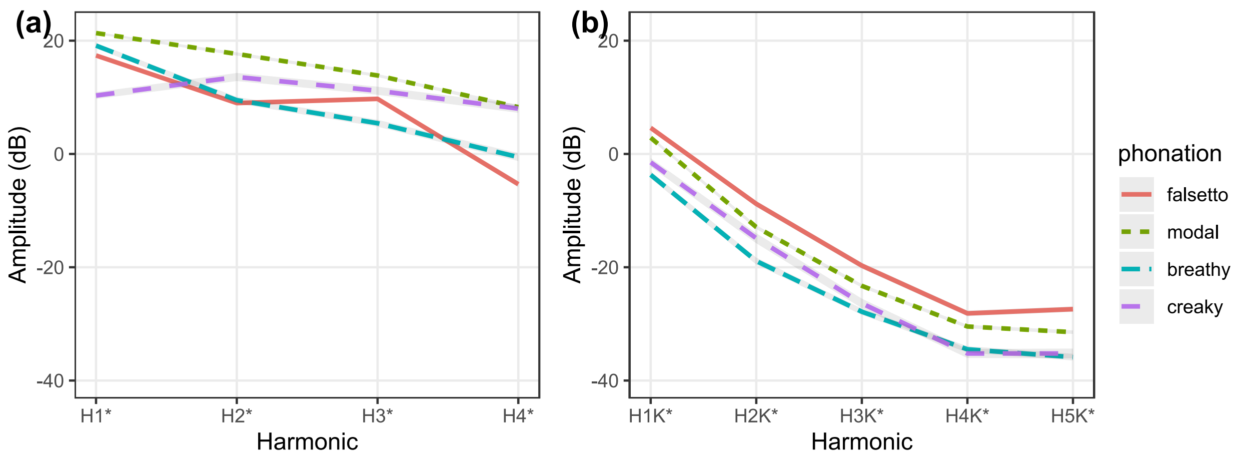

Figure 1. (a) Harmonic energy of H1*, H2*, H3*, and H4*; (b) Harmonic energy of H1K*, H2K*, H3K*, H4K*, and H5K*.

Table 1. Model comparison between falsetto (F) vs. modal (M), breathy (B), and creaky (C).

	H1*-H2*	H2*-H3*	H3*-H4*	H1K*-H2K*	H2K*-H3K*	H3K*-H4K*	H4K*-H5K*
no f0 control	F > M, C	F < M, C, B	F > M, C, B	F < M	n.s.	F > M	F < M, C, B
f0 control	F > C; F < B	F < B	F > B	F < M	n.s.	F > M	F < M

References

Colton, R. H. (1972). Spectral characteristics of the modal and falsetto registers. *Folia Phoniatrica*, *24*, 337–344.

Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, *121*(4), 2283–2295.

Kreiman, J., Shue, Y.-L., Chen, G., Iseli, M., Gerratt, B. R., Neubauer, J., & Alwan, A. (2012). Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *Journal of the Acoustical Society of America*, *132*, 2625–2632.

Shue, Y.-L. (2010). *The voice source in speech production: Data, analysis and models* [PhD Thesis]. UCLA.

Titze, I. R. (1994). *Principles of Voice Production*. Prentice Hall Inc.

van den Berg. (1968). Mechanisms of the larynx and the laryngeal vibrations. In *Manual of Phonetics* (pp. 278–308). North Holland.

Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America*, *139*(4), 1493–1507.